Journal of Marine Research, 55, 1069–1101, 1997

On the development of thermohaline correlations as a result of nonlinear diffusive parameterizations

by Raffaele Ferrari[1,2] and W. R. Young[2]

ABSTRACT

Some oceanographic mixing parameterizations assume that transports depend nonlinearly on the buoyancy gradient; e.g., diffusivities are proportional to some power of the buoyancy gradient. In this paper we examine the consequences of these nonlinear-diffusion parameterizations by solving an initial value problem in which the $t = 0$ thermohaline fields are prepared as random and uncorrelated distributions of temperature and salinity. Solutions of the nonlinear diffusion equation as a 'rundown' problem show that correlations develop between the temperature and salinity. These correlations are such that the evolving thermohaline gradients tend to be strongly compensating in their joint effect on buoyancy.

1. Introduction

Oceanographers and meteorologists use models in which the transport of some quantity depends nonlinearly on horizontal gradients of the buoyancy. Perhaps the oldest branch of this family descends from Stommel's (1961) "two box" idealization of the pole-to-equator, thermohaline circulation. Stommel posited a transport law in which the exchange of mass between the boxes is proportional to the buoyancy difference between the boxes. For buoyancy itself then, the buoyancy exchange (=buoyancy difference times mass exchange) will be proportional to the square of the buoyancy difference. Passive tracers will be exchanged between the boxes at a rate which is proportional to the product of the buoyancy difference and the tracer difference.

Welander (1971) improved on Stommel's *ad hoc* specification of an exchange law by applying scaling arguments to the thermocline equations. Welander's reasoning suggests that the mass exchange of the large-scale thermohaline overturning is proportional to the one-third power of the buoyancy difference, and the two-thirds power of the vertical diffusivity.

Another class of examples is concerned with the parameterization of heat fluxes produced by baroclinic instability. The recent article by Visbeck *et al.* (1997) reviews this problem and recommends the adoption of an eddy diffusivity which is proportional to the

1. Politecnico di Torino, Torino, Italy.
2. Present address: Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0230, U.S.A.

horizontal buoyancy gradient; in this case, as in Stommel's box model, the flux of buoyancy is proportional to the square of buoyancy gradients. The extensive baroclinic turbulence simulations of Pavan and Held (1996) suggest an even stronger dependence of flux on gradient; viz., buoyancy flux is proportional to the fourth power of the buoyancy gradient.

A third class of examples concerns the transport of heat and salt in mixed layers and other shallow systems with strong vertical mixing. Young (1994) has developed a mixed layer model in which the shear dispersion of heat and salt produces cubically nonlinear diffusive terms in the vertically averaged conservation equations. This mechanism, which is most potent on relatively small scales for which there is ageostrophic down-pressure gradient flow, is buoyancy driven shear dispersion acting on buoyancy itself (Erdogan and Chatwin, 1967, or the review by Young and Jones, 1991). Models such as this have found application in estuarine dynamics (e.g., Godfrey, 1980).

These nonlinear diffusion models have been introduced using idealized box models, or by arduous multiple scale expansions, or as empirical scaling laws. The different scenarios span the full range of space and time scales which are of interest to oceanographers: from the planetary scale of the thermohaline circulation, through the deformation scale dynamics of baroclinic eddies, down to the ageostrophic circulations in shallow stratified systems. One largely pedagogical goal of this article is to find some middle ground between these different approaches and present arguments emphasizing the unity and simplicity of the underlying physical ideas. Indeed, the central idea is the same in all cases: horizontal buoyancy gradients create down-pressure gradient flows that release the potential energy stored in stratification. Because the strength of the velocities depends on the buoyancy gradient, so must the transport of all advected tracers (including buoyancy itself).

A second goal of this article is to understand some implications of these nonlinear transfer models. Given that fluxes depend nonlinearly on the buoyancy gradient, what are the likely observable consequences of this nonlinearity? An example is given by Chen and Young (1995) who use a nonlinear transfer model to explain the widely observed phenomenon of 'density compensation' between temperature and salt gradients in the mixed layer. In frontal zones, oceanic density compensation is observed on scales of 10 to 100 kilometers (see, for example, Roden (1977, 1980, 1986); Niiler (1984); Colin de Verdiere *et al.* (1986); Yuan and Talley (1992)). But, particularly in the mixed layer, there are also observations of the phenomenon on scales of 1 to 10 kilometers (for example, Flament *et al.* (1985); Samelson and Paulson (1988); and Rudnick (1996)). The model of Chen and Young is based on the assumption that random straining by mesoscale turbulence creates property fronts in the mixed layer and that the equilibrium width of this population of fronts is determined by a balance between the strain and a small-scale nonlinear diffusive process. Because of the nonlinearity, there is a systematic relation between the 'strength' of a front (i.e., the buoyancy jump across the front) and the frontal width; stronger fronts tend to be thicker fronts. As a statistical consequence of this correlation

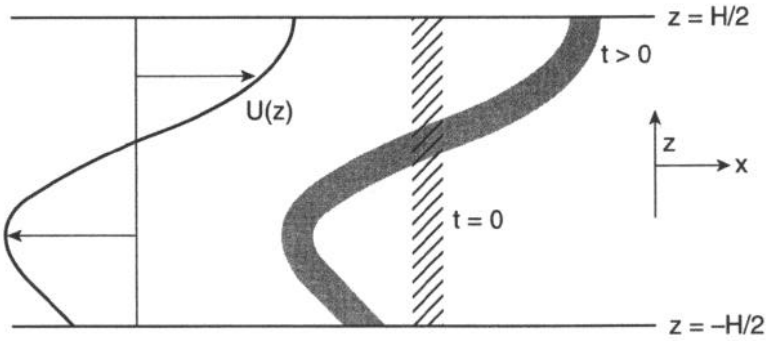

Figure 1. Definition sketch: the tracer is contained in a channel $-H/2 < z < H/2$. The velocity has no net horizontal transport, $\overline{u} = 0$.

between strength and thickness, Chen and Young show that the temperature and salinity gradients are correlated in the sense that there is partial cancellation in their joint effect on buoyancy gradient. In Section 4 we revisit the issue of density compensation and view it from the perspective of an initial value problem.

