The Efficiency of Mixing in Turbulent Patches: Inferences from Direct Simulations and Microstructure Observations

W. D. SMYTH, J. N. MOUM, AND D. R. CALDWELL

College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon

(Manuscript received 17 April 2000, in final form 28 August 2000)

ABSTRACT

The time evolution of mixing in turbulent overturns is investigated using a combination of direct numerical simulations (DNS) and microstructure profiles obtained during two field experiments. The focus is on the flux coefficient Γ, the ratio of the turbulent buoyancy flux to the turbulent kinetic energy dissipation rate ϵ. In observational oceanography, a constant value $\Gamma = 0.2$ is often used to infer the buoyancy flux and the turbulent diffusivity from measured ϵ. In the simulations, the value of Γ changes by more than an order of magnitude over the life of a turbulent overturn, suggesting that the use of a constant value for Γ is an oversimplification. To account for the time dependence of Γ in the interpretation of ocean turbulence data, a way to assess the evolutionary stage at which a given turbulent event was sampled is required. The ratio of the Ozmidov scale L_O to the Thorpe scale L_T is found to increase monotonically with time in the simulated flows, and therefore may provide the needed time indicator. From the DNS results, a simple parameterization of Γ in terms of L_O/L_T is found. Applied to observational data, this parameterization leads to a 50%–60% increase in median estimates of turbulent diffusivity, suggesting a potential reassessment of turbulent diffusivity in weakly and intermittently turbulent regimes such as the ocean interior.

1. Introduction

In this paper, we investigate the efficiency of mixing by turbulent patches in the ocean thermocline. In particular, we are interested in variations of the flux coefficient Γ in time over the life cycle of a turbulent event. (The flux coefficient is closely related to mixing efficiency: it is the ratio of work done against gravity to energy dissipated via friction.) Our approach employs a combination of microstructure observations and direct numerical simulations.

In oceanic applications, Γ is often treated as a constant, usually with the value 0.2, and used to estimate heat fluxes from measurements of the kinetic energy dissipation rate ϵ (e.g., Smyth et al. 1996). However, an accumulation of observational evidence suggests that Γ is, in fact, significantly variable (Gargett and Moum 1995; Moum 1996a; Ruddick et al. 1997). Many factors have the potential to influence Γ; examples include the Prandtl number and the strength of the ambient stratification. Here, we investigate the possibility that Γ evolves systematically in time as turbulent overturns grow, break, and decay.

We consider the particular case of Kelvin–Helmholtz (KH) billows, thought to be a common source of turbulence in geophysical flows (e.g., DeSilva et al. 1996). The steady compressive strain that develops in the braids separating adjacent vortices generates sharp scalar gradients, and therefore rapid mixing (e.g., Broadwell and Breidenthal 1982). We find that preturbulent billows mix at a rate comparable to that occurring after the onset of turbulence. Because little energy is lost to viscosity prior to transition, the flux coefficient is actually larger in the laminar state than in the turbulent state. The evolution of mixing efficiency in Kelvin–Helmholtz billows has also been described independently by Caulfield and Peltier (2000) and Staquet (2000), both of which appeared after the present work was submitted. Our results extend those of Caulfield and Peltier (2000) and Staquet (2000) in two important ways: first, we investigate the oceanographically relevant case $\mathrm{Pr} > 1$; second, we provide a physical explanation for the time dependence of the mixing efficiency.

We investigate the time dependence of Γ using two very different approaches: direct numerical simulations (DNS) and field observations of turbulent microstructure. These two methods were chosen because their respective strengths and weaknesses are complementary. The time dependence of Γ is easy to assess in DNS output. However, the validity of DNS as a model of ocean turbulence is limited to low Reynolds number and specified initial and boundary conditions. In contrast,

Corresponding author address: Dr. William D. Smyth, College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331.
E-mail: smyth@oce.orst.edu

microstructure measurements sample real ocean turbulence, but neither Γ nor stage of evolution can be measured directly. These limitations of the observational data can be circumvented by means of two assumptions:

1) There exists an ''observational clock,'' a measureable quantity that varies monotonically as a turbulent event evolves, and can therefore be used to calibrate the age of the event.
2) Γ can be approximated by the Osborn–Cox formula $\Gamma_d = \kappa CN^2/\epsilon$, where κ is the scalar diffusivity, C is the Cox number (to be defined later), and N is the Brunt–Väisälä frequency (Osborn and Cox 1972; Oakey 1982). Although the underlying assumptions are seldom valid in geophysical turbulence, the Osborn–Cox approximation is usually assumed to provide a useful estimate of Γ.

Insofar as assumptions (1) and (2) are valid, the time dependence of Γ can be assessed using observational data. Here, DNS results will be used to examine the time dependence of Γ directly and also to test assumptions (1) and (2) above. Our expression for the observational clock is a slight variation on the ratio of Ozmidov scale to Thorpe scale, denoted $\hat{R}_{OT}$, which has been suggested previously as an indicator of patch age (e.g., Wijesekera and Dillon 1997). Having established the range of validity of these assumptions, we then proceed to examine the relationship between Γ_d and $\hat{R}_{OT}$ in the observational data. The results suggest a parameterization for Γ in terms of $\hat{R}_{OT}$. We complete the study by applying this parameterization to the observational data. This results in a 50%–60% increase in experiment mean values of turbulent diffusivity.

Our research plan assumes that DNS of breaking KH billows furnishes a valid model of thermocline patches. This depends both on the accuracy of our numerical model and on the validity of the notion that patches are driven by shear instability. The accuracy of the model has been established elsewhere (Smyth and Moum 2000a,b; Smyth 1999). The idea that patches are driven by shear instability has been suggested by many authors (e.g., Osborn 1980), but has not been tested. Thus, we begin by comparing various turbulence statistics from DNS of KH waves with observations of thermocline patches. Although the Reynolds numbers attained in the simulations are at the small end of the range of observations, most statistical indicators of turbulence dynamics compare favorably.

Section 2 describes the DNS model, the microstructure observations, and the problem of comparing the two in a meaningful way. An overview of turbulence evolution in a typical simulation is given in this section. An important technical issue that arises is the appropriate choice for the bulk density gradient for use in computing the Ozmidov scale, and we propose a new method for defining that gradient. Section 3 is devoted to the comparison of turbulence statistics between the model and the observations. Having established the applicability of the model, we then proceed to evaluate the time dependence of $\hat{R}_{OT}$ (i.e., to test assumption 1 above). In section 4, we examine the evolution of flux coefficient, and show that newly formed, preturbulent overturns can contribute significantly to the net potential energy gain. We then compare the flux coefficient to its approximation Γ_d in order to test assumption 2 above. Next, we examine $\Gamma_d(\hat{R}_{OT})$ in observational data, confirming that preturbulent billows contribute significantly to the net potential energy gain. Finally, we discuss the implications of our results for the estimation of turbulent diffusivities from observational data. Our conclusions are summarized in section 5.

2. Methodology

a. DNS of breaking Kelvin–Helmholtz billows

The numerical methods employed to generate the present dataset have been described in detail previously (Smyth 1999; Smyth and Moum 2000a,b); here, we provide a brief summary. Our mathematical model employs the Boussinesq equations for velocity, density, and pressure in a nonrotating physical space measured by the Cartesian coordinates x, y, and z:

$$\mathbf{u}_{,t} = \mathbf{u} \times (\boldsymbol{\nabla} \times \mathbf{u}) - \boldsymbol{\nabla}\Pi + g\theta\mathbf{k} + \nu\nabla^2\mathbf{u}, \quad (1)$$

$$\Pi = \frac{p}{\rho_o} + \frac{1}{2}\mathbf{u} \cdot \mathbf{u}. \quad (2)$$

Subscripts following commas denote partial differentiation. The variable p represents pressure and ρ_o is a constant characteristic density. The thermodynamic variable θ represents the fractional specific volume deviation, or minus the fractional density deviation, that is, $\theta = -(\rho - \rho_o)/\rho_o$. In a fluid where density is controlled only by temperature, θ is proportional to the temperature deviation (with proportionality constant equal to the thermal expansion coefficient). The gravitational acceleration g has the value 9.8 m s^{-2}, and $\hat{k}$ is the vertical unit vector. Viscous effects are represented by the usual Laplacian operator, with kinematic viscosity $\nu = 1.0 \times 10^{-6}$ m^2 s^{-1}.

The augmented pressure field Π is specified implicitly by the incompressibility condition

$$\boldsymbol{\nabla}\cdot\mathbf{u} = 0, \quad (3)$$

and the scalar θ evolves in accordance with

$$\theta_{,t} = -\mathbf{u}\cdot\boldsymbol{\nabla}\theta + \kappa\nabla^2\theta, \quad (4)$$

in which κ represents the molecular diffusivity of θ. For thermal stratification in the ocean, a typical value for κ is 1.4×10^{-7} m^2 s^{-1}. In the present simulations, we vary κ in order to gain access to a wider range of Reynolds numbers.

We assume periodicity in the horizontal dimensions:

$$f(x + L_x, y, z) = f(x, y + L_y, z) = f(x, y, z), \quad (5)$$

in which f is any solution field and the periodicity intervals L_x and L_y are constants. At the upper and lower boundaries ($z = \pm 1/2L_z$), we impose an impermeability condition on the vertical velocity:

$$w|_{z=\pm 1/2L_z} = 0, \tag{6}$$

and zero-flux conditions on the horizontal velocity components u and v and on θ:

$$u_{,z}|_{z=\pm 1/2L_z} = v_{,z}|_{z=\pm 1/2L_z} = \theta_{,z}|_{z=\pm 1/2L_z} = 0. \tag{7}$$

These imply a condition on Π at the upper and lower boundaries:

$$[\Pi_{,z} - g\theta]_{z=\pm 1/2L_z} = 0. \tag{8}$$

The domain dimensions were chosen based on the known geometry of the outer scales of KH billows. The domain length L_x was chosen so as to accommodate two wavelengths of the primary KH instability. Inclusion of only a single wavelength would save memory, but the merging of adjacent vortices, which plays an important role in the development of turbulence, would then be suppressed. The wavelength was determined from linear stability analyses (e.g., Smyth and Peltier 1989). Here L_z was chosen as $L_x/2$, sufficient to prevent the upper and lower boundaries from influencing the turbulence evolution significantly (Smyth and Peltier 1993). The domain width L_y was chosen to be $L_x/8$. This aspect ratio is justified in light of the small spanwise wavelength of the dominant secondary instabilities that lead the flow to a turbulent state (Klaassen and Peltier 1991; Caulfield and Peltier 2000). Sensitivity experiments in which L_y was halved and doubled yielded no difference in the turbulence statistics of interest here. The model is initialized with a parallel flow in which shear and stratification are concentrated in the shear layer, a horizontal layer surrounding the plane $z = 0$:

$$\tilde{u}(z) = \frac{u_o}{2}\tanh\frac{2z}{h_o}; \qquad \tilde{\theta}(z) = \frac{\theta_o}{2}\tanh\frac{2z}{h_o}. \tag{9}$$

The constants h_o, u_o, and θ_o represent the initial thickness of the shear layer and the changes in velocity and density across it. These constants can be combined with the fluid parameters ν and κ and the geophysical parameter g to form three dimensionless groups whose values determine the stability of the flow at $t = 0$:

$$\mathrm{Re}_o \equiv \frac{u_o h_o}{\nu}, \qquad \mathrm{Ri}_o \equiv \frac{g\theta_o h_o}{u_o^2}, \qquad \mathrm{Pr} \equiv \frac{\nu}{\kappa}. \tag{10}$$

The initial macroscale Reynolds number Re_o expresses the relative importance of viscous effects. In the present simulations, Re_o is of order a few thousand, large enough that the initial instability is nearly inviscid. The bulk Richardson number Ri_o quantifies the relative importance of shear and stratification. If $\mathrm{Ri}_o < 0.25$, the initial mean flow possesses unstable normal modes (Hazel 1972). Some authors have defined Re_o and Ri_o based on $u_o/2$, $\theta_o/2$, and $h_o/2$ rather than u_o, θ_o, and h_o, (e.g., Smyth and Peltier 1993; Caulfield and Peltier 2000). For comparison with such studies, our values of Re_o should be divided by four (Ri_o is unchanged). In order to obtain a fully turbulent flow efficiently, we add to the initial mean profiles a perturbation field designed to excite the most-unstable primary and secondary instabilities. Details of the initial perturbation are discussed elsewhere (Smyth and Moum 2000b). Care is taken to ensure that the initial perturbation is weak enough to obey linear physics. In that case, the precise form of the perturbation has little effect on the statistical quantities of interest here.

TABLE 1. Parameter values describing a sequence of six simulations of breaking Kelvin–Helmholtz billows. Here N_x, N_y, and N_z are the array dimensions, L_x, L_y, and L_z are the domain dimensions, and Δx ($= \Delta y = \Delta z$) is the grid interval. Initial conditions are characterized by the length scale h_o, the velocity scale u_o, and the temperature scale θ_o. The Prandtl number is Pr. The inital bulk Reynolds number is Re_o. Results from simulations 4–6 are displayed only in Figs. 10 and 11.

	Simulation					
	1	2	3	4	5	6
Pr	7	4	1	7	1	2
Re_o	1354	1967	4978	701	1965	1354
N_x	512	512	512	256	256	512
N_y	64	64	64	32	64	64
N_z	256	256	256	128	128	256
L_x (m)	2.73	3.29	5.24	1.96	3.29	2.73
L_y (m)	0.34	0.41	0.65	0.25	0.82	0.34
L_z (m)	1.36	1.63	2.62	0.98	1.63	1.36
Δx (10^{-2} m)	0.53	0.64	1.03	0.77	1.29	0.53
h_o (m)	0.20	0.24	0.38	0.14	0.24	0.20
u_o (10^{-3} m s^{-1})	6.38	8.34	13.27	4.98	8.34	6.38
θ_o (10^{-6})	2.00	2.41	3.83	1.44	2.41	2.00
Symbol	●	*	○	●	○	*

Spatial discretization is Fourier pseudospectral in both horizontal direction, second-order centered finite differences in the vertical. The fields are stepped forward in time using a second-order Adams–Bashforth method. Spectra confirming the accuracy of the numerics, along with comparisons with laboratory experiments, are provided elsewhere (Smyth and Moum 2000a,b; Smyth 1999).