Pierre Welander was fascinated by the dynamic asymmetry between heat and salt; he took it as a challenge to formulate problems in which this asymmetry is still important, even though the two constituents have the same diffusivity. Welander's insight is that even if T and S differ only in their boundary conditions, then instability of a statically stable state is still possible (Welander, 1989). We like to think that Welander might have found the following contribution interesting as an extreme example of what can happen when T and S are identical in all respects (including boundary conditions). We will see in Section 4 that the kinematic combination $T + S$ (sometimes referred to as 'spice,' e.g., Munk, 1981) still has an interesting evolution which theoretical oceanographers tend to overlook by focusing exclusively on the active combination $T - S$.

2. Shear dispersion: A comparison of different models of vertical mixing

We begin our discussion by collecting some basic results on shear dispersion. Consider the dispersion of some tracer (concentration $c(x, z, t)$) in a shallow layer. One can think of this layer as the mixed layer of the ocean, or an estuary, or some other configuration with high aspect ratio; i.e., with depth H much less than the horizontal scale L. The main point here is that there are two very different timescales. The fast timescale is mixing over the depth H and the slow time scale is that associated with horizontal transports.

With the notation of Figure 1, the mathematical model is

$$c_t + uc_x = \text{mix}, \tag{2.1}$$

where 'mix' indicates some model of the vertical mixing. We compare three different models of mixing and show that they all lead to the same conclusion. The three mixing

models are

$$\begin{aligned} \text{mix}_1 &= Dc_{zz}, \\ \text{mix}_2 &= \tau^{-1}(\overline{c} - c), \\ \text{mix}_3 &= \text{vertical homogenization at intervals of } \tau. \end{aligned} \tag{2.2a,b,c}$$

The model in (2.2a) is the familiar diffusion model, and in this case our problem is identical to Taylor's shear dispersion (1953). The second model in (2.2b) is 'relaxation to vertical average': the vertical average of the tracer over the layer is denoted by

$$\overline{c}(x, t) \equiv \frac{1}{H} \int_{-H/2}^{H/2} c(x, z, t)\, dz. \tag{2.3}$$

Then, with (2.2b), the tracer concentration is continuously nudged toward the local vertical average in (2.3) (e.g., Young, 1994). If the layer in Figure 1 is being mixed by eddies whose scale is comparable to H then model (2.2b) might be more realistic than the diffusion model (2.2a).

The 'intermittent mixing model' in (2.2c) works by letting the tracer advect for a time τ (that is, the right-hand side of (2.1) is zero) and then instantaneously vertically homogenizing the tracer. Thus, at times $t = n\tau$, $c(x, z, t) \rightarrow \overline{c}(x, t)$. This might model event-driven mixing, such as the passage of storms, or mixing which occurs only at some point in a tidal cycle (e.g., Linden and Simpson, 1988).

Our conclusion is that, in the long term, the large-scale dispersion of the tracer can be described by an effective diffusion equation for $\overline{c}(x, t)$:

$$\overline{c}_t + \overline{u}\overline{c}_x = D_{\text{eff}}\overline{c}_{xx}. \tag{2.4}$$

The effecive diffusivity in (2.4) is given by $D_{\text{eff}} = \alpha U^2\tau$ where α is a model-dependent dimensionless constant, U is the scale of $u(z)$, and τ is the vertical homogenization time. For the model in (2.2a), $\tau = H^2/D$ and then (2.4) is shear dispersion (Taylor, 1953).

Thus, the details of the mixing model on the RHS of (2.1) are not important; one expects that the combination of the sheared velocity field, $u(z)$, and the vertical mixing will always produce an effective diffusion coefficient in the vertically averaged tracer equation in (2.4). In this case it does not pay to get too exercised about which of the models in (2.4) is 'better' or 'more realistic.' In various circumstances all three models are plausible representations of small-scale mixing processes, and all three models result in an effective diffusion equation that describes the dispersion of tracer on large scales and long times.