For the present analyses, we have chosen six simulations using different values of Re_o, Ri_o, and Pr, as described in Table 1. Molecular diffusivity was varied so that the Prandtl number ranged from 1 to 7 (the highest value being most realistic for heat diffusing in water). The choice of Re_o was constrained by the need to maintain adequate resolution of the small scales; values ranged from 701 to 4978, with larger values corresponding to smaller Pr. The initial bulk Richardson number, Ri_o, was set to 0.08 in all cases reported here. [Some effects of varying Ri_o are discussed by Caulfield and Peltier (2000) and Staquet (2000).] As the evolution progressed, both the bulk Richardson and Reynolds numbers increased in proportion to the increasing depth of the shear layer (Smyth and Moum 2000b). For all runs, the initial maximum shear u_o/h_o was set to 0.035 s^{-1}, and the initial minimum frequency $N = \sqrt{g\theta_o/h_o}$

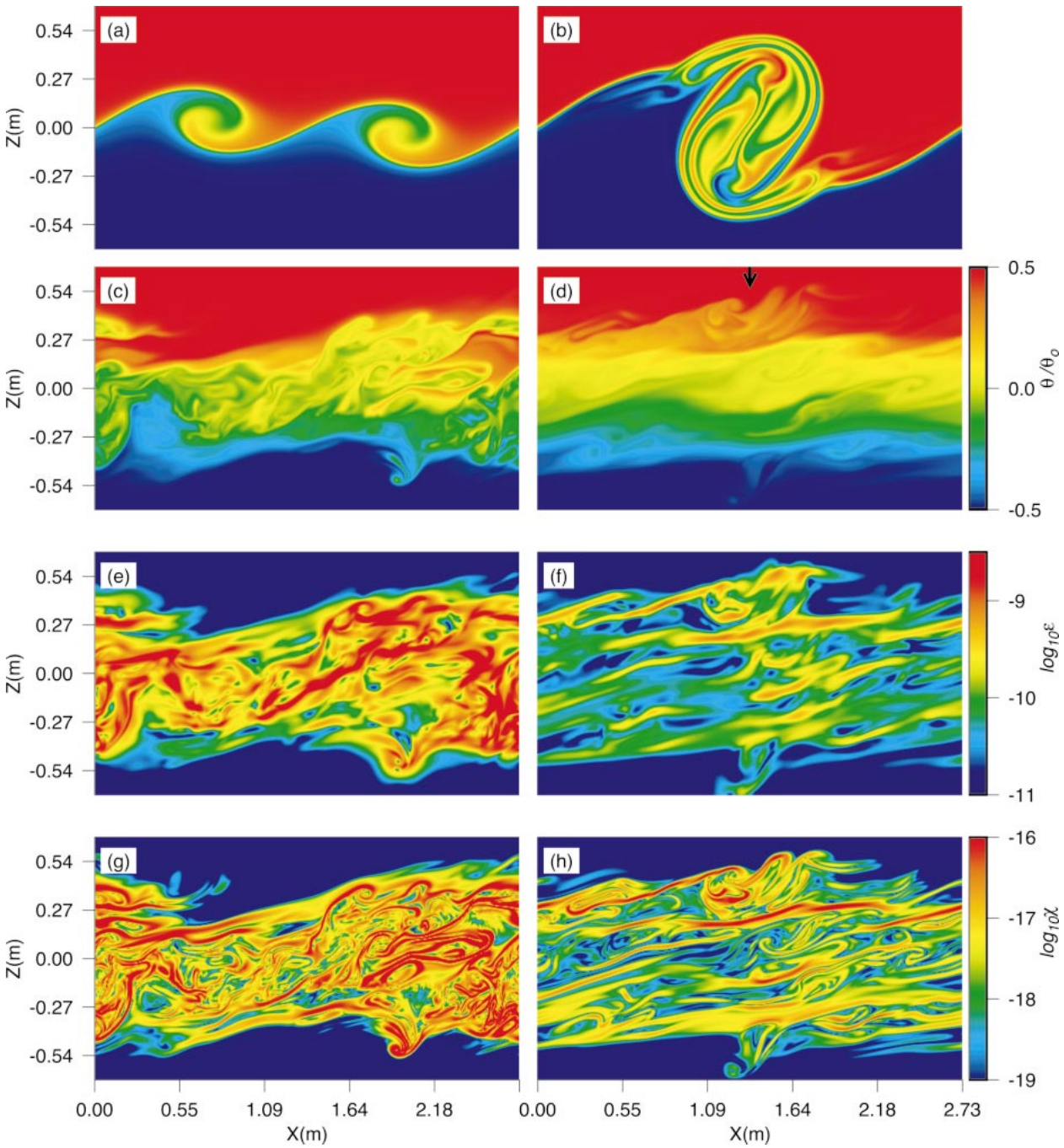

FIG. 1. Cross sections of instantaneous flow fields from simulation 1. Temperature is represented in nondimensional form as θ/θ_o in (a)–(d): (a) $t = 565$ s, (b) $t = 1414$ s, (c) $t = 4242$ s, and (d) $t = 6222$ s. The arrow in (d) corresponds to Fig. 2. The turbulent kinetic energy dissipation rate ϵ is shown at (e) $t = 4242$ s and (f) $t = 6222$ s. The scalar variance dissipation rate χ is also shown at (g) $t = 4242$ s and (h) $t = 6222$ s.

was set to 0.0099 s^{-1}. Most of our conclusions are demonstrated using only the first three simulations shown in Table 1.

Cross sections of the evolving flow fields in simulation 1 are shown in Fig. 1. Figures 1a–d show scaled temperature cross sections taken at four instants during the simulation. Figures 1e and 1f show the kinetic energy dissipation rate ϵ at the latter two instants, and Figs. 1g and 1h show the scalar variance dissipation rate at the same two instants. [The dissipation rates ϵ and χ are defined explicitly in section 2c(3).] At $t = 565$ s (Fig. 1a), a pair of primary KH billows has grown and

begun to overturn. By $t = 1414$ s, the primary billows are in the process of merging to form a single large billow (Fig. 1b). Note the extremely sharp gradients occurring to the right and left of the billow; these are sites of rapid scalar mixing. At $t = 4242$ s, the transition to turbulence is complete, and the fields exhibit a highly complex structure (Figs. 1c,e,g). The scalar dissipation rate varies on much smaller scales than does the kinetic energy dissipation rate in consequence of the high Prandtl number. By $t = 6222$ s, turbulence has decayed significantly (Figs. 1d,f,h). Gravity waves propagate on localized layers of strong stratification. Between the stratified layers, the gradient Richardson number is small, and secondary shear instabilities drive small-scale overturns. This layered structure is commonly observed in the ocean (e.g., Phillips 1972; Posmentier 1977; Ruddick et al. 1989; Balmforth et al. 1998).

b. Microstructure observations

Two observational datasets are used in the present study. The first was collected during the FLUX STATS cruise (FLX91) on 1–7 May 1991 approximately 1000 km off the coast of northern California (30°N, 135°15′W). Measurements made by the free-falling profiler Chameleon included pressure (depth), temperature, conductivity, temperature–gradient fluctuations using an FP07 microbead thermistor, horizontal velocity–gradient fluctuations using airfoil shear probes, and vertical velocity fluctuations using a Pitot tube (Moum 1990; Hebert and Moum 1994; Moum 1996a,b). Because the thermistor did not resolve the entire spectrum of temperature fluctuations at the fall speed used ($\sim$1 m s^{-1}), the calculation of temperature variance dissipation rate required corrections based on the Batchelor form of the scalar variance spectrum (Peters et al. 1988).

From 400 profiles, patches were selected using principles similar to those of Moum (1996b). (a) Using the density profile, parts of the water column were eliminated that had no Thorpe displacements [see section 2c(1)] across them, leaving patch candidates. (b) Patches thinner than 0.15 m were rejected and (c) patches separated by less than 0.15 m were joined. (d) Patches with excessive noise in the microstructure shear were eliminated. (e) Patches in which density and temperature profiles were very different were removed by requiring that L_T^σ, the estimate of the Thorpe scale [section 2c(1)] calculated from the density profile was one to four times L_T, the estimate calculated from the temperature profile. (Patches for which $L_T^\sigma/L_T < 1$ represent salinity intrusions, while those for which $L^\sigma/L_T > 4$ occur in weakly stratified locations where noise in the conductivity sensor is a factor.) (f) Patches with mean $\epsilon < 4 \times 10^{-9}$ m^2 s^{-3} were discarded. A total of 3425 patches remained. This selection is considerably larger than the selection arrived at by Moum (1996b), which was derived from the same dataset using subjective methods. The increase in selection size is due mainly to the fact that the selection process has now been automated, but is also due in part to two differences in the selection criteria. First, we have dispensed with the requirement that energy-containing scales not exceed 3 m, partly on the basis of the DNS results. Second, we have added the condition e above.

A second set of 1155 patches was selected from the Tropical Instability Wave Experiment (TIWE), conducted at the equator at 140°W in December 1991 (Lien et al. 1995). Only data from yeardays 324–327 were used. The resulting patches were all located in the upper-equatorial thermocline between 60- and 200-m depth, and spanned both the upper and lower flanks of the Pacific Equatorial Undercurrent.

c. Model–data comparisons

In this subsection, we define terms needed for the model–data comparisons, and also address some technical subtleties that arise therein.

1) THORPE REORDERING

In order to obtain an estimate of the ''background'' temperature structure from a patch observation, that is, using only a single vertical profile, we reorder that profile into a statically stable configuration (Thorpe 1977). The resulting function, which we denote $\theta_T(z)$, represents an estimate of the temperature profile that would be observed if the fluid were allowed to relax adiabatically to its state of minimum potential energy. The distance each parcel traverses to effect this reordering is denoted $-\delta_T$, where δ_T is an estimate of the vertical displacement from equilibrium depth. (In this study, we employ the convention that the displacement δ_T is associated with the *observed* position of a fluid parcel, not with its position in the reordered profile.) These estimates are of limited validity, of course, because they use only information from a single vertical profile. An alternative procedure, in which the entire three-dimensional volume is reordered, has been suggested by Winters et al. (1995) (also see Scinocca 1995) and employed in the present context by Smyth and Moum (2000b), Caulfield and Peltier (2000), and Staquet (2000). The one-dimensional version, besides being applicable to profiler data, has the virtue of registering only temperature deviations associated with overturns. For example, a field of nonoverturning wave motions is removed by three-dimensional reordering, but is entirely invisible to the one-dimensional procedure.

In the analysis of microstructure profiles, the Thorpe displacement is the basis for defining the boundaries of a given turbulent patch (e.g., Moum 1996b). Over most of a typical profile, the local stratification will be stable and the displacement zero. A ''patch'' is therefore defined as a region of continuously nonzero δ_T. In analyzing the DNS data, we treat the data subsample at each time t and lateral location (x, y) as if it were a

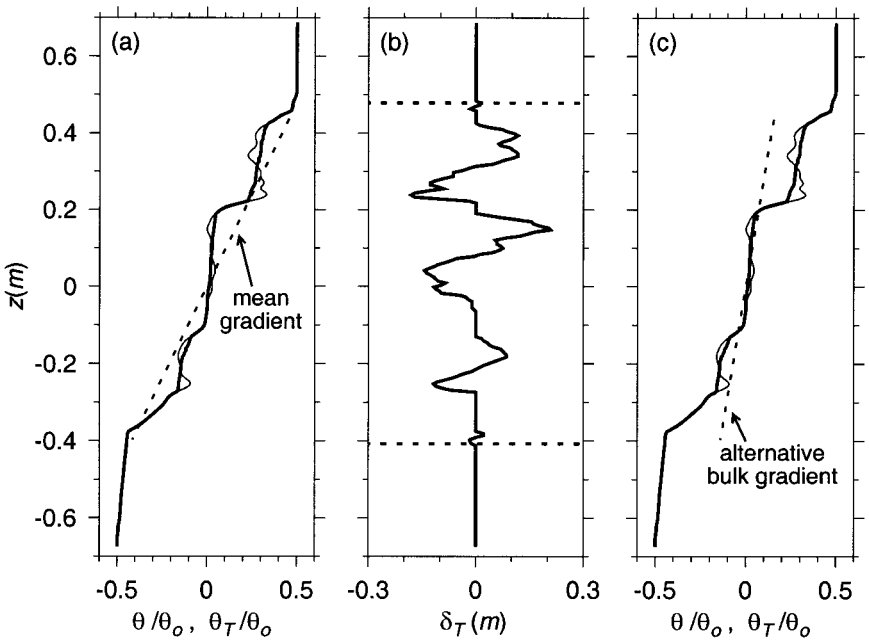

FIG. 2. Alternative definitions of the bulk temperature gradient in layered turbulence. Scaled temperature profiles were taken from the Pr = 7 case; $x = L_x/2$, $y = L_y/2$, $t = 6222$ s (arrow in Fig. 1d). (a) The thin (thick) solid curve represents the original (Thorpe-reordered) temperature profile. The dashed line shows the mean gradient $\langle\hat{\theta}_z\rangle$ taken over the extent of the patch. (b) The Thorpe displacement δ_T. Horizontal dashed lines identify upper and lower patch boundaries. (c) Same as (a) except that the dashed line now represents the alternative bulk gradient $\hat{\theta}_z$.

microstructure profile. We search inward from the upper and lower boundaries to find the outermost locations at which $\delta_T \neq 0$. (This procedure is complicated slightly due to numerical noise in the unstratified fluid outside the turbulent region, which can lead to large, physically meaningless Thorpe displacements. We therefore set δ_T to zero in any region adjacent to the upper or lower boundary in which the reordered temperature profile is uniform, in the sense that its vertical gradient is less than $0.02\theta_o/h_o$.) An example is given in Fig. 2. Figure 2a shows the original and reordered temperature profiles; Fig. 2b shows the Thorpe displacement and the patch boundaries. A useful measure of the typical parcel displacement is provided by the Thorpe scale:

$$L_T = \langle\delta_T^2\rangle^{1/2}. \tag{11}$$

Here and elsewhere, angle brackets denote a vertical average taken over the region between the patch boundaries, so that L_T represents the root-mean-square (rms) Thorpe displacement. For the example shown in Fig. 2, $L_T = 0.08$ m.

2) BULK STRATIFICATION

A universal problem in the interpretation of patch observations is the definition of a ''bulk'' density gradient that quantifies the stratification against which the turbulence must work (e.g., Hebert et al. 1992). Most theories of stratified turbulence assume a uniform background gradient, a circumstance that is never realized exactly in nature. The usual approach is to simply average the reordered density gradient between the upper and lower boundaries of the patch (e.g., Moum 1996b). This approach works well in most cases, but can give ambiguous results when the patch contains more than one distinct overturning region.

The example shown in Fig. 2 is taken from late in a simulation, at a point where the turbulence is strongly layered (Fig. 1d). Here, the average gradient (dashed line in Fig. 2a) represents the stratification only in a coarse sense. Close inspection reveals that the reordered density gradient alternates between regions of strong stability in which there is no overturning, and regions of weak stability in which overturning occurs (Fig. 2b). It may be argued that the work required for overturning is determined by the relatively weak stratification within the overturning layers. In this sense, the gradient shown in Fig. 2a overestimates the required bulk gradient. Figure 2c shows an alternative choice that more closely represents the gradients in the overturning regions. We now describe the alternative approach.

A characteristic property of turbulence in uniform stratification, and one that guides our choice for the revised bulk gradient, is the equality of the Thorpe scale and the Ellison scale. The latter is given by

$$L_E = \frac{\langle(\theta - \theta_T)^2\rangle^{1/2}}{\widehat{\theta}_{,z}}, \tag{12}$$

in which the numerator is the rms fluctuation from the background temperature and $\widehat{\theta}_{,z}$ is a measure of the background gradient, which we leave unspecified for the moment. Here and elsewhere, the subscript preceded by a comma indicates differentiation. Traditionally, $\widehat{\theta}_{,z}$ is specified as the mean gradient across the turbulent layer. For the general case in which the background stratification is not uniform, equality of the Ellison and Thorpe scales is achieved by choosing

$$\widehat{\theta}_{,z} = \frac{\langle(\theta - \theta_T)^2\rangle^{1/2}}{L_T}. \tag{13}$$

This is the gradient shown in Fig. 2c. The associated Brunt–Väisälä frequency is given by

$$\hat{N}^2 = g\widehat{\theta}_{,z}. \tag{14}$$

Here and elsewhere, the hat indicates a quantity computed using the alternative bulk gradient $\widehat{\theta}_{,z}$.

A more informative derivation of (13) follows from the fact that

$$\langle(\theta - \theta_T)^2\rangle = \langle\theta_{T,z}^2\delta_T^2\rangle - \langle\theta_{T,z}\theta_{T,zz}\delta_T^3\rangle + \cdots, \tag{15}$$

which is derived via a Taylor series expansion in δ_T. Because the reordered temperature profile tends to be nearly linear in overturning regions (e.g., Fig. 2), the first term in the expansion generally dominates. Substituting (15) and (11) into (13), we have

$$\widehat{\theta}_{,z}^{\,2} \approx \frac{\langle\theta_{T,z}^2\delta_T^2\rangle}{\langle\delta_T^2\rangle}, \tag{16}$$

that is, $\widehat{\theta}_{,z}$ approximates the rms average of the temperature gradient weighted by the squared parcel displacement. Thus, $\widehat{\theta}_{,z}$ satisfies the intuitive notion that the bulk gradient should be most representative of the overturning regions.

Our alternative bulk gradient has the additional virtue of being relatively insensitive to the choice of patch boundaries. If the patch boundaries are chosen carefully so as to include only the overturns, $\widehat{\theta}_{,z}$ is nearly identical to the mean gradient $\langle\theta_{,z}\rangle$. However, if the boundaries also enclose some stable, nonoverturning fluid as in Fig. 2, $\widehat{\theta}_{,z}$ is unaffected, whereas the mean gradient is undesirably increased (e.g., Hebert et al. 1992). If the patch boundaries enclose more than one overturning layer, $\widehat{\theta}_{,z}$ provides an average over those layers while ignoring the intervening stable fluid. The problem of deciding whether or not to treat each overturning region as a separate patch is thus alleviated. Early in each simulation, overturning is found throughout the shear layer (e.g., Figs. 1b, 1c). In these cases, $\widehat{\theta}_{,z}$ is indistinguishable from the mean gradient. Later in the life cycle, though, when layering is advanced (e.g., Fig. 1d, Fig. 2), $\widehat{\theta}_{,z}$ provides an improved estimate of the background gradient in overturning regions.

Ultimately, the best choice of averaging operator depends on the application. Here, our eventual goal is to identify measurable quantities that can be used to gauge the evolutionary stage of an observed turbulent event. We will see that this goal is best achieved when bulk stratification is computed as $\widehat{\theta}_{,z}$.

3) TURBULENT DISSIPATION RATES AND RELATED QUANTITIES

The turbulent kinetic energy dissipation rate is given by $\epsilon = 2\nu s_{ij}s_{ij}$, where

$$s_{ij} = \frac{1}{2}(u'_{i,j} + u'_{j,i}). \tag{17}$$

Primes indicate deviations from the horizontal mean. The horizontal mean is removed in order to approximate the effect of high-pass filtering that is performed on the observational data.

When motions in the dissipation range (scales less than roughly 10 times the Kolmogorov scale) are isotropic, ϵ may be approximated as follows (Hinze 1975):

$$\epsilon_z = \frac{15}{4}\nu(u'^2_{,z} + v'^2_{,z}). \tag{18}$$

This approximation can be calculated using only the two components of the strain tensor that are available from vertical profiles measured with a shear probe. When the buoyancy Reynolds number, $R_b = \epsilon/\nu N^2$, is insufficiently large, however, the isotropic approximation can give misleading results. Gargett et al. (1984) found that an isotropic inertial subrange becomes clearly evident only when R_b exceeds 200. Smyth and Moum (2000a) demonstrated that dissipation range statistics become isotropic when R_b exceeds $O(10^2)$. Although R_b is frequently less than 100 in both model and observations, isotropy is assumed of necessity in the observations. For consistency, we do the same in analyzing the model data. The buoyancy Reynolds number is thus approximated by

$$\hat{R}_b = \frac{\langle\epsilon_z\rangle}{\nu\hat{N}^2}. \tag{19}$$

Note that the isotropic approximation to ϵ is needed for the calculation of $\hat{R}_b$. This can lead to a dangerous situation wherein isotropy is incorrectly assumed due to an overestimate of $\hat{R}_b$ (Smyth and Moum 2000a).

The dissipation rate for temperature variance is given by

$$\chi = 2\kappa\langle|\nabla\theta'|^2\rangle, \tag{20}$$

and the corresponding isotropic approximation is

$$\chi_z = 6\kappa\langle\theta'^2_z\rangle \tag{21}$$

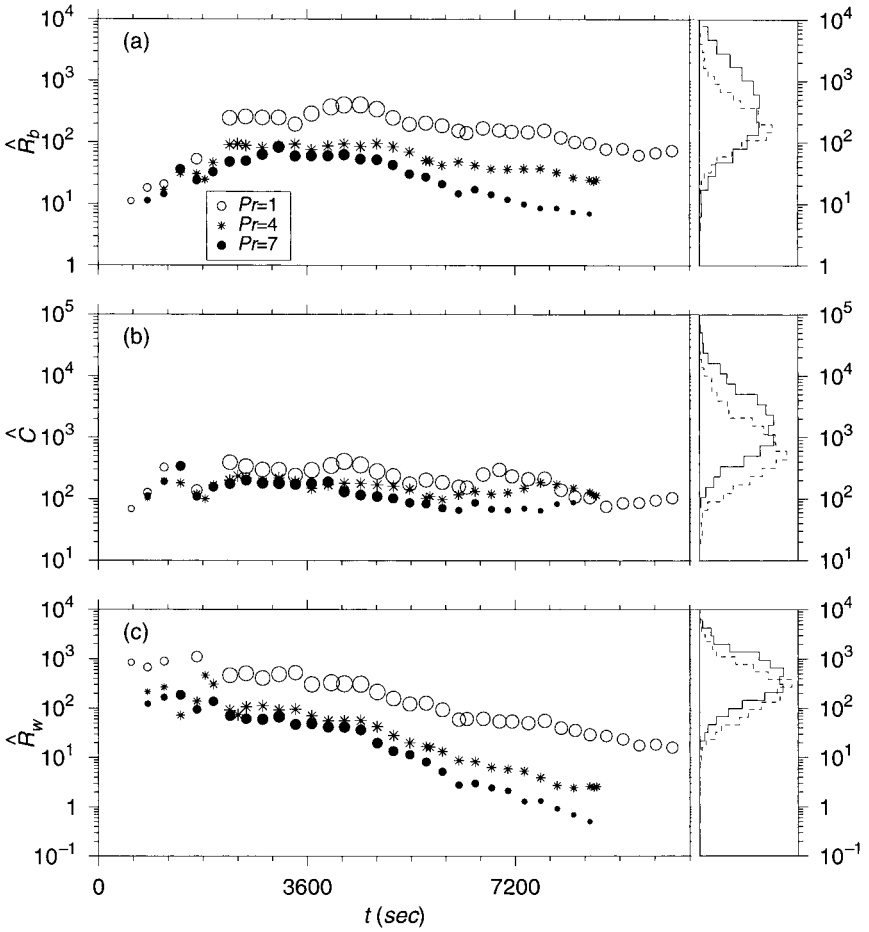

FIG. 3. Evolution in time of (a) the buoyancy Reynolds number, (b) the Cox number, and (c) the vertical Reynolds number. Symbols indicate the first three DNS runs: Pr = 1 (open circles), Pr = 4 (asterisks), and Pr = 7 (bullets). Symbol size is proportional to $\log_{10}\hat{R}_b$. Bootstrap 95% confidence limits on each average are smaller than the symbol size. At the right of each figure is a histogram of the corresponding quantity taken from the observational data. In each case, the solid (dashed) curve represents the TIWE (FLX91) data. The scalings of the histograms are arbitrary.

The Cox number is defined using the isotropic formula $\hat{C} = 3\langle\theta'^2_z\rangle/\widehat{\theta^2_{,z}}$, where $\widehat{\theta^2_{,z}}$ is as defined in (13).

3. Quantitative comparisons of turbulence statistics

This section has two main purposes. First, we assess the applicability of the DNS model by means of statistical comparison with observational data. We show that, although the Reynolds numbers attained by the model lie at the low end of the range of observed values, most statistical indicators of turbulence dynamics compare favorably. Our second purpose in this section is to use the DNS data to validate the first of the two assumptions needed to extract the time dependence of the flux coefficient from the observations (cf. section 1).

The results from the simulations are saved at regular time intervals. At each such time, the flow is sampled by taking 512 vertical profiles through the computational domain (64 in the streamwise direction, 8 in the spanwise direction). The spatial separation between profiles is 8 grid intervals, or about 24 times the *minimum* Batchelor scale. Profiles spaced more closely than this are unlikely to be statistically independent. The results are then analyzed in a manner consistent with the analysis of ocean microstructure profiles. Occasional profiles contain no overturning and are therefore discarded. However, nearly every profile contains a patch, for which we record values of various quantities of interest such as the mean dissipation rate, the Thorpe scale, etc. Our main interest will be in the time evolution of the median values of these quantities for each simulation.

a. Reynolds and Cox numbers

Figure 3a shows the time evolution of the buoyancy Reynolds number. To compute each of these data points,

we obtained values of $\langle \epsilon_z \rangle$ and $\hat{N}^2$ for each profile, combined them to yield values (usually 512) of $\hat{R}_b$, then took the median. The different simulations are distinguished by symbols. Open circles represent the Pr = 1 simulation, asterisks denote the Pr = 4 case, and bullets represent the Pr = 7 case. Symbol size is proportional to the logarithm of $\hat{R}_b$. (That information is redundant in the case of Fig. 3a, but will help in the interpretation of other figures.) The observations provide no information about the stage of evolution, and are therefore represented by histograms positioned for comparison with the time-dependent DNS results. Thus, we see from Fig. 3a that the buoyancy Reynolds numbers attained by the simulated flows lie within the lower half of the range of the observations.

Values of the buoyancy Reynolds number as large as $O(10^4)$ have been observed in the upper-equatorial thermocline (also see Wijesekera et al. 1993; Wijesekera and Dillon 1997). Our Pr = 7 simulation attains a maximum buoyancy Reynolds number near 70, while the Pr = 1 cases reach maxima in the range 200–400. These maxima are within the range of the observations, and it may therefore be expected that the simulated patches will have properties in common with the patches observed in the ocean. A rough criterion for the dissipation range to be free of buoyancy effects is $\hat{R}_b > 20$. All of our simulations meet this criterion for a substantial time interval after the onset of turbulence.

Values of the Cox number attained in the simulations are confined to the smallest end of the observed range. Since the Cox number depends on small-scale temperature structure, the artifice of lowering the Prandtl number makes little difference. In the late stages of the simulations (and, presumably, in the smallest observed patches), both $\hat{C}$ and $\hat{R}_b$ are overestimated due to anisotropy effects (Smyth and Moum 2000a).

The limitations of the simulations are most obvious upon examination of the vertical Reynolds number (Fig. 3c), computed as $\hat{R}_w = w_{\rm rms}^2/\hat{N}\nu$, where $w_{\rm rms}^2$ is the mean-squared vertical velocity. The Pr = 1 case delivers realistic values of $\hat{R}_w$, but the Pr > 1 runs both start at the low end of the observed range and decrease from there. This result indicates caution in interpreting the late stages of the Pr > 1 runs: they may enter a physical regime that is not found in the observations. We will see that the model fares considerably better in the comparison of other turbulence characteristics.

b. Measures of patch size

Overall patch thickness, that is, the median vertical distance between the upper and lower patch boundaries, is denoted L_p. In all simulations, patch thickness grows rapidly at first, then much more slowly as turbulence subsides (Smyth and Moum 2000b). Our interest here is in the ratio of the patch thickness to other characteristic length scales. For example, the ratio of L_p to the Thorpe scale is a measure of the complexity of the large scales. L_p/L_T is near two when only a single overturn is present, but becomes larger when many large overturns are contained within the patch boundaries (Fig. 4a). Our DNS values of L_p/L_T all lie within the range of the observations. There is a general tendency for L_p/L_T to increase with time, which suggests that the smallest (largest) of the observed values correspond to young (old) turbulence.

Similar results obtain for the ratio of L_p to the maximum Thorpe scale (Fig. 4b). (Note that $L_{T\rm max}$ is not the maximum Thorpe displacement in the entire flow; it is the median of the maxima occurring in individual profiles.) The lower limit of unity corresponds to the presence of a single large overturn. As in the case of L_p/L_T, the ratio tends to increase with time, but remains within the range of observed values. A third measure of complexity is the ratio of maximum to rms Thorpe displacement, $L_{T\rm max}/L_T$ (Fig. 4c). This quantity remains within a restricted range of values that agrees very well with the observations.

The ratios L_p/L_T and $L_p/L_{T\rm max}$ do not increase monotonically with time. Large, rapid decreases at $t = 3500$ and 6500 s (particularly in the Pr = 1 case) are produced by the spontaneous generation of large vortices long after the initial instability has subsided. This is made possible, despite the absence of forcing, by the tendency of temperature to mix more throughly than momentum. That differential mixing causes the gradient Richardson number in the interior of the turbulent layer to become smaller than the critical value for shear instability, even though the bulk Richardson number is much larger (e.g., Scinocca 1995; Werne and Fritts 1999; Smyth and Moum 2000b; Caulfield and Peltier 2000).