There are many ways of deriving (2.4). Let us take a simple and instructive route, and use the episodic mixing model in (2.2c). Without any loss of generality we can take $\overline{u} = 0$. (If $\overline{u} \neq 0$ we can change frame so that the vertically averaged velocity is zero.) Suppose that one maintains a uniform gradient of tracer, G, and that at $t = 0^+$ a mixing event has just

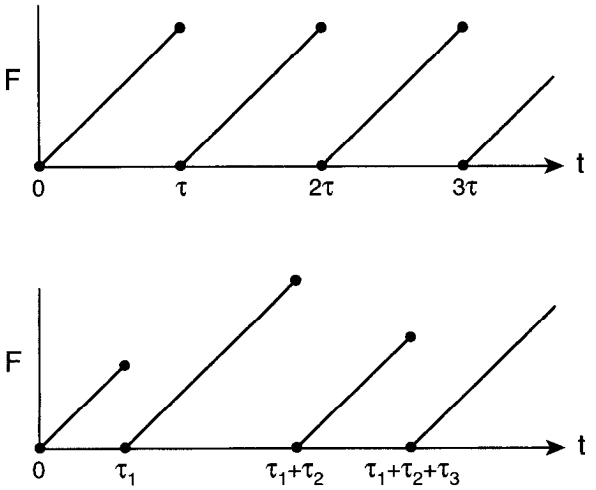

Figure 2. The flux as a function of time is a "sawtooth." (a) If the mixing events occur periodically, the flux is a periodic function of time. (b) If the events occur at random intervals, then the sawtooth is irregular.

occurred so that the tracer is vertically homogenized

$$c(x, z, 0^+) = Gx. \tag{2.5}$$

Then, in the interval before the next mixing event, the tracer is advected by $u(z)$ so that

$$c(x, z, t) = G[x - u(z)t] \qquad \text{when} \qquad 0 < t < \tau. \tag{2.6}$$

The result in (2.6) is the solution of (2.1), with zero RHS, and (2.5) as an initial condition. Now at $t = \tau$ the tracer is again vertically homogenized so that the concentration returns to (2.5). It looks like nothing has happened as a result of the advection and mixing. But the interesting point is that there has been a flux of tracer during the interval between mixing events:

$$\begin{aligned} F(x, t) &= \frac{1}{H}\int_{-H/2}^{H/2} u(z)G[x - u(z)t]\, dz, \\ &= -\overline{u^2}Gt, \qquad \text{when} \qquad 0 < t < \tau, \end{aligned} \tag{2.7a,b}$$

where we have used $\overline{u} = 0$ in passing from (2.7a) to (2.7b). The flux as a function of time has the sawtooth pattern shown in Figure 2; after each mixing event the flux starts at zero and then grows linearly till the next event homogenizes the tracer and resets the flux to zero.

The final step is to note that the time average of the flux in the top panel of Figure 2 is

$$\langle F \rangle = -\tfrac{1}{2}\overline{u^2}\tau G. \tag{2.8}$$

The expression in (2.8) is the flux-gradient relation which is typical of a diffusive process; the diffusivity is $D_{\text{eff}} = \overline{u^2}\tau/2$.

It is educational to repeat the calculation above for the other two mixing models in (2.2); in both cases one looks for a steady solution of the form $c(x, z, t) = \Gamma x + c'(z)$. Substitution gives an easy equation for $c'(z)$ and then the final step is to calculate the vertically averaged flux and identify the constant of proportionality in the flux-gradient relation as the effective diffusivity. We can summarize the results of these calculations by giving explicit expressions for the effective diffusivity in the three cases

$$\begin{aligned} D_{\text{eff1}} &= D^{-1}\overline{\psi^2}, \\ D_{\text{eff2}} &= \tau\overline{u^2}, \\ D_{\text{eff3}} &= \tfrac{1}{2}\tau\overline{u^2}, \end{aligned} \tag{2.9a,b,c}$$

where in (2.9a), $\psi(z)$ is the streamfunction of the velocity in Figure 1; i.e., $u = -\psi_z$. At the level of scale analysis $\psi \sim HU$ and all three results in (2.9) have $D_{\text{eff}} \sim U^2\tau$ where τ is the vertical mixing time.

A distinguishing characteristic of shear dispersion is that weaker vertical mixing (i.e., larger τ) produces stronger horizontal mixing. At first this seems curious, but the sawtooth graph of $F(t)$ in Figure 2 shows how this inverse dependence arises. The sheared velocity field 'unmixes' the tracer by tilting over the initial field in (2.5) and the horizontal flux grows continuously as this unmixing proceeds. If the time between mixing events is increased, then one 'accumulates' more flux during the advective process.

The flux-gradient relation in (2.8) was obtained by considering a uniform gradient, G, and averaging in time over many cycles. We now argue that the same diffusivity, $\overline{u^2}\tau/2$, approximately describes the dispersion of tracer in situations in which the gradient is not constant. That is, the vertically averaged tracer diffusion equation is

$$\overline{c}_t = (\overline{u^2}\tau/2)\overline{c}_{xx}; \tag{2.10}$$

one can add additional complications such as sources etcetera. The approximation being made here is that the flux-gradient relation in (2.8) applies locally, provided that the concentration $\overline{c}$ is 'slowly varying.' To understand 'slowly varying,' observe that in the interval between mixing events a fluid particle travels through a distance $\ell \sim U\tau$. If the concentrtion gradient is approximately uniform over this length scale, then (2.8) will apply locally. The length scale associated with the variation of the concentration gradient is c_x/c_{xx}. Thus we reach the conclusion that (2.10) applies provided that $U\tau c_{xx}/c_x \ll 1$.