An alternative measure of eddy size is provided by the buoyancy scale, $\hat{L}_b = w_{\rm rms}/\hat{N}$ (Fig. 5). This represents the vertical displacement that a fluid parcel could attain if all of its vertical kinetic energy were expended as work against gravity. For the Pr > 1 cases, $\hat{L}_b$ agrees very well with the Thorpe scale; for the Pr = 1 case, $\hat{L}_b$ generally exceeds L_T by a factor of two. All values are within the range of observations.

c. The energy-containing scale

In stationary turbulence, the size of the large eddies that carry most of the kinetic energy may be approximated as q^3/ϵ, in which q^2 is two-thirds of the turbulent kinetic energy per unit mass (e.g., Tennekes and Lumley 1972). In the microstructure observations, only the vertical component $w_{\rm rms}$ of the eddy-scale turbulent velocity fluctuations is measured. As a result, an assumption of large-scale isotropy is used to justify the approximation $q \approx w_{\rm rms}$. For consistency with the observations, we assume isotropy in the simulated flows and thus approximate q by $w_{\rm rms}$, and also ϵ by ϵ_z, denoting the resulting estimate as L_{En}, namely,

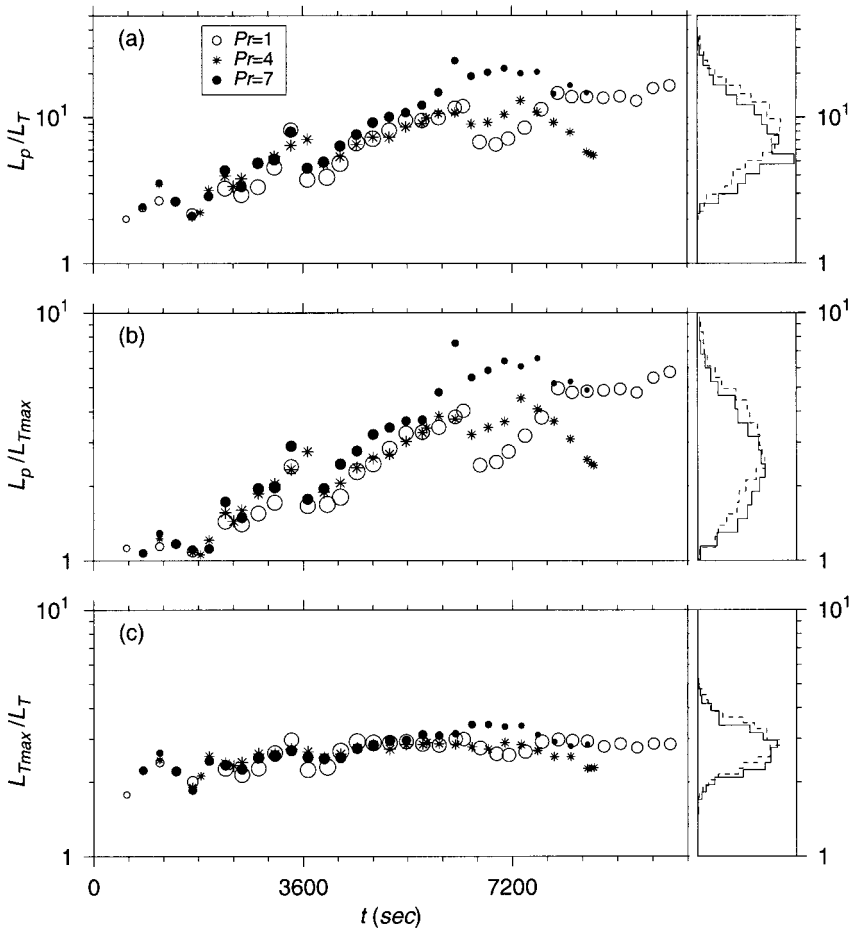

FIG. 4. Evolution of the ratios (a) L_p/L_T, (b) $L_p/L_{T\max}$, and (c) $L_{T\max}/L_p$ Symbols and histograms are defined as in Fig. 3.

$$L_{En} = \frac{w_{\rm rms}^3}{\langle \epsilon_z \rangle}. \tag{22}$$

The length scale L_{En} provides a very poor representation of the energy-containing scales in the preturbulent phase of the flow evolution (Fig. 6). Here, the disturbance carries a great deal of kinetic energy that has not yet begun to dissipate, so that L_{En} exceeds the Thorpe scale (a more direct measure of large eddy size) by 2–3 orders of magnitude. [This is the reason that we have dispensed with the patch selection criterion $L_{En} < 3$ m employed in Moum (1996a).] As turbulence develops, the ratio L_{En}/L_T decreases rapidly to values of order unity. In the final stages of decay, particularly in the Pr > 1 cases, the ratio decreases further to values significantly less than unity.

So far, our policy in analyzing the DNS results has been to employ only information available in measured microstructure profiles (except for the time). We now set that policy aside in order to investigate the effects of anisotropy on our estimates of the energy-containing scale. It is likely that the assumption of isotropy is often invalid, particularly in the large, energy-carrying scales (Smyth and Moum 2000a). In the late stages of our simulations, motions are predominantly streamwise, so that $w_{\rm rms}$ may severely underestimate q. To test this possibility, we compute L_{En} using all three components of the velocity fluctuations. For consistency, we also employ the full dissipation rate ϵ rather than ϵ_z, although the effect of anisotropy is much less pronounced here than on the velocity. We denote the result L'_{En}, namely,

$$L'_{En} = \frac{\langle q^2 \rangle^{3/2}}{\langle \epsilon \rangle}. \tag{23}$$

The result of these adjustments is that, after the onset of turbulence, L'_{En} remains remarkably proportional to L_T in all simulations, with proportionality constant slightly in excess of unity (Fig. 6b).

The results shown in Fig. 6 suggest that observations of $L_{En}/L_T < 1$ are due in part to anisotropy. Given that KH billows extract energy from the background shear by virtue of anisotropy in the production range, it seems likely that these scales are significantly anisotropic at all Reynolds numbers. In contrast, Gargett (1988) proposes that energy-containing scales of stratified turbu-

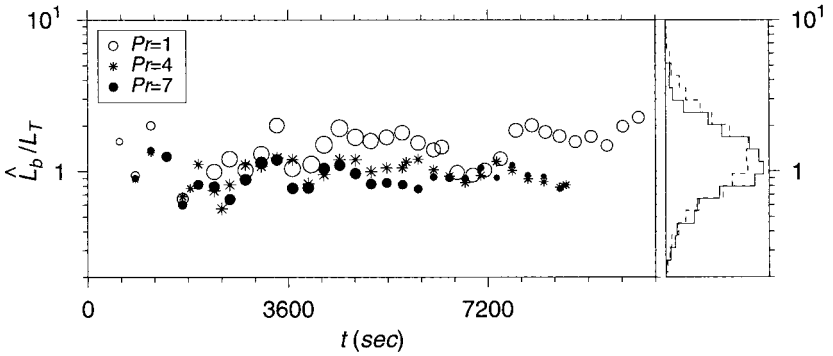

FIG. 5. Evolution of the ratio of buoyancy to Thorpe scales. Symbols and histograms are defined as in Fig. 3.

lence become isotropic at sufficiently high Reynolds number. However, in Gargett's scaling it is the horizontal and vertical velocity perturbations themselves that become equal, and then only in order of magnitude, so that a significant difference between the *cubed* velocity components that appear in L_{En} is not precluded. We emphasize that, when the analyses are done in a consistent manner (i.e., when isotropy is assumed, as in Fig. 6a), there is no discrepancy between the model results and the observations.

d. The Ozmidov scale

The Ozmidov scale

$$\hat{L}_O = \left(\frac{\langle \epsilon_z \rangle}{\hat{N}^3}\right)^{1/2} \tag{24}$$

provides an estimate of the length scale above which motions are strongly affected by buoyancy. The ratio $L_{En}/\hat{L}_O$ varies over several orders of magnitude in the course of each simulation, beginning at values in excess of 100, decreasing to order unity during the transition to turbulence, then decreasing further as the turbulence decays (Fig. 7a). This range of values is consistent with the observations. Much variability is removed when the isotropic approximations for q and ϵ are replaced by the exact values (Fig. 7b). The trend remains monotonically decreasing, but that decrease covers only 1–2 orders of magnitude.

e. $\hat{L}_O/L_T$ as an indicator of event age

We complete this section by testing the first of the two assumptions needed to extract the time dependence of the flux coefficient from the observations (cf. section 1). The ratio of the Ozmidov scale to the Thorpe scale has been suggested by several authors as a convenient indicator of the age of an observed turbulent event (e.g., Dillon 1982; Wijesekera and Dillon 1997). This information is crucial to the interpretation of microstructure data in light of the extreme intermittency of ocean turbulence. It has been suggested (e.g., Gibson 1980) that interior ocean mixing is dominated by rare, powerful events that are almost never observed, events of which the observed turbulent overturns are likely to be mere remnants. In that case, turbulence statistics derived from sparse microstructure observations may severely underestimate the level of mixing in the interior ocean. Other authors (e.g., Caldwell 1983; Gregg 1987) have contended that this problem is not severe, that microstructure observations actually contain a fair sampling of young, energetic events. In Fig. 8a, we demonstrate that $\hat{R}_{OT}$ represents an effective clock for these simulated turbulent events. Young, preturbulent overturns are characterized by $\hat{R}_{OT} < 1/2$. Shortly after transition, the ratio increases to order unity. As turbulence decays, $\hat{R}_{OT}$ increases further to values substantially greater than 1. These results are entirely consistent with the conclusions of Wijesekera and Dillon (1997), which were derived from observational data by using entropy as a time measure. The solid curve on Fig. 8a shows a least squares fitted power law of the form $\hat{R}_{OT} = (t/t_o)^b$, with $t_o = 3298$ s and $b = 1.140$.

For comparison, we also examine the ratio L_O/L_T (Fig. 8b), in which L_O is computed using the mean density gradient across the patch rather than the weighted mean described in section 2c(2). In other words, L_O is computed as

$$L_O = \left(\frac{\langle \epsilon_z \rangle}{N^3}\right)^{1/2}, \tag{25}$$

in which $N^2 = g\langle \theta_z \rangle$. The ratio L_O/L_T tends to grow monotonically, but that tendency is generally less pronounced than is found with $\hat{L}_O/L_T$ (Fig. 8a), and is thus more readily dominated by short-period fluctuations. In quantitative terms, the coefficients appearing in the power law are $t_o = 6833$ s and $b = 0.752$. Thus, the mean growth exponent of $\hat{L}_O/L_T$ over these three simulations is 52% larger than that of L_O/L_T. The difference is partly because the weighted gradient $\hat{\theta}_z$ decreases in response to the growth of localized, secondary overturns (Smyth and Moum 2000b), whereas the mean gradient

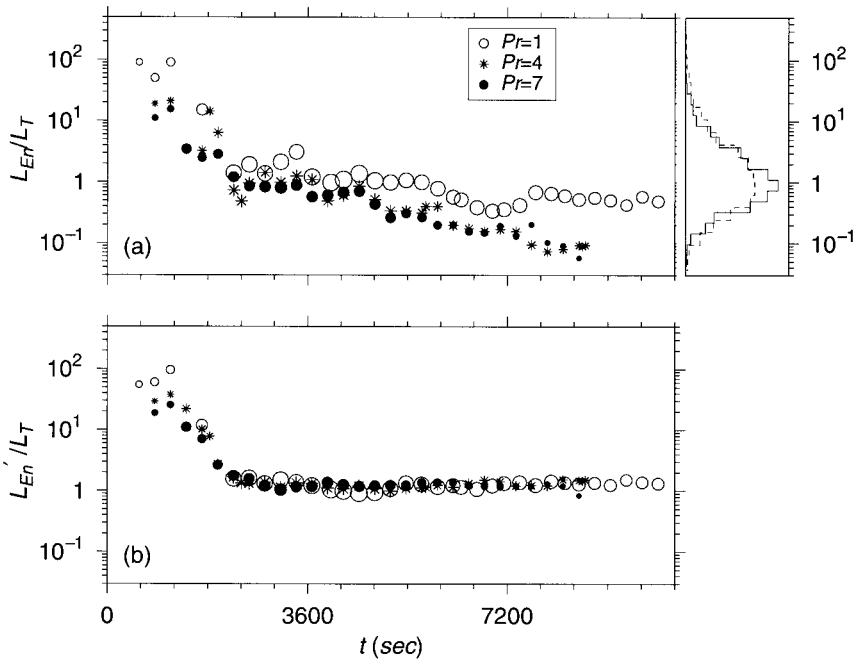

FIG. 6. Evolution of the ratio of the energy-containing scale to the Thorpe scale. (a) The energy-containing scale is approximated as with the observational data, assuming that the large scales are isotropic, i.e., $L_{En} = w_{\text{rms}}^3/\langle\varepsilon_Z\rangle$. Symbols and histograms are defined as in Fig. 3. (b) The assumption of isotropy is discarded, resulting in $L'_{En} = \langle q^2\rangle^{3/2}/\langle\varepsilon\rangle$. There is no corresponding observational data.

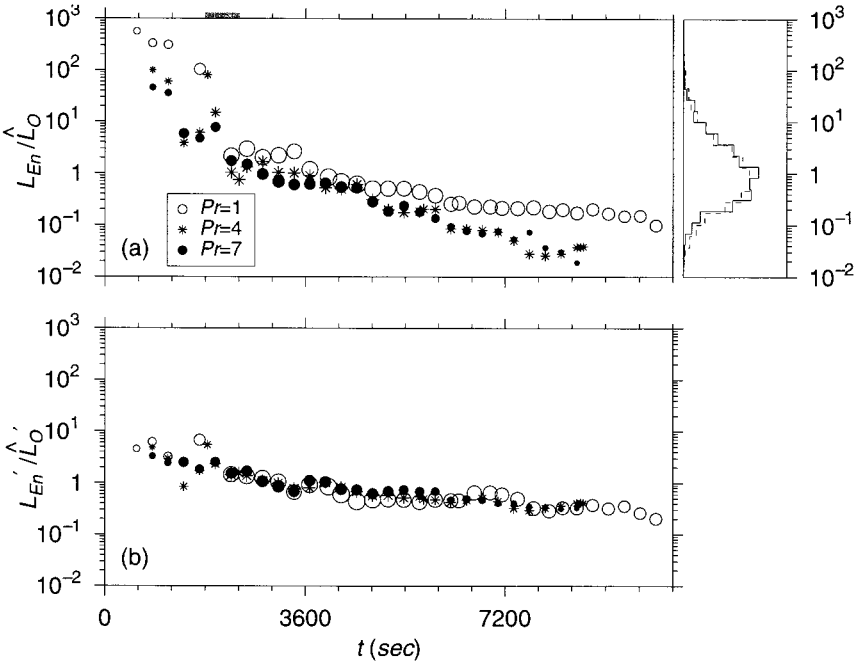

FIG. 7. Evolution of the ratio of the energy-containing scale to the Ozmidov scale. (a) The energy-containing scale is approximated as with the observational data, assuming that all scales are isotropic, i.e., $L_{En} = w_{\text{rms}}^3/\langle\varepsilon_Z\rangle$. The shaded bar at the top of (a) marks the approximate time of the transition to turbulence. Symbols and histograms are defined as in Fig. 3. (b) The assumption of isotropy is discarded, resulting in $L'_{En} = \langle q^2\rangle^{3/2}/\langle\varepsilon\rangle$ and $\hat{L}'_O = (\langle\varepsilon\rangle/\bar{N}^3)^{1/2}$. There is no corresponding observational data.