The conclusions above can be reinforced by solving an initial value problem. Suppose that $c(x, z, 0) = \exp[ikx]$. Then the solution of the advection equation (2.1) in the first interval, when $0 < t < \tau$, is

$$c(x, z, t) = \exp[ik(x - u(z)t)]. \tag{2.11}$$

Now at $t = \tau$ the field in (2.11) is instantaneously vertically homogenized. As a specific example, take $u = sz$. With this velocity profile, vertically averaging (2.11) at $t = \tau$ gives

$$c(x, z, \tau^+) = \text{sinc}\,(\kappa)\exp[ikx], \tag{2.12}$$

where $\text{sinc}\,\kappa \equiv \kappa^{-1}\sin\kappa$ and $\kappa \equiv kHs\tau/2$. The tracer distribution in (2.12) is proportional to the initial condition, but there is k-dependant attenuation factor. Iterating the process one finds that after n mixing events

$$\begin{aligned} c(x, z, n\tau^+) &= \text{sinc}^n\,(\kappa)\exp[ikx], \\ &= \exp[n\ln[\text{sinc}\,\kappa] + ikx]. \end{aligned} \tag{2.13a,b}$$

If $kHs\tau/2 \ll 1$ (i.e., on large horizontal scales), then $\ln[\text{sinc}\,\kappa] \approx -(kHs\tau)^2/24$. Thus, since $t \approx n\tau$, one can rewrite (2.13b) as

$$c(x, z, n\tau^+) \approx \exp[-D_{\text{eff}}k^2t + ikx], \tag{2.14}$$

where $D_{\text{eff}} = H^2s^2\tau/24$ is the effective diffusivity in (2.9c) with $u = sz$. In this example we see explicitly how the small horizontal wavenumbers decay at the rate predicted by the diffusion equation (2.10).

Before we leave the passive scalar problem we mention an embellishment of the intermittent mixing model. Suppose that the interval between mixing events is a random variable; immediately after an event has occurred, the time to the next event is determined by selecting τ from a probability distribution with probability density function $\mathscr{P}(\tau)$ (see the lower panel in Fig. 2). What is the effective diffusivity in this case? One can show that the generalization of (2.9c) is

$$D_{\text{eff}} = \frac{1}{2}\frac{E(\tau^2)}{E(\tau)}\overline{u^2}, \tag{2.15}$$

where $E(f(\tau)) \equiv \int_0^\infty f(\tau)\mathscr{P}(\tau)\,d\tau$ is the average using the probability density function of intervals between mixing events (normalization is $E(1) = 1$). It is clear from Figure 2 that more tracer is transported when, by chance, there is a large interval with no mixing events. Intuitively, this is the reason for the appearance of the second moment, $E(\tau^2)$, in (2.15).

3. Shear dispersion in a stratified, nonrotating fluid

Now let us turn to a problem in which the tracer is buoyancy. This case is different from the ones discussed above because buoyancy is dynamically active. As far as possible we follow the same route as Section 2. Thus, we consider the horizontal fluxes associated with maintaining uniform horizontal gradients of temperature, salinity and buoyancy. There are some algebraic complications introduced by considering the possibility that the temperature and salinity gradients are not parallel. But this generalization is very useful, and so we make no assumptions concerning the colinearity of the thermohaline gradients.

Consider a nonrotating stratified fluid. We use the Boussinesq approximation and write

the density as $\rho = \rho_0[1 - g^{-1} B(x, y, z, t)]$ where $B(x, y, z, t)$ is the buoyancy. Assuming that the equation of state is linear, and using suitable definitions, the buoyancy is

$$B = T - S, \tag{3.1}$$

where T is the 'temperature' and S is the 'salinity.' With our definitions, S, T, B and g all have the dimensions of acceleration.

The equations of motion are then

$$\begin{aligned} \frac{D\mathbf{u}}{Dt} &= -\nabla p + B\hat{\mathbf{z}} + \text{mix}_3, \\ \nabla \cdot \mathbf{u} &= 0, \\ \frac{DS}{Dt} &= \text{mix}_3, \\ \frac{DT}{Dt} &= \text{mix}_3, \end{aligned} \tag{3.2}$$

where $\mathbf{u} = (u, v, w)$ and "mix_3" indicates that we use the episodic mixing model in (2.2c) and apply this instantaneous homogenization to momentum and the stratifying components. Suppose that at $t = 0^+$ a mixing event has just occurred so that T and S are vertically homogeneous and the momentum is also completely mixed. Thus the initial condition is

$$S(x, y, z, 0) = \mathbf{G}_S \cdot \mathbf{x}, \qquad T(x, y, z, 0) = \mathbf{G}_T \cdot \mathbf{x}, \qquad \mathbf{u}(x, y, z, 0) = (0, 0, 0). \tag{3.3}$$

In (3.3), the two gradient vectors $\mathbf{G}_S$ and $\mathbf{G}_T$ are constant vectors in the horizontal plane and $\mathbf{x} = (x, y, z)$.

There is an exact solution of (3.2) and (3.3) which illustrates how dense fluid flows under light fluid and so releases the potential energy stored in the stratification:

$$\begin{aligned} \mathbf{u} &= -\mathbf{G}_B z t, \\ p &= \mathbf{G}_B \cdot \mathbf{x} z + \mathbf{G}_B \cdot \mathbf{G}_B z^2 \frac{t^2}{4}, \\ S &= \mathbf{G}_S \cdot \mathbf{x} + \mathbf{G}_B \cdot \mathbf{G}_S z \frac{t^2}{2}, \\ T &= \mathbf{G}_T \cdot \mathbf{x} + \mathbf{G}_B \cdot \mathbf{G}_T z \frac{t^2}{2}, \end{aligned} \tag{3.4a,b,c,d}$$

where $\mathbf{G}_B \equiv \mathbf{G}_T - \mathbf{G}_S$ is the horizontal buoyancy gradient. The solution above applies only before the system is reset by vertical mixing, i.e., only when $0 < t < \tau$.