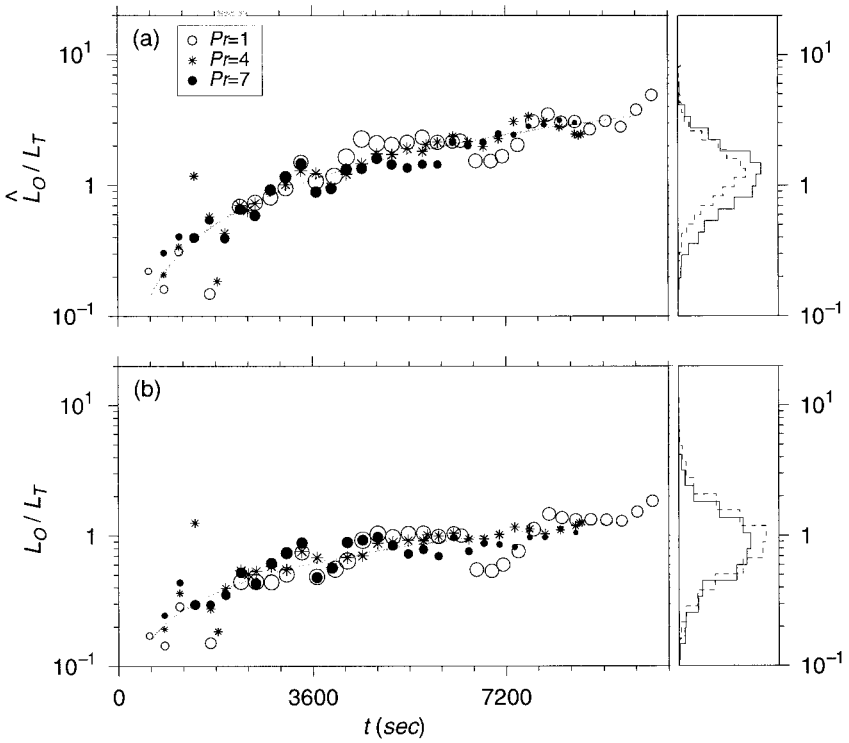

FIG. 8. Evolution of the ratio of the Ozmidov scale to the Thorpe scale. Symbols and histograms are defined as in Fig. 3. The shaded bar at the top of (a) marks the approximate time of the transition to turbulence. (b) The mean scalar gradient is used to calculate the Ozmidov scale, i.e., $L_O = (\langle \varepsilon_z \rangle / N^3)^{1/2}$. The solid curves represent power-law growth with exponents (a) 1.140 and (b) 0.752.

$\langle \theta_z \rangle$ is unaffected. The result of this is that the ratio of Ozmidov to Thorpe scale functions best as an age indicator when the former length scale is computed using the weighted scalar gradient $\hat{\theta}_z$, as in (24). An alternative expression for $\hat{R}_{OT}$ is

$$\hat{R}_{OT} = \frac{L_O}{L_E^{3/4} L_T^{1/4}}. \tag{26}$$

In summary, we have seen that while Reynolds numbers achieved by the model are at the small end of the observed range, most other turbulence statistics tested are consistent with observations. These include measures of the large-scale geometry, as well as statistics relating background stratification, kinetic energy, and dissipation rates. We conclude that our simulated overturns provide an acceptable model for turbulent patches in the thermocline.

Three quantities, L_p/L_T, $L_{En}/\hat{L}_O$, and $\hat{R}_{OT}$, look promising as age indicators. Here L_p/L_T can be significantly nonmonotonic (Fig. 4a) and is also sensitive to the choice of patch boundaries. The evolution of $L_{En}/\hat{L}_O$ is largely a consequence of the changing anisotropy of the energy containing scales, and may thus depend on the details of the large-scale structure (Fig. 7). In contrast, the monotonic increase of $\hat{R}_{OT}$ (Fig. 8a) is an expression of increasing geometric complexity, a fundamental property of turbulence. We therefore expect that $\hat{R}_{OT}$ will be generally applicable as a time indicator. The fact that the observations include the range of values of these ratios found in our preturbulent patches suggests that ''young'' patches are present (albeit rare) in the observations. Further evidence will be given in the following section.

4. Mixing efficiency

Using the DNS results, we have demonstrated a robust, monotonic time dependence in the ratio $\hat{R}_{OT} = \hat{L}_O / L_T$. This suggests that it may be possible to use $\hat{R}_{OT}$ as a clock to indicate the phase of the turbulence in the absence of a real time base, as in typical oceanic observations. This is significant in determining a mixing efficiency for the turbulence, which is used to estimate eddy diffusivity and which has been purported to vary with the stage of decay of the turbulence. To assess the significance of time variability of mixing efficiency, we first examine its time evolution in our DNS results. The calculation is done both exactly and using the approximation employed in the analysis of observational data. We then turn to the observations to interpret the variations of mixing efficiency. In each case, we find that mixing efficiency evolves distinctly with turbulence

age; young, preturbulent billows mix very efficiently and can therefore make a significant contribution to the total potential energy gain.

a. Assessment of mixing efficiency in DNS and observational data

We begin with a discussion of the subtleties encountered in defining the ''efficiency'' of mixing by stratified turbulence. Stratified turbulence is a highly complex phenomenon, and is usually thought of as a combination of two much simpler phenomena: gravity waves, and ''turbulence'' as it occurs in nonstratified flow. There is no way to rigorously disentangle the wavelike and turbulence-like aspects of stratified turbulence, but since gravity waves are essentially nondissipative, we conceptually associate mixing with the ''turbulent'' aspect of the flow.

The work associated with mixing is best understood in the context of transfers of kinetic energy to and from the fluctuating velocity field. In the present model, perturbation kinetic energy evolves in accordance with

$$\frac{d}{dt}\left\langle \frac{1}{2}\overline{u_i' u_i'} \right\rangle_D = S - B - \langle \overline{\varepsilon} \rangle_D, \tag{27}$$

in which $S = -\langle \overline{u'_i w'}\, \overline{u}_{i,z} \rangle_D$ is the source (or production) term, $B = -g\langle \overline{w'\theta'} \rangle_D$ is the rate of potential energy increase, and the final term represents the loss of perturbation kinetic energy to viscous dissipation. Here, the subscript D on the angle brackets indicates that the vertical average is taken over the entire domain (not just the patch). Overbars indicate the horizontal mean, primes indicate fluctuations about that mean, and repeated indices are summed over.

The efficiency of a mechanical system is defined as the ratio of work done to energy input. In the present case, that would be B/S, commonly called the flux Richardson number (denoted R_f). The difficulty with applying this concept to stratified turbulence is that both the production and buoyancy terms are reversible, that is, they represent two-way transfer of energy. The ratio B/S is therefore often negative. This complication is symptomatic of the presence of gravity waves, which mediate the transfer of kinetic energy from the mean flow to the turbulence. Half of this problem was solved by Winters et al. (1995), who showed how the buoyancy term can be separated into reversible and irreversible parts, B_r and B_i. Unfortunately, that procedure does not work on the production term.

Further progress may be made by shifting attention from production to dissipation. The reversibility of the production term is linked to the fact that production occurs at large scales, and is therefore dominated by wavelike motions. In contrast, viscous dissipation is driven by small-scale motions that have much less tendency to be wavelike, and the dissipation term is sign-definite by definition. One may therefore compute the positive definite quantity

$$\Gamma = B_i/\langle \overline{\varepsilon} \rangle_D, \tag{28}$$

which represents the ratio of irreversible potential energy increase to energy loss via friction. This quantity provides a useful definition of the ''efficiency'' of mixing, and it has been used as such by researchers in stratified turbulence (e.g., McEwan 1983). Unfortunately, it conflicts with the historical definition of efficiency, as pointed out by Moum (1996b). In that paper, Γ was referred to instead as the ''flux coefficient,'' a designation that is also appropriate here since Γ is the factor by which the dissipation rate is multiplied to obtain the irreversible buoyancy flux (which is also the rate of irreversible potential energy gain). Therefore, the flux coefficient will be our vehicle for understanding the efficiency with which an evolving stratified turbulence event raises the potential energy of the fluid.

We now describe two definitions of the flux coefficient, each of which is an approximation to (28). The first is called the ''instantaneous'' flux coefficient Γ_i and is computable only from the three-dimensional information available in simulations. The second quantity is Γ_d, an approximation to Γ_i that can be assessed using profiler measurements.

The computation of the irreversible potential energy increase begins with the identification of the minimum potential energy state of a given density distribution, that is, the spatial configuration of the density distribution that is in static equilibrium (Winters et al. 1995). This equilibrium state is easily computed from model data by sorting the density field in a manner that is the three-dimensional analog of the Thorpe reordering described in section 2c(1). In this state, density varies only in the vertical, and its vertical gradient is negative semidefinite. The associated (volume-averaged) potential energy is defined, up to an additive constant, by

$$P_b = -\frac{g}{L_z}\int_0^{L_z} z\theta_*(z)\, dz, \tag{29}$$

in which θ^* is the profile of scaled temperature defining the equilibrium state. The total potential energy is

$$P_t = -\frac{g}{V}\int z\theta(x, y, z)\, dV, \tag{30}$$

where the integral is over the computational volume. The available potential energy is just the difference: $P_a = P_t - P_b$.

A fluid volume obeying the boundary conditions (5)–(8), at rest in the equilibrium state, gains potential energy by molecular diffusion at the (volume-averaged) rate

$$\Phi(t) = \frac{g\kappa}{L_z}\overline{\theta}\,\Big|_{\text{bottom}}^{\text{top}} \tag{31}$$

in which $\overline{\theta}\,\big|_{\text{bottom}}^{\text{top}}$ denotes the difference in horizontally

averaged θ between the upper and lower boundaries (Winters et al. 1995). The quantity we will compute to represent the instantaneous flux coefficient is

$$\Gamma_i = \frac{dP_b/dt - \Phi}{\langle \overline{\epsilon} \rangle_D}. \tag{32}$$

The subtraction of Φ isolates the irreversible work associated with fluid motion from that which would occur in the resting state, a distinction that can be significant in weak turbulence.

Computation of Γ_i requires more information than is available in microstructure observations. We now consider an alternative approach first used by Osborn and Cox (1972). The model equations and boundary conditions given in section 2a imply the following budget for the volume-averaged temperature variance:

$$\frac{d}{dt}\left\langle \frac{1}{2}\overline{\theta'^2} \right\rangle_D = -\langle \overline{w'\theta'}\,\overline{\theta}_{,z} \rangle_D - \frac{1}{2}\langle \overline{\chi} \rangle_D. \tag{33}$$

This may be rearranged to give an expression for the net scalar flux

$$\langle \overline{w'\theta'} \rangle_D = -\frac{1}{2\widetilde{\theta}_{,z}}\left(\langle \overline{\chi} \rangle_D + \frac{d}{dt}\langle \overline{\theta'^2} \rangle_D \right). \tag{34}$$

The quantity $\tilde{\theta}_{,z}$ is a measure of the bulk scalar gradient, given explicitly by

$$\widetilde{\theta}_{,z} = \frac{\langle \overline{w'\theta'}\,\overline{\theta}_{,z} \rangle_D}{\langle \overline{w'\theta'} \rangle_D}. \tag{35}$$

Unfortunately, $\tilde{\theta}_{,z}$ cannot be assessed reliably from observations as it requires prior knowledge of the buoyancy flux. We must therefore approximate $\tilde{\theta}_{,z}$ using either the mean gradient $\langle \overline{\theta}_{,z} \rangle_D$ or the alternative bulk gradient $\widehat{\theta}_{,z}$ In the DNS data, $\tilde{\theta}_{,z}$ exceeds the mean gradient by as much as a factor of 3 in preturbulent overturns, but becomes more nearly equal to the mean gradient as turbulence develops. The behavior of $\widehat{\theta}_{,z}$ is the opposite: it approximates the mean gradient early on, but is smaller than the mean gradient late in the flow evolution as turbulence becomes layered (Fig. 2). Thus, the mean gradient provides a better estimate of $\tilde{\theta}_{,z}$ than does $\widehat{\theta}_{,z}$ Accordingly, we follow the usual practice of using the mean gradient in the Osborn–Cox approximation.

In addition, stationarity is usually assumed in the analysis of observational data, so that the time derivative in (34) may be neglected. Dividing (34) by the volume-averaged dissipation rate and multiplying by g then gives:

$$\frac{g\langle \overline{w'\theta'} \rangle_D}{\langle \overline{\epsilon} \rangle_D} \equiv \Gamma_d = \frac{g\langle \overline{\chi} \rangle_D}{2\langle \overline{\theta}_{,z} \rangle_D \langle \overline{\epsilon} \rangle_D}. \tag{36}$$

Of necessity, the volume averages $\langle \overline{(\,)} \rangle_D$ are replaced by vertical averages between the upper and lower patch boundaries and isotropic approximations are employed for the dissipation rates:

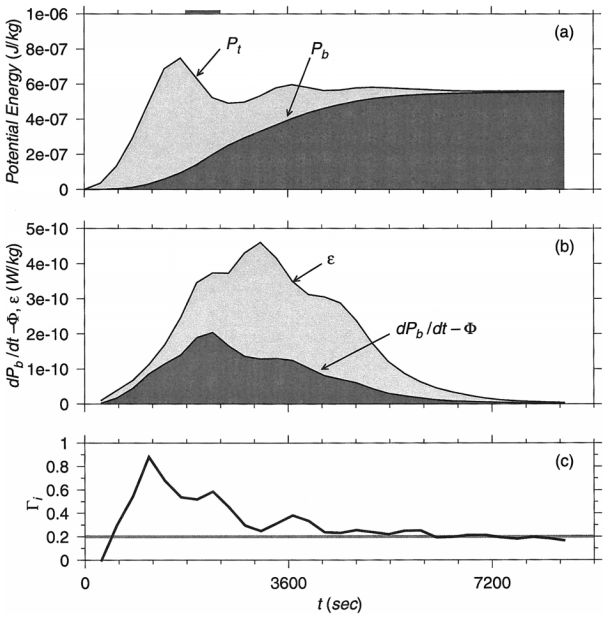

FIG. 9. Evolution of various quantities describing work done against gravity for the Pr = 7 simulation. (a) The upper curve denotes the total potential energy as given in (30), the lower curve denotes the background potential energy (29), and the distance between the curves represents the available potential energy. The shaded bar at the top of (a) marks the approximate time of the transition to turbulence. (b) Volume-averaged perturbation kinetic energy dissipation rate (upper curve) and rate of irreversible potential energy increase (lower curve). The ratio of the lower to the upper abcissa is Γ_i. (c) Γ_i.

$$\Gamma_d = \frac{g\langle \chi_z \rangle}{2\langle \theta_{,z} \rangle \langle \epsilon_z \rangle}. \tag{37}$$

b. Flux coefficient evolution in simulated flows

In Fig. 9 we show various aspects of potential energy evolution for the case Pr = 7. Early in the simulation, most of the work done against gravity is reversible, that is, it contributes to the available potential energy (Fig. 9a). The latter grows to a maximum near $t = 1600$ s, then decreases rapidly (i.e., the billow breaks). The background potential energy grows monotonically and, ultimately, becomes equal to the total potential energy as vertical motions decay.