One can easily verify that (3.4) is a solution of (3.2) and (3.3) by substitution: form $B = T - S$ by subtracting (3.4c) and (3.4d) and then the pressure in (3.4b) is determined by the

hydrostatic relation. There is a horizontal pressure gradient, $\nabla p = \mathbf{G}_B z$, parallel to the buoyancy gradient and proportional to z. This pressure gradient drives the accelerating, vertically sheared flow in (3.4a): this is simply dense fluid slumping under lighter fluid. Finally, one can show that the tracer advection equations (3.2c) and (3.2d) are satisfied. The temperature and salinity fields are vertically homogeneous at $t = 0$, but because of the accelerating vertical shear in the velocity field they develop vertical structure which grows like t^2.

Notice that in (3.4) we have assumed that there is no vertically integrated transport, $\bar{\mathbf{u}} = 0$. (We use the notation in Fig. 1, so that the fluid occupies the region $H/2 > z > -H/2$). The condition $\bar{\mathbf{u}} = 0$, which might be enforced by distant vertical sidewalls, is used to determine the constant of integration which arises when one determines the pressure by integrating the hydrostatic relation. Thus, the net horizontal flux of mass in the layer is zero. However there is an *exchange* of mass (analogous to a lock exchange, in which dense water flows below light water). This is the fundamental mechanism behind all of the different models mentioned in the introduction. Even in rotationally dominated systems, such as Stommel's (1961) thermohaline problem, there are inevitably processes which release the gravitational potential energy stored in horizontal nonuniformities of the stratification. Indeed, an analysis of the energy balance of the solution (3.4) is instructive. One finds that the potential energy is decreasing as the fluid 'restratifies.' All of the kinetic energy that appears in (3.4a) comes from the local lowering of the center of mass; the pressure work terms are identically zero.

The main point here is that the velocities associated with the release of potential energy are larger when the horizontal buoyancy gradients are larger. Consequently, as we show below in (3.5), the net flux of heat and salt depends nonlinearly on the buoyancy gradient.

We are uncertain of the origins of the solutions in (3.4); a rotating version of this problem (but with only one stratifying component) appeared as a question in the 1977 general exam of the MIT-WHOI Joint Program in Oceanography. Different elaborated versions of (3.4) have been used by Simpson and Linden (1989), Roemmich *et al.* (1994) and Tandon and Garrett (1994) as illustrations of how horizontal buoyancy gradients drive vertically sheared horizontal flows which in turn create vertical buoyancy gradients ('restratification').

Although the various balances in (3.4) are easy to understand intuitively, there are some nonintuitive implications for the fluxes. For instance, using (3.4a) and (3.4c) the salt flux which occurs during the interval between mixing events is:

$$\mathbf{F}_S = 1/H \int_{-H/2}^{H/2} \mathbf{u} S \, dz = -\frac{1}{24} (\mathbf{G}_S \cdot \mathbf{G}_B) \mathbf{G}_B H^2 t^3 \quad \text{when} \quad 0 < t < \tau. \tag{3.5}$$

Because the scalar is active, the flow is accelerating between mixing events and so the flux in (3.5) grows like t^3, rather than following the linear-in-t growth shown in Figure 2. The final step is to calculate how much salt is transported by integrtaing (3.5) in time, and then

averaging over many mixing cycles. One finds that the analog of (2.8) is

$$\langle \mathbf{F}_S \rangle = -1/96 H^2 \tau^3 (\mathbf{G}_S \cdot \mathbf{G}_B)\mathbf{G}_B. \tag{3.6}$$

The flux of salt is proportional to the cube of the tracer gradients and, remarkably, the flux of salt is in the direction of the buoyancy gradient.[3]

With (3.6), and the analogous result for temperature, we can now write down the coupled equations that describe the buoyancy driven shear dispersion of vertically averaged heat and salt:

$$T_t = \gamma \nabla \cdot [(\nabla B \cdot \nabla T)\nabla B], \qquad S_t = \gamma \nabla \cdot [(\nabla B \cdot \nabla S)\nabla B], \tag{3.7}$$

where, from (3.6), $\gamma \equiv H^2\tau^3/96$. Notice that if we subtract (3.7a) from (3.7b) then we obtain a closed equation for the buoyancy viz.

$$B_t = \gamma \nabla \cdot [(\nabla B \cdot \nabla B)\nabla B]. \tag{3.8}$$

One might then be tempted to ignore the passive combination of T and S, called 'spice' by Munk (1981), as trivial. However, the passive combination is sometimes used to interpret oceanographic data and so it is worthwhile to develop some understanding of spice dynamics which goes beyond the early works of Stommel (1962) and Veronis (1972); this is a goal of Section 4.

In Section 2 the effective shear dispersion coefficient was obtained using three different models of vertical mixing. One can easily go through the same exercise here. In fact, the one-dimensional version of the nonlinear diffusion equation (3.8) was first obtained by Erdogan and Chatwin (1967) using molecular diffusion and viscosity as a mixing model.[4] Young (1994) obtained (3.7) using relaxation to vertical average as a model of mixing. These alternative models differ from the episodic mixing model because heat, salt and momentum do not have to mix at the same rates. Thus Chatwin and Erdogan have both a viscosity, ν, and a buoyancy diffusivity, κ. Young's model relaxes the momentum to its vertical average on a time scale τ_U, while T and S relax with a time scale τ. The coupled nonlinear diffusion equations in (3.7) result from all three models; the details of the mixing model are contained in the coefficient γ in (3.7).

The flux-gradient relation in (3.6) is obtained by assuming that the gradients $\mathbf{G}_S$ and $\mathbf{G}_T$ are uniform. Thus, validity of (3.7) requires that ∇T and ∇S are 'slowly varying' so that the solution in (3.4) applies locally. To understand the implications of 'slowly varying,' observe that between mixing events, fluid travels through a distance of order $\ell \sim B_x H \tau^2$. If the buoyancy gradient 'looks' uniform over this distance then we expect (3.4) to apply locally. The obvious length associated with the variation of the buoyancy gradient is B_x/B_{xx}.

3. Eddy mixing arguments would say that the flux of salt is parallel to the salt gradient. But in this problem the velocity in (3.4a) is always in the direction of the buoyancy gradient, and so are all the shear dispersive fluxes. However, notice that, from (3.6), $\mathbf{F}_S \cdot \mathbf{G}_S < 0$; thus the flux of salt tends to be down the salt gradient.

4. Because Erdogan and Chatwin used molecular diffusivity, rather than episodic mixing, the vertical mixing time scale τ is proportional to H^2. In this case, the analog γ is proportional to the eighth power of the layer depth and inversely proportional to the factor $\nu^2\kappa$ where ν is the viscosity and κ the buoyancy diffusivity.

Requiring that this length scale be less than ℓ shows that 'slowly varying' means that the nondimensional number $B_{xx}H\tau^2$ is small.

Simpson and Linden (1989) give an interesting theoretical and experimental discussion of the role of buoyancy curvature, B_{xx}, in driving frontogenesis. In the simple solution in (3.4) the buoyancy curvature is zero and there is no frontogenesis: the B-surfaces tilt over and maintain equal spacing; the horizontal buoyancy gradient is constant. But if the initial condition has nonzero buoyancy curvature, then Simpson and Linden show that the buoyancy gradient can increase. If there is no mixing then one suspects that B_x might become infinite in a finite time; this singularity is frontogenesis. With the episodic mixing model there is a race between the next homogenization event and the singularity formation: given an initial condition which happens first? Simpson and Linden show that the time to singularity formation scales like $1/\sqrt{B_{xx}H}$. This race is then decided by the same nondimensional parameter, $B_{xx}H\tau^2$, that we identified in our discussion of the validity of (3.7).

Finally, let us discuss the range of oceanic parameters for which the nonlinear diffusion equations (3.7) might apply. In our derivation we have ignored the effects of rotation (for which, see Young (1994) and Talley (1998)) and we have averaged over the depth, say $H =$ 100 m, of the mixed layer. For the time between mixing events we take $\tau \sim 10^4$ s. On scales of $\Delta x \sim 1$ km, horizontal buoyancy gradients $\mathbf{G}_B$ of order 10^{-7} s^{-2} are common in the mixed layer (this corresponds to a $\Delta T \sim 0.1°$ over a distance $\Delta x \sim 10^3$ m). With $H \sim$ 100 m, the velocity obtained from (3.4a) is $\mathbf{u} \sim \mathbf{G}_B H\tau \sim 0.1$ m s^{-1}. With these same numbers, if we estimate B_{xx} as $\mathbf{G}_B/\Delta x$, again with $\Delta x = 1$ km, we find $B_{xx}H\tau^2 \sim 1$. Thus the nonlinear diffusion model is marginally valid (and likewise the frontogenetic mechanism of Simpson and Linden (1989) cannot be discounted in the ocean mixed layer).

4. Creation of density compensated thermohaline structure by nonlinear diffusion

In this section we explore the implications of the nonlinear diffusion equations (3.7) in an idealized setting. Suppose that spatial variations in temperature and salinity are created at some instant by random forcing and that these nonuniformities subsequently disappear as a result of the nonlinear diffusion. This is a "rundown" problem in which the down-gradient diffusion eventually erases the initial variations in temperature and salinity. There is no internal mechanism for producing new randomness; all the randomness is inherited from the initial conditions. One can imagine that statistical problems such as this are relevant to what happens when a shallow layer of water is forced by random meteorological inputs.

We show that even though the initial salinity and initial temperature are uncorrelated, the result of the nonlinear diffusion is to create correlations between the thermohaline fields and their gradients. The physical reason is that in the initial conditions there will be regions in which temperature and salinity gradients happen to partially compensate in their joint effect on buoyancy. The nonlinear diffusion coefficient will be small in those regions, and so the initially compensating gradients will persist. Likewise, the initial condition will also contain regions in which the thermohaline gradients accidentally reinforce in their joint

effect on buoyancy; those regions are then subject to strong diffusion and the gradients will disappear quickly. The consequence of this selective decay is that the thermohaline structure becomes 'spicy' as the system runs down.

However, selective decay is not the whole story: we will show that the nonlinear diffusion also actively produces compensated temperature and salinity gradients.