In Fig. 9b, we see that the rate of irreversible work done against gravity peaks quite early in the simulation. In fact, it peaks as the billow is breaking. Note that about one-third of the net potential energy gain in this simulation occurred *before* the transition to turbulence! The dissipation of kinetic energy, in contrast, peaks near $t = 3000$ s, after the transition to turbulence is complete. This offset between maximum irreversible work and maximum kinetic energy dissipation leads to large values of Γ_i early in the simulation (Fig. 9c; Winters et al. 1995; Caulfield and Peltier 2000; Staquet 2000). This highly efficient mixing occurs primarily in the braids

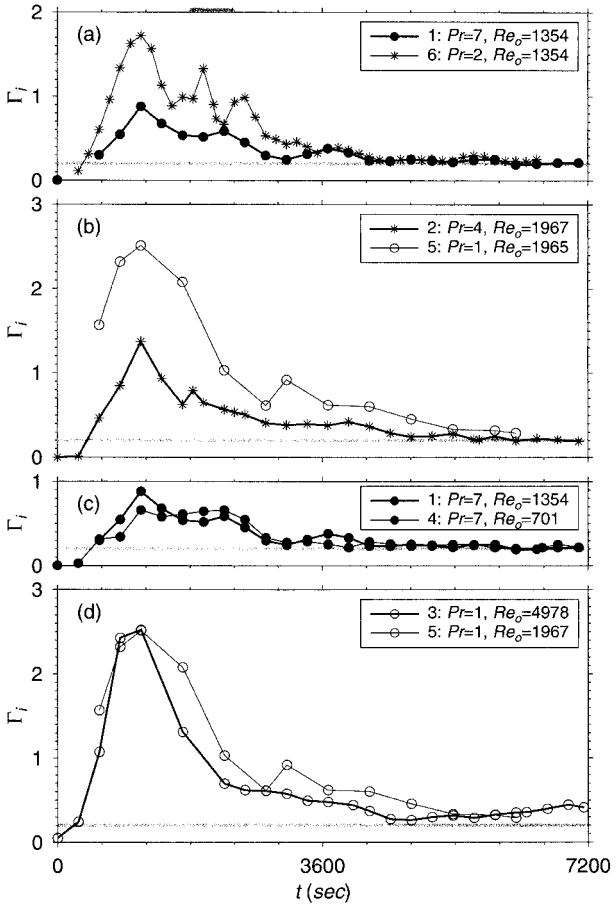

FIG. 10. Time evolution of Γ_i for six simulations. Thin lines indicate simulations 4, 5, and 6. The shaded bar at the top of the uppermost frame marks the approximate time of the transition to turbulence.

connecting adjacent KH billows. After the onset of turbulence, Γ_i decreases and appears to asymptote to the canonical value 0.2 (Caulfield and Peltier 2000; Smyth and Moum 2000b; Staquet 2000).

The tendency for Γ_i to peak during the preturbulent phase of billow evolution then asymptote to a lower value is common to all of our simulations. However, the quantitative evolution differs significantly between the various cases investigated due to differences in both Prandtl number and initial Reynolds number. At any given time, Γ_i tends to be larger in the low-Pr cases. This could be a result of increased molecular diffusivity, or it could be due to the higher Reynolds number reached in those simulations.

In order to properly assess the relevance of these results for ocean turbulence, it is important that we understand the Reynolds and Prandtl number dependence of Γ_i more fully. We investigate further using result from three additional simulations in order to compare cases in which one parameter was fixed while the other was varied (Fig. 10). In Fig. 10a, we compare two runs having the same initial Reynolds number but different Prandtl numbers. In the preturbulent and transitional phases, the high-Pr case displays considerably smaller values of Γ_i. As turbulence decays, both simulations asymptote to $\Gamma_i \approx 0.2$. In Fig. 10b, we show a similar comparison. In this case, the (common) Reynolds number is increased and the Prandtl numbers are correspondingly lower: Pr = 4 and Pr = 1. Again, the higher Pr case displays smaller flux coefficients over most of the simulations.

We next compare two runs in which Pr was set to 7 while Re_o was varied (Fig. 10c). There is no significant difference in the evolution of Γ_i. Finally, we compare two cases in which Pr was set to unity and Re_o differed (Fig. 10d). Here, there is a hint that the case with lower Reynolds number exhibited higher Γ_i, but the difference is much less pronounced than was observed as a result of varying Pr. Note that both cases with Pr = 1 appeared to asymptote to states in which Γ_i was significantly higher than 0.2.

We conclude that any dependence of Γ_i on initial Reynolds number is too subtle to be captured in the present dataset. In contrast, the dependence on Prandtl number is clear. In the preturbulent phase and during transition, Γ_i varies inversely with Pr. In the decay phase, Prandtl number dependence is not evident until Pr is reduced to unity, at which point the asymptotic value of Γ_i is closer to 0.3 or 0.4 than to 0.2. For Pr $>$ 1, the late stage value of Γ_i is indistinguishable from 0.2 in these data.

In Fig. 11, we show the net flux coefficient Γ_{net}, calculated as the ratio of the net work done against gravity to the net kinetic energy lost to dissipation. The numerator and denominator in (32) are integrated separately over the duration of the simulation, then divided. For the Pr = 7 case, an additional run with smaller Reynolds number (small bullet) yielded a slightly reduced value of Γ_{net}. For the Pr = 1 case, an additional run with smaller Reynolds number (small circle) yielded a slightly increased value of Γ_{net}.

As in case of Γ_i (Fig. 10), the dependence of Γ_{net} on Prandtl number is quite robust, whereas the dependence on Reynolds number is weak and inconsistent. These results suggest that Reynolds number is not a dominant factor determining the flux coefficient. We conclude that the difference in flux coefficient evolution among the various cases considered here is due primarily to the difference in Prandtl number.

To understand the extraordinary efficiency of mixing in preturbulent overturns, we must consider the underlying physics. As noted above, turbulence does not mix scalars directly; rather, it accelerates molecular fluxes by amplifying scalar gradients through compressive strain in the velocity field (Batchelor 1959). The association of strong mixing with turbulence is based on the fact that strain tends to be much stronger in turbulent flow than in laminar flow.

However, gradient amplification depends not only on the magnitude of the strain, but also on the orientation between the strain and the scalar gradient. Amplification

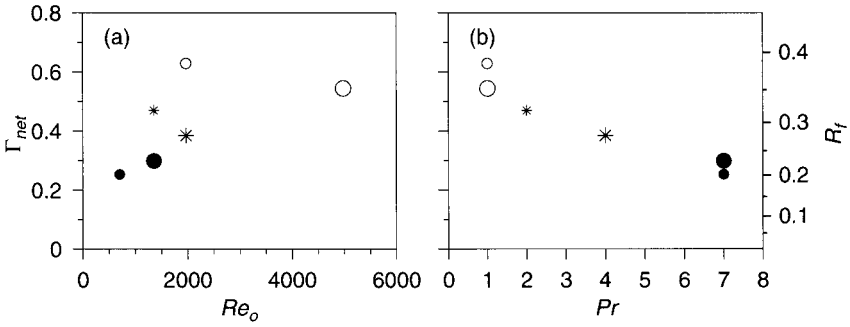

FIG. 11. Net values of the flux coefficient and R_f for seven DNS runs. Each data point represents one complete simulation. The large circle, asterisk, and bullet denote the Pr = 1, Pr = 4, and Pr = 7 cases shown in previous figures (cases 3, 2, and 1, respectively). Smaller symbols represent cases 4 (bullet), 5 (circle), and 6 (asterisk), as described in Table 1.

is most effective when the gradient is aligned parallel to the compressive principal strain. If the strain field is steady, scalar gradients will tend toward this alignment. Early theories of scalar mixing (e.g., Batchelor 1959) assumed that the turbulent strain field evolves slowly enough that scalar gradients would remain close to this ideal state. We have learned, however, that this assumption is incorrect: turbulent strain fields evolve far too rapidly for the scalar gradient to maintain anything close to ideal alignment (Smyth 1999). As a result, turbulent mixing is much less efficient than one would expect based on the magnitude of the strain alone. Conversely, the weak but steady strain field in preturbulent billows can be a surprisingly effective mixing mechanism. The state of maximum Γ_i is shown in Fig. 1b. The reader may verify visually that the sharpest scalar gradients occur at this time (cf. Figs. 1a, 1c, 1d).

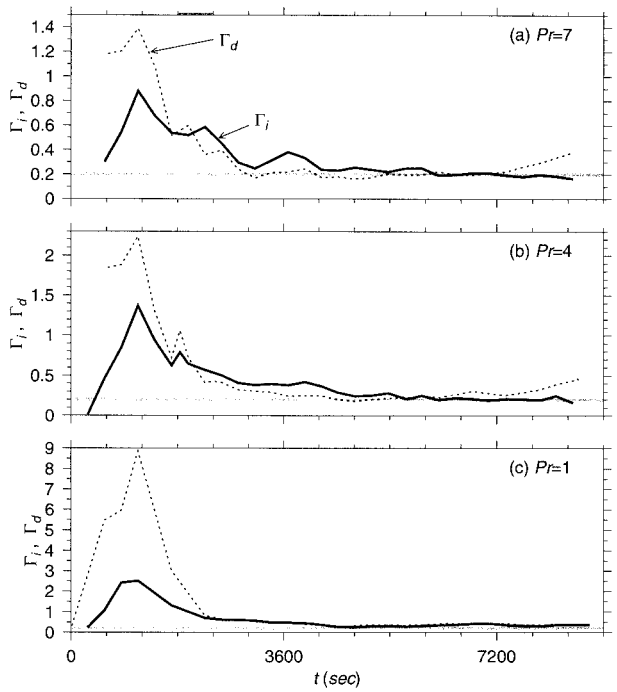

FIG. 12. Comparison of the instantaneous flux coefficient Γ_i and its approximation Γ_d for simulations (a) 1, (b) 2, and (c) 3. Solid curves: Γ_i. Dashed curves: Γ_d, computed as in (36). Horizontal lines indicate the standard value 0.2. The shaded bar at the top of (a) marks the approximate time of the transition to turbulence.

c. Γ_d as an approximation to the flux coefficient

In preparation for evaluating flux coefficient evolution in the observational data, we now test the second of the two assumptions described in section 1. This requires assessing the usefulness of Γ_d as an approximation for Γ_i. Here Γ_d and Γ_i follow similar evolutionary patterns, the main differences being $\Gamma_d > \Gamma_i$ during the preturbulent phase and the tendency of Γ_d to increase during the final stages of turbulence decay (Fig. 12). The large discrepancy found in the preturbulent phase is not surprising. The stationarity assumption inherent in the Osborn–Cox formulation is invalid during this phase. Also, examination of model results has shown that the ''correct'' scalar gradient for the Osborn–Cox formula, $\tilde{\theta}_z$ (35), exceeds the mean gradient by a factor 2–3 during this phase. Despite this, we use the mean gradient for consistency with the observational results. (The erroneous factor decreases to 1.0–1.5 after the transition to turbulence.) The increase observed late in the simulation is related to the inaccuracy of the isotropic approximations in this very weak turbulence. Given these discrepancies, it is reassuring to find that Γ_d is actually a useful predictor of Γ_i in all simulations. The correspondence is reasonable while turbulence is strong, and even prior to transition Γ_d gives a qualitative indication of the large flux coefficient that characterizes that early phase.

The simulated values of Γ_d are within the range of observed values, though they tend to lie toward the low end of the range (Fig. 13). Instances of $\Gamma_d > 1$ are common in the observations, especially those from the Equatorial Undercurrent, but occur in the simulations only when Pr < 7. In Fig. 13b, we assess the accuracy of Γ_d as a predictor of the true net flux coefficient, Γ_{net}.

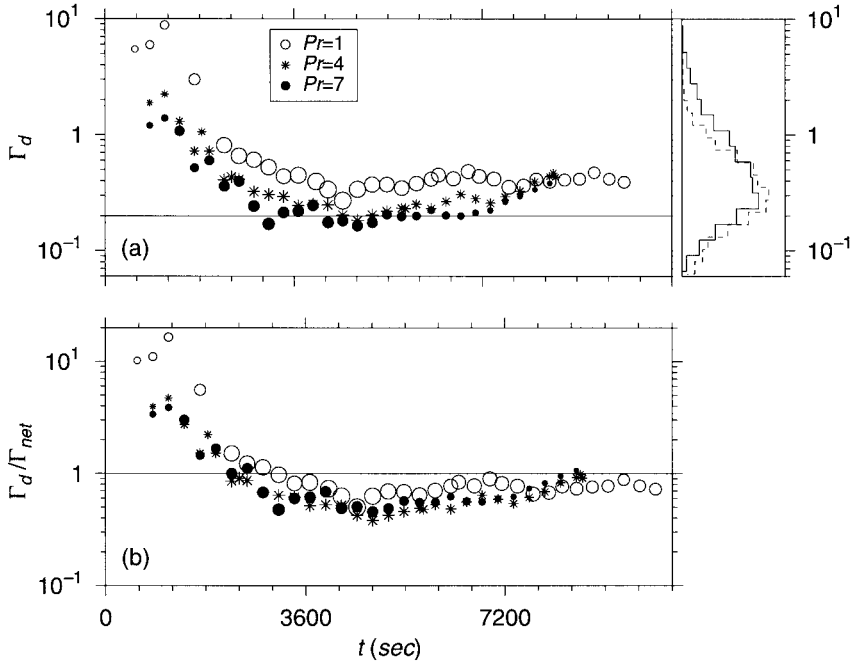

FIG. 13. (a) Evolution of the Osborn–Cox approximation to the flux coefficient Γ_d for three DNS runs. Symbols and histograms are as defined in Fig. 3. The horizontal line indicates the standard value 0.2. (b) Evolution of Γ_d normalized by the true net flux coefficient (cf. Fig. 12).

In all cases, values of Γ_d obtained before the onset of turbulence overpredict Γ_{net}, while values corresponding to fully-developed turbulence tend to be low. Overestimation of Γ_{net} in the preturbulent state is due both to the intrinsically large mixing efficiency of laminar billows and to the limitations of Γ_d as an approximation to the true flux coefficient. The tendency to underestimate Γ_{net} later in the evolution reflects the low mixing efficiency of the fully turbulent flow.

d. "Evolution" of Γ_d in observations

We turn now to an assessment of the evolution of the flux coefficient derived from observational data. Two central assumptions needed for this assessment have now been tested using the DNS data. We have shown that the length scale ratio $\hat{R}_{OT}$ provides an effective indicator of the "age" of a turbulent event (section 3e). We have also seen that Γ_d provides, within limits, a useful estimate of the flux coefficient that can be computed on the basis of observational data (section 4c). Although we cannot follow the time evolution of a particular event as in the DNS analyses, we can examine the relationship between Γ (as approximated by Γ_d) and event age (as expressed by $\hat{R}_{OT}$) in samples drawn from many events. If turbulence age is indeed the dominant factor determining Γ, that fact should be evident in these analyses. Conversely, if the data do not reveal the expected relationship between Γ_d and $\hat{R}_{OT}$, the applicability of the DNS results to ocean turbulence must be questioned.