We consider the simplest problem in which the temperature and salinity have spatial variations in only the x-direction. Eqs. (3.7) then reduce to

$$T_t = \gamma(B_x^2 T_x)_x, \qquad S_t = \gamma(B_x^2 S_x)_x. \tag{4.1a,b}$$

The one-dimensional case is sufficient to illustrate the development of thermohaline correlations from random initial conditions and the simplicity of a one-dimensional model enables us to perform extensive Monte Carlo simulations and make analytic headway. But before discussing the evolution of random initial thermohaline fields, we derive some analytic solutions of the nonlinear diffusion equations. This exercise explains some counterintuitive features of the model and establishes a framework for interpreting the Monte Carlo simulations.

a. Self-similar solutions of the nonlinear diffusion model. To simplify the algebra we introduce buoyancy $B \equiv T - S$ and spice $Q \equiv T + S$ as new independent variables. In terms of these variables the nonlinear diffusion equations (4.1) are

$$B_t = \gamma(B_x^3)_x, \qquad Q_t = \gamma(B_x^2 Q_x)_x, \tag{4.2a,b}$$

Differentiating (4.2) with respect to x gives alternative diffusion equations for the buoyancy gradient $h \equiv B_x$ and the spice gradient $j \equiv Q_x$:

$$h_t = \gamma(h^3)_{xx}, \qquad j_t = \gamma(h^2 j)_{xx}. \tag{4.3a,b}$$

Physically, $h(x, t)$ is the active tracer whose dynamics is governed by a nonlinear power-law diffusion. The evolution of the passive tracer j, or for that matter of any other passive linear combination of T_x and S_x, depends on the buoyancy gradient h and on the initial profiles.

The nonlinear coupled equations (4.2) and (4.3) have analytic solutions only for particular choices of the initial conditions. A simple and instructive choice is

$$h(x, 0) = C\delta(x), \qquad j(x, 0) = j_0. \tag{4.4a,b}$$

In their discussion of thermal conduction in fluids, Landau and Lifshitz (1993) give a self-similar solution of (4.3a) with the initial condition in (4.4a):

$$h = C\sigma(\gamma t)^{-1/4}(2/\pi)\sqrt{1 - \eta^2}H(1 - \eta^2), \qquad \eta \equiv \sigma x/(\gamma t)^{1/4}, \tag{4.5}$$

where $H(x)$ is the Heaviside step function and $\sigma \equiv (\pi^2/48C^2)^{1/4}$. The initial condition has a sharp buoyancy front of strength C; the diffusive evolutions preserves the strength of the

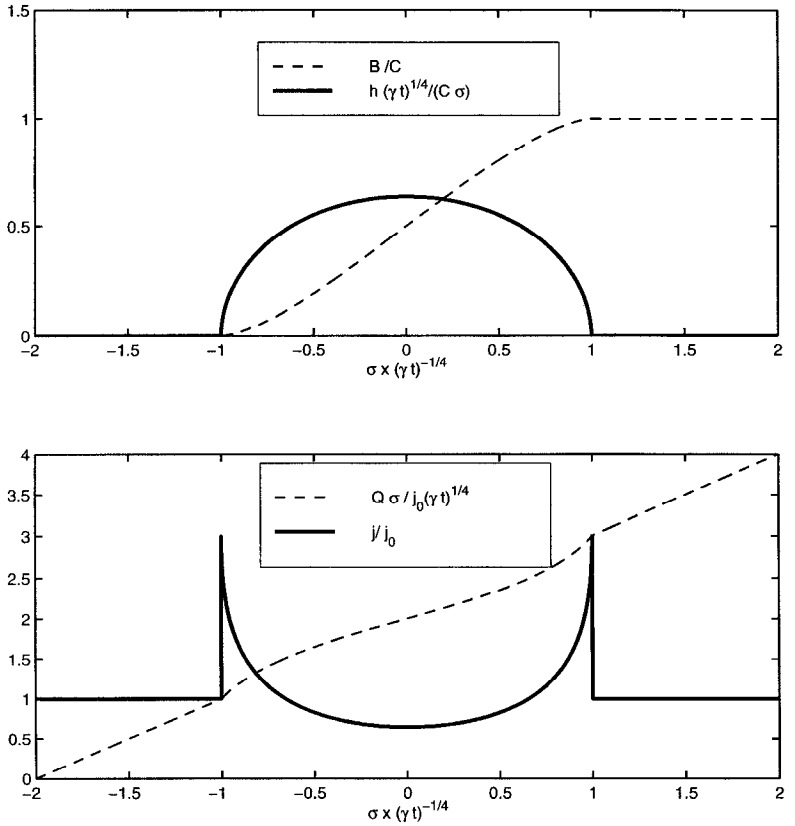

Figure 3. Self-similar solutions of the Eqs. (4.3). (Upper panel) Scaled profiles of buoyancy B and buoyancy gradient h. (Lower panel) Scaled profiles of spice Q and spice gradient j.

front, C, but the width of the front increases as $t^{1/4}$ (see Fig. 3). The buoyancy gradient is zero when $|\eta| > 1$ i.e., for $x > (\gamma t)^{1/4}/\sigma$. In other words, there is an expanding front separating the two regions which have not yet felt the influence of the initial discontinuity in buoyancy. This behavior is very different from the well-known Gaussian solution of the analogous problem with constant diffusivity, for which all regions of space are immediately affected by an initial point disturbance.

There is also a self-similar solution of (4.3b) with the initial condition in (4.4b). The details are in the Appendix and the result is

$$j = \begin{cases} 3j_0 \cosh\left[\sqrt{2} \arcsin \eta\right]/\cosh\left(\pi/\sqrt{2}\right), & \text{if } |\eta| < 1; \\ j_0, & \text{if } |\eta| > 1. \end{cases} \tag{4.6}$$

The solution in (4.6) is shown in the lower panel of Figure 3. In the regions where $h = 0$ the initially uniform spice gradient is undisturbed. In the region where $h \neq 0$ the profile of spice is deformed. Remarkably, near the edges of the disturbed region, the spice gradient j

is larger than its initial value (by up to a factor of 3). The creation of enhanced spice gradients is in contrast with the continuous dissipation of buoyancy gradients.

It is helpful to physically interpret this simple problem. The initial conditions correspond to a vertically averaged mixed layer composed of two water masses of different density separated by a sharp front: the solution (4.5) describes the nonlinear diffusion of this front. The initial condition also has a uniform gradient in spice. The spice is a passive tracer and is diffused only where there is a buoyancy gradient. The dashed curve in the lower panel of Figure 3 shows the developing profile of spice. In the region where the buoyancy gradient h is nonzero the spice gradient, j, is nonuniform. In the central region j is reduced; but at the edges j is three times larger than its initial value. Thus nonlinear diffusion can result in a substantial localized amplification of pre-existing spice gradients. This process has an important role to play in the statistical problems discussed next.

b. A scaling argument for the decay of buoyancy variations. We now focus our attention on the evolution of random initial conditions in the temperature and salinity fields. We denote the correlation length of the initial random buoyancy distribution by ℓ_0 and the RMS buoyancy fluctuations by $\mathcal{B}_0$. Thus, the initial buoyancy gradient is of order $\mathcal{B}_0/\ell_0$. We assume that the domain has a horizontal length $L \gg \ell_0$ so that there are many independent samples of the buoyancy field. We now give a simple scaling argument for the decay in time of the RMS buoyancy fluctuations $\mathcal{B}(t)$ and the increase in time of the correlation length $\ell(t)$ as the random initial distribution runs down under the dynamics in (4.1).

Applying dimensional considerations to the diffusion equation (4.2a) gives a relation connecting $\mathcal{B}(t)$ and $\ell(t)$

$$\frac{d\mathcal{B}}{dt} \sim \gamma \mathcal{B}^3 \ell^{-4}. \tag{4.7}$$

A second scaling relation is obtained by arguing that diffusion, either linear or nonlinear, is a conservative averaging process. In a length $\ell(t)$ there are

$$N(t) \sim \ell(t)/\ell_0 \tag{4.8}$$

independent samples of the random initial condition. Then, from the central limit theorem, the RMS buoyancy fluctuations obtained by diffusively averaging the N independent samples scales as $\mathcal{B}/\sqrt{N}$ or, from (4.8),

$$\mathcal{B}(t) \sim \sqrt{\ell_0/\ell(t)}\,\mathcal{B}_0. \tag{4.9}$$

Eqs. (4.7) and (4.9) are two relations for the two unknowns $\mathcal{B}(t)$ and $\ell(t)$. It is easy to see that there is a power law solution[5]

$$\mathcal{B}(t) \sim \mathcal{B}_0(t/\tau_\gamma)^{-1/10}, \qquad \ell(t) \sim \ell_0(t/\tau_\gamma)^{1/5}, \tag{4.10a,b}$$

5. The same arguments show that for a linear diffusion equation with random initial conditions $\ell \sim t^{1/2}$ and $\mathcal{B} \sim t^{-1/4}$.

where $\tau_\gamma \equiv \ell_0^4/\gamma \mathscr{B}_0^2$ is the fundamental time scale of the problem.

The argument leading to (4.10) suggests that the solution of (4.1), with random initial distributions of T and S, will exhibit a self-similar scaling regime characterized by a $t^{1/5}$ growth of the buoyancy correlation length. Up to this point we have considered only the buoyancy $B = T - S$. Before continuing with theoretical arguments we need to test the scaling argument in (4.10) by comparison with numerical solutions of (4.1).

c. The numerical simulation. For the simulations we will use dimensionless variables. Buoyancy, temperature and salinity will be measured in terms of the initial RMS buoyancy variations ($\mathscr{B}_0$), lengths in terms of the initial correlation length (ℓ_0), and time in terms of $\tau_\gamma \equiv \ell_0^4/\gamma \mathscr{B}_0^2$. τ_γ is the time it takes to mix tracers over a distance ℓ_0 using nonlinear diffusion driven by the initial RMS buoyancy gradients. In terms of nondimensional variables one simply sets $\gamma = 1$ in (4.1).

We take the initial temperature and salinity to have zero correlation and we introduce a parameter, α, to quantify the difference in the RMS contributions of salinity and temperature to that of buoyancy:

$$\langle TS \rangle = 0, \qquad \alpha = \langle T^2 \rangle - \langle S^2 \rangle, \tag{4.11a,b}$$

where $\langle\ \rangle$ is an integral over the domain

$$\langle B^2 \rangle \equiv \frac{1}{L} \int_0^L B^2 \, dx. \tag{4.12}$$

The initial conditions are established by selecting T and S at each grid point using a uniform probability density function with zero mean such that

$$\langle B^2 \rangle = 1, \qquad \langle T^2 \rangle = \frac{1+\alpha}{2}, \qquad \langle S^2 \rangle = \frac{1-\alpha}{2}. \tag{4.13a,b,c}$$

Thus the initial correlation length is of the order of the grid spacing and the profiles of temperature and salinity are just white noise. Since the domain has a finite size, the