In Fig. 14, we show the individual microstructure observations from the TIWE and FLX91 datasets. In each case, the relationship between Γ_d and $\hat{R}_{OT}$ is very well-defined and reveals an overall monotonic decrease. Differences in detail between the observational and DNS results are not surprising. Very large values of Γ_d at small $\hat{R}_{OT}$ may be due to the tendency of Γ_d to overestimate Γ in young turbulence (Fig. 12). Unlike the model results, observational values of Γ_d show no tendency to asymptote to 0.2 at long times. We attribute this to inability to resolve the finest-scale density fluctuations in the observational data. Because of this, values of L_T smaller than about 10 cm are rarely observed. Large values of $\hat{R}_{OT}$ are therefore attained only as a result of large Ozmidov scales, which tend to correlate with small Γ_d. In the model results, the mean value of Γ at high $\hat{R}_{OT}$ is also influenced by cases with small $\hat{L}_O$ and even smaller L_T, which tend to have larger mixing efficiency. Despite the differences between the numerical and observational analyses, the main result of this section, namely that newly created billows mix more efficiently than fully developed turbulence, is revealed by both calculations.

e. Implications for observational estimates of the turbulent diffusivity

Both numerical and observational data support the conclusion that the flux coefficient decreases dramatically as a billow grows and breaks down into turbulence. This has significant implications for observational es-

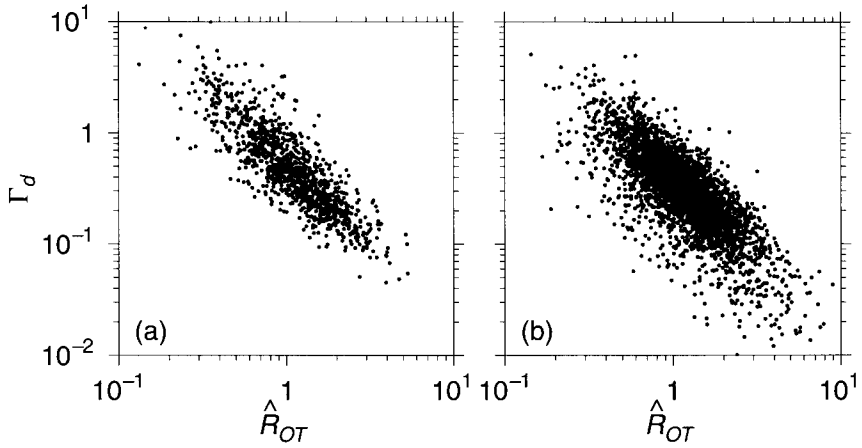

FIG. 14. Here Γ_d as a function of the length scale ratio $\hat{R}_{OT} = L_O/L_T$, for turbulent patches observed in (a) TIWE and (b) FLX91. The ordinate is interpreted as a surrogate for time.

timates of turbulent heat fluxes and diffusivities. Observational estimates of turbulent mass diffusivity are often made using

$$K_\rho = \Gamma\varepsilon/N^2, \tag{38}$$

which is derived by writing $B_i = K_\rho N^2$ in (28). The usual assumption is that $\Gamma = 0.2$ (e.g., Smyth et al. 1996). This assumption is based primarily on laboratory results (e.g., Ivey and Imberger 1991) and observed mean values (e.g., Moum 1996b). What has been neglected to date is the possible correspondence with the state of the turbulence. The time dependence of the flux coefficient in the present results suggests that diffusivity estimates based on $\Gamma = 0.2$ may be significantly in error. The same applies to estimates of vertical heat and mass fluxes made using the same method (Gregg 1987). The assumption $\Gamma = \Gamma_d$ represents a viable alternative, but we have seen that this estimate is of limited validity, particularly for young events.

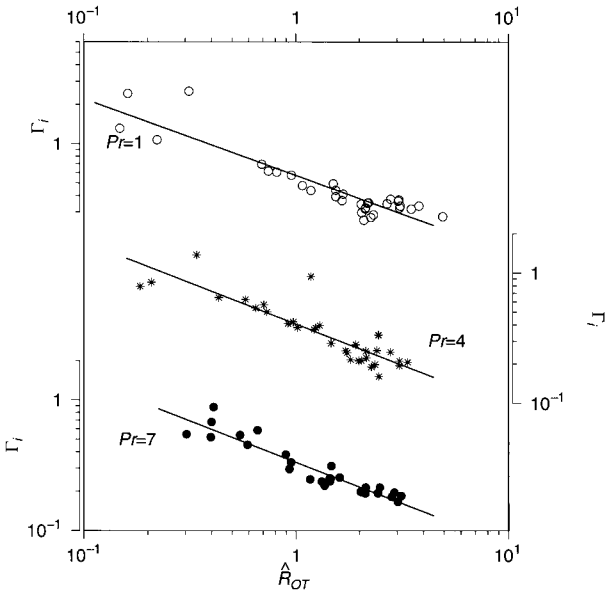

FIG. 15. Instantaneous flux coefficient Γ_i (based on three-dimensional reordering) displayed as a function of $\hat{R}_{OT}$ for simulations 1 (bullets), 2 (asterisks), and 3 (open circles). The horizontal axis is reversed to facilitate interpretation as a time indicator. Results from different simulations are displaced vertically for clarity. Also shown are straight lines illustrating least squares fitted power-law relationships of the form $\Gamma_i = a\hat{R}_{OT}^b$. Values of the coefficients a and b are given in Table 2.

As a step toward improved flux estimates, we look now at the relationship between Γ_i and $\hat{R}_{OT}$ from the numerical data (Fig. 15). Since $\hat{R}_{OT}$ is measurable, this relation suggests a way to estimate the appropriate value of Γ_i for a given observed turbulent event. The result may then be substituted for Γ in (38) to obtain the turbulent diffusivity. In all three simulations, the results are well approximated by a power-law relation of the form

$$\Gamma_i = a\hat{R}_{OT}^b. \tag{39}$$

Least squares fitted values for the constants a and b given in Table 2. The exponent b varies remarkably little between the three cases, whereas the multiplier a varies systematically, with larger values corresponding to smaller Pr. We concluded in section 4b that Prandtl number is the main determinant of the flux coefficient, and therefore that simulation #1 best approximates ocean turbulence in this respect. Accordingly, we suggest that instantaneous flux coefficients be estimated

TABLE 2. Coefficients for least squares fitted relations between the instantaneous flux coefficient and the ratio of Ozmidov to Thorpe scale: $\Gamma_i = a\hat{R}_{OT}^b$. The Prandtl number is included for reference.

	Simulation		
	1	2	3
a	0.3339	0.4052	0.5725
b	−0.6293	−0.6292	−0.5859
Pr	7	4	1

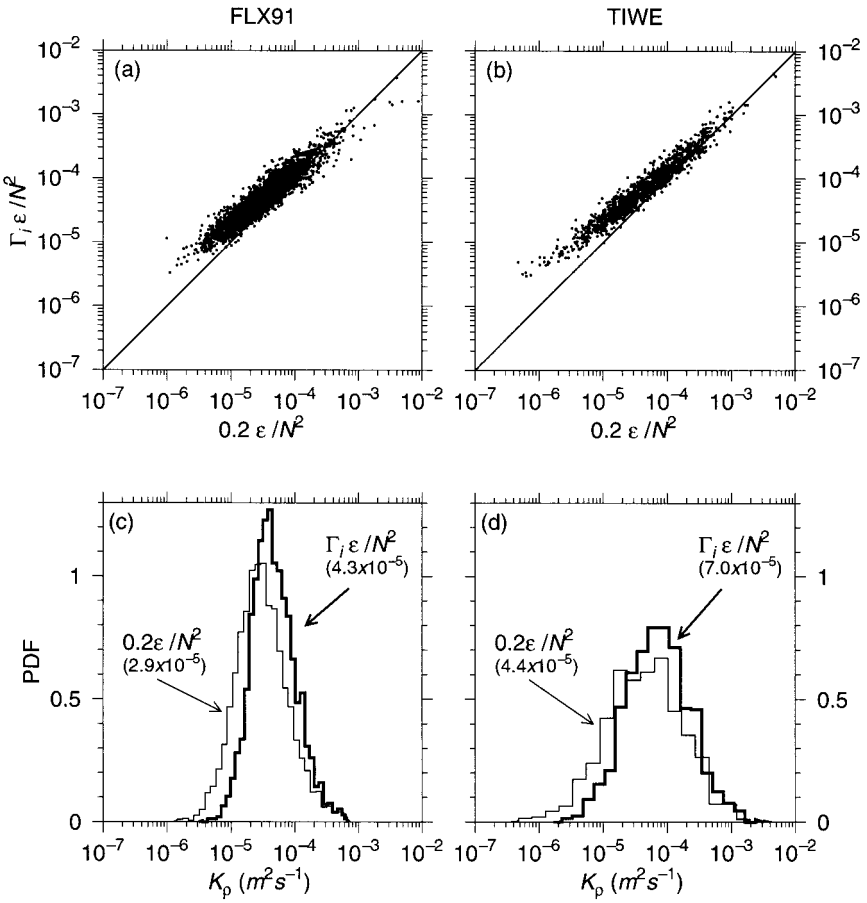

FIG. 16. Estimated turbulent diffusivities $\Gamma_i \varepsilon / N^2$ vs $0.2\varepsilon N^2$ for the (a) FLX91 and (b) TIWE datasets. Here Γ_i is approximated using (40). Probability density functions for the diffusivity estimates for (c) FLX91 and (d) TIWE. Bracketed quantities represent the median of each distribution.

using (39), with the parameter values derived from simulation 1, namely,

$$\Gamma_i \approx 0.33 \hat{R}_{OT}^{-0.63}. \tag{40}$$

We now compute turbulent diffusivity estimates from the observational datasets using (38), with $\Gamma = \Gamma_i$ provided by (39), and compare the results with those obtained assuming $\Gamma = 0.2$ and $\Gamma = \Gamma_d$. In Fig. 16, we show the relationship between diffusivities calculated using $K_\rho = 0.2\varepsilon/N^2$ and those calculated using $K_\rho = \Gamma_i \varepsilon/N^2$. In the latter case, Γ_i is estimated from the observational data using (40).

Figures 16a and 16c show the results for the FLX91 observations. When ϵ/N^2 is large, there is no significant difference in the two estimates. For small ϵ/N^2, however, use of the revised flux coefficient increases the diffusivity estimate substantially. The result is a 50% increase in the median of K_ρ. (The geometric mean behaves similarly.) In the TIWE observations (Figs. 16b and 16d), turbulence levels are generally higher due to the strong mean shear found in the equatorial undercurrent. Nevertheless, the results of using (40) to parameterize the flux coefficient are similar to the midlatitude case. The median of K_ρ now increases by 60%.

Figure 17 shows the corresponding comparison with the results obtained when Γ_d is used to estimate the turbulent diffusivity. For the FLX91 data, the two methods give similar results. For the TIWE data, diffusivities obtained using the Osborn–Cox method are larger than those obtained using the parameterization (40). The present results suggest that these estimates are exaggerated due to the tendency of the Osborn–Cox approximation to overestimate the flux coefficient for young overturns. The difference is not evident in the FLX91 data, in which young overturns are somewhat less prevalent (cf. Fig. 8a).

5. Conclusions

We have employed a combination of numerical and observational methods to investigate mixing as it evolves over the lifetime of a turbulent overturn. Our main focus has been the evolution of the instantaneous flux coefficient Γ_i, which is closely related to the effi-

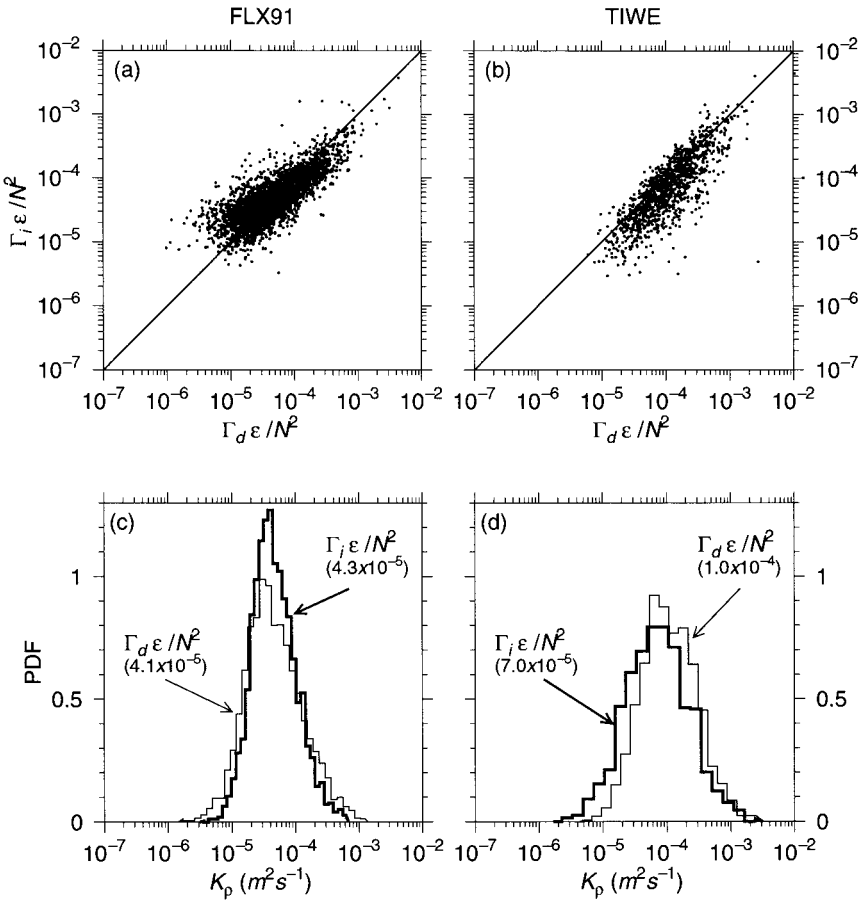

FIG. 17. Scatterplots of the estimated turbulent diffusivities $\Gamma_i \varepsilon/N^2$ vs $\Gamma_d \varepsilon/N^2$ for the (a) FLX91 and (b) TIWE datasets. Here Γ_i is approximated using (40). Probability density functions for the diffusivity estimates for (c) FLX91 and (d) TIWE. Bracketed quantities represent the median of each distribution.

ciency of mixing. Because numerical and observational technologies are each immune to problems that beset the other, the combination reveals a picture that is much more robust and comprehensive than would be possible using either approach alone. The DNS data enable us to test assumptions that are needed for the interpretation of the observations, while the observational data allow access to realistic parameter regimes that present-day computers cannot simulate.

These analyses have clearly illustrated the difficulty of defining an appropriate bulk stratification for realistic stratified turbulence. This complication arises because theories of stratified turbulence invariably assume a uniform background stratification, an idealization that is never realized exactly in nature. The usual practice is to use some average of the background density gradient in the region of interest to represent the bulk stratification. Here, we have described three separate definitions: $\langle \theta_z \rangle$, $\widehat{\theta_z}$, and $\tilde{\theta}_z$, all of which are equivalent in uniform stratification but differ when stratification varies with depth. The appropriate choice of averaging operation depends upon the application. The most common choice is the simplest: the arithmetic mean, $\langle \theta_z \rangle$. In section 4a, we described the alternative gradient $\tilde{\theta}_z$, in which the average is weighted by the local buoyancy flux. Use of this average, were it practical, would improve the accuracy of the Osborn–Cox formulation. Unfortunately, the averaging operation requires knowledge of the buoyancy flux, which is usually what one is trying to estimate. A more useful alternative gradient was described in section 2c(2): an rms average weighted by the local overturning scale. This yields an average gradient that is weighted to favor overturning regions, and an approximate expression for it, $\widehat{\theta_z}$, (13), is easily calculated from profiler data. Use of this bulk gradient in the computation of the Ozmidov scale improves the effectiveness of L_O/L_T as a gauge of turbulence evolution (section 3e). Also, $\widehat{\theta_z}$, is conveniently insensitive to the choice of patch boundaries. In summary, $\widehat{\theta_z}$, was used in the computation of $\hat{L}_O$ and of the other turbulence statistics compared in section 3. The simple arithmetic mean was used in section 4 to calculate the Osborn–Cox approximation to the flux coefficient Γ_d.

Our first task in these analyses has been to test the validity of the KH model for turbulent patches via comparison of various turbulence statistics. The results are consistent with the hypothesis that KH billows represent a valid model for the observed patches, both in the Equatorial Undercurrent and in the midlatitude thermocline. The length scale ratios we have computed and compared with observations are all geometrical expressions of the underlying dynamics of turbulence in a stratified environment. Although Reynolds numbers are limited by computer capacity, the length scale ratios have proven to be entirely consistent with observations. The only exception to this occurs late in the decay phase, when the simulated flows enter a layered state that is not captured in the present observational data, although such states are often observed in sheared, stratified turbulence (Ruddick et al. 1989). These states are excluded from the present observations due to the difficulty of measuring salinity (and thus density) on sufficiently small scales. Statistics related to background shear are missing from the comparisons due to the absence of finescale shear information in the observational data. Both this and the sparseness of the observational data make it difficult to design a test that discriminates rigorously among different types of turbulence. It could be that turbulence from other sources, for example, convective or advective instability, would fare equally well in some of the tests we have applied here.

The most serious discrepancy between model results and observations is in the Reynolds number. DNS values lie near or below the lowest observed values. This is the main limitation of DNS as a tool for studying turbulence. The implicit assumption in the interpretation of these results is that the Reynolds number is high enough to approximate a high Reynolds number limit, so that further increases would make no significant difference to the results. We are confident that this assumption is valid when the results of interest concern only the physics of the dissipation range (e.g., Smyth 1999; Smyth and Moum 2000a). Once the dissipation range is independent of the large scales, further increases in Reynolds number have little effect. In contrast, mixing efficiency is not purely a property of the dissipation range. The flux coefficient of interest to us here relates dissipation to the buoyancy flux, which is driven by large eddies. It would therefore not be surprising if Γ turned out to depend strongly on the Reynolds number. However, present indications are that this is not so. Experiments in which the Reynolds number was varied revealed no significant variation of Γ_i.

In summary, there is nothing here to disprove the hypothesis that the relationships among the four variables Γ_i, Γ_d, $\hat{R}_{OT}$, and time are the same in the observed flows as in the modeled flows, at least up until the final stages of turbulence decay. Acceptance of this hypothesis justifies using the DNS results as a guide to the interpretation of the observational data. Accordingly, we have proceeded as follows:

1) The dependence of Γ_i on time has been assessed for the simulated flows.
2) The usefulness of $\hat{R}_{OT}$ as a surrogate for time and of Γ_d as a surrogate for Γ_i (assumptions 1 and 2 of the Introduction) has been tested, again using the simulated flows.
3) The dependence of Γ_d on $\hat{R}_{OT}$, has been assessed using the observational data, and compared with the results of step 1 above.

We have shown that Γ_i is of order unity or greater for preturbulent overturns, but then decreases dramatically with the onset of turbulence (Figs. 9 and 10), asymptoting to a value not far from 0.2.

We have confirmed that, in a turbulence event driven by shear instability, the ratio $\hat{R}_{OT}$ increases monotonically with time (Fig. 8a; Wijesekera and Dillon 1997). A convenient alternative expression of the time dependence shown in Fig. 8a is given by

$$t_N \equiv \int_{t_o}^{t} \frac{N(t')}{2\pi}\, dt' = 0.2 + 3.2 \log_{10}\hat{R}_{OT} \pm 0.5, \quad (41)$$

which produces an estimate of the time, in buoyancy periods, since the maximum $\hat{R}_b$ was reached (at t_o, shortly after the transition to turbulence). The estimate is also valid when $t_N < 0$, that is, for preturbulent overturns. This result pertains to breaking KH billows with $\mathrm{Pr} = 7$ and moderate Reynolds number, and may require adjustment as a broader range of flows is investigated. The error estimate 0.5 was derived using DNS case #1, which is most closely representative of turbulent patches in the thermocline.

We have also seen that Γ_d provides a reliable estimate of the flux coefficient during much of the flow evolution (Fig. 12). However, this estimate is severely affected by nonstationarity during the initial overturning phase. The observational data reveal that Γ_d is a well-defined, decreasing function of $\hat{R}_{OT}$.

The result of these analyses is a self-consistent picture, supported by both model and observations, in which the flux coefficient decreases dramatically over the lifetime of a turbulent overturn. Our description of this evolution extends the previous results of Winters et al. (1995), Caulfield and Peltier (2000), Smyth and Moum (2000b), and Staquet (2000). The evolution of Γ_i is determined more by the Prandtl number than by the initial Reynolds number in these experiments. Early in the evolution, peak mixing efficiency shows a consistent increase with decreasing Prandtl number. In the decay phase, the asymptotic value of Γ is near 0.2 for $\mathrm{Pr} = 7$, 4, and 2, but is considerably larger (0.3–0.4) for $\mathrm{Pr} = 1$.

We have explained mixing efficiency evolution in terms of the alignment between the scalar gradient and the strain field. In preturbulent flow, the strain field evolves slowly, so that scalar gradients are able to maintain the optimal orientation for compression and mixing. After transition, the strain field is greatly amplified, but

it also fluctuates rapidly, so that the relative orientation of the scalar gradient is generally far from optimal (Smyth 1999). As a result, the mixing efficiency of fully developed turbulence is smaller than would be expected based on the magnitude of the strain alone.

The evolution of the flux coefficient in the present experiments may be summarized by the parameterization $\Gamma_i = 0.33\hat{R}_{OT}^{-0.63}$. Using this quantity as an alternative to the traditional $\Gamma = 0.2$ in the definition $K_\rho = \Gamma\varepsilon/N^2$, we have evaluated the turbulent diffusivity K_ρ in both observational datasets. In each case, the result was a 50%–60% increase in the estimated median diffusivity. Further development of this parameterization will require extension of the DNS database to higher Reynolds numbers and more general flow geometries, as well as consideration of the effects of parameters other than $\hat{R}_{OT}$.

The preturbulent phase may be of secondary importance in circumstances where turbulence-driving shear persists over long times, such as in the Equatorial Undercurrent. In other cases, however, forcing is likely to be more intermittent, with the result that the initial stage of billow development contributes a more significant fraction of the net mixing. Possible examples of this include transient depression of the Richardson number due to interference of gravity waves, the passage of an internal soliton, and the approach of a gravity wave packet to a critical layer. The present results suggest that such events may exhibit mixing efficiency significantly in excess of 0.2.

Acknowledgments. This work has benefited from conversations with Hemantha Wijesekera, Larry Mahrt, Peter Diamessis, and Colm Caulfield. The research was funded by the National Science Foundation under Grants OCE9521359 and OCE9711872. Computations were performed on the Connection Machine CM5 and CM500e facilities at Oregon State University's Environmental Computing Center.

REFERENCES

Balmforth, N., S. L. Smith, and W. Young, 1998: Dynamics of interfaces and layers in a stratified turbulent fluid. *J. Fluid Mech.*, **355,** 329–358.

Batchelor, G. K., 1959: Small-scale variation of convected quantities like temperature in turbulent fluid. *J. Fluid Mech.*, **5,** 113–133.

Broadwell, J., and R. Breidenthal, 1982: A simple model of mixing and chemical reaction in a turbulent shear layer. *J. Fluid Mech.*, **125,** 397–410.

Caldwell, D., 1983: Oceanic turbulence: Big bangs or continuous creation? *J. Geophys. Res.*, **88,** 7543–7550.

Caulfield, C., and W. Peltier, 2000: Anatomy of the mixing transition in homogeneous and stratified free shear layers. *J. Fluid Mech.*, **413,** 1–47.

DeSilva, I., H. Fernando, F. Eaton, and D. Hebert, 1996: Evolution of Kelvin–Helmholtz billows in nature and laboratory. *Earth Planet. Sci. Lett.*, **143,** 217–231.

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

Gargett, A. E., 1988: The scaling of turbulence in the presence of stable stratification. *J. Geophys. Res.*, **93,** 5021–5036.

——, and J. Moum, 1995: Mixing efficiencies in turbulent tidal fronts: Results from direct and indirect measurements of density flux. *J. Phys. Oceanogr.*, **25,** 2583–2608.

——, T. Osborn, and P. Nasmyth, 1984: Local isotropy and the decay of turbulence in a stratified fluid. *J. Fluid Mech.*, **144,** 231–280.

Gibson, C., 1980: Fossil temperature, salinity and vorticity turbulence in the ocean. *Marine Turbulence*, J. Nihoul, Ed., Elsevier, 221–257.

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

Hazel, P., 1972: Numerical studies of the stability of inviscid parallel shear flows. *J. Fluid Mech.*, **51,** 39–62.

Hebert, D., and J. Moum, 1994: Decay of a near-inertial wave. *J. Phys. Oceanogr.*, **24,** 2334–2351.

——, ——, C. Paulson, and D. Caldwell, 1992: Turbulence and internal waves at the equator. Part II: Details of a single event. *J. Phys. Oceanogr.*, **22,** 1346–1356.

Hinze, J., 1975: *Turbulence*. 2d ed. McGraw-Hill, 790 pp.

Ivey, G., and J. Imberger, 1991: On the nature of turbulence in a stratified fluid. Part I: The energetics of mixing. *J. Phys. Oceanogr.*, **21,** 650–658.

Klaassen, G., and W. Peltier, 1991: The influence of stratification on secondary instability in free shear layers. *J. Fluid Mech.*, **227,** 71–106.

Lien, R.-C., D. Caldwell, M. Gregg, and J. Moum, 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.

McEwan, A., 1983: Internal mixing in stratified fluids. *J. Fluid Mech.*, **128,** 59–80.

Moum, J. N., 1990: Profiler measurements of vertical velocity microstructure in the ocean. *J. Atmos. Oceanic Technol.*, **7,** 323–333.

——, 1996a: Efficiency of mixing in the main thermocline. *J. Geophys. Res.*, **101** (C5), 12 057–12 069.

——, 1996b: Energy-containing scales of turbulence in the ocean thermocline. *J. Geophys. Res.*, **101** (C6), 14 095–14 109.

Oakey, N., 1982: Determination of the rate of dissipation of turbulent energy from simultaneous temperature and velocity shear microstructure measurements. *J. Phys. Oceanogr.*, **12,** 256–271.

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

——, and C. S. Cox, 1972: Oceanic fine structure. *Geophys. Fluid Dyn.*, **3,** 321–345.

Peters, H., M. Gregg, and J. O'Toole, 1988: On the parameterization of equatorial turbulence. *J. Geophys. Res.*, **93,** 1199–1218.

Phillips, O., 1972: Turbulence in a strongly stratified fluid: Is it unstable? *Deep-Sea Res.*, **19,** 79–81.

Posmentier, E., 1977: The generation of salinity finestructure by vertical diffusion. *J. Phys. Oceanogr.*, **7,** 298–300.

Ruddick, B., T. McDougall, and J. Turner, 1989: The formation of layers in a uniformly stirred density gradient. *Deep-Sea Res.*, **36,** 579–609.

——, D. Walsh, and N. Oakey, 1997: Variations in apparent mixing efficiency in the North Atlantic Central Water. *J. Phys. Oceanogr.*, **27,** 2589–2605.

Scinocca, J., 1995: The mixing of mass and momentum by Kelvin–Helmholtz billows. *J. Atmos. Sci.*, **52,** 2509–2530.

Smyth, W., 1999: Dissipation range geometry and scalar spectra in sheared, stratified turbulence. *J. Fluid Mech.*, **401,** 209–242.

——, and W. Peltier, 1989: The transition between Kelvin–Helmholtz and Holmboe instability: An investigation of the overreflection hypothesis. *J. Atmos. Sci.*, **46,** 3698–3720.

——, and ——, 1993: Two-dimensional turbulence in homogeneous and stratified shear layers. *Geophys. Astrophys. Fluid Dyn.*, **69,** 1–32.

——, and J. Moum, 2000a: Anisotropy of turbulence in stably stratified mixing layers. *Phys. Fluids*, **12,** 1343–1362.

——, and ——, 2000b: Length scales of turbulence in stably stratified mixing layers. *Phys. Fluids*, **12,** 1327–1342.

——, D. Hebert, and J. Moum, 1996: Local ocean response to a multiphase westerly wind-burst. Part 2: Thermal and freshwater responses. *J. Geophys. Res.*, **101,** 22 495–22 512.

Staquet, C., 2000: Mixing in a stably stratified shear layer: Two- and three-dimensional numerical experiments. *Fluid. Dyn. Res*., **27,** 367–404.

Tennekes, H., and J. Lumley, 1972: *A First Course in Turbulence*. The MIT Press, 300 pp.

Thorpe, S., 1977: Turbulence and mixing in a Scottish loch. *Philos. Trans. Roy. Soc. London,* **A286,** 125–181.

Werne, J., and D. Fritts, 1999: Stratified shear turbulence: Evolution and statistics. *Geophys. Res. Lett.,* **26,** 439–442.

Wijesekera, H., and T. Dillon, 1997: Shannon entropy as an indicator of age for turbulent overturns in the oceanic thermocline. *J. Geophys. Res.*, **102** (C2), 3279–3291.

——, ——, and L. Padman, 1993: Some statistical and dynamical properties of turbulence in the ocean pycnocline. *J. Geophys. Res.*, **98,** 22 665–22 679.

Winters, K., P. Lombard, J. Riley, and E. A. D'Asaro, 1995: Available potential energy and mixing in density-stratified fluids. *J. Fluid. Mech.*, **289,** 115–128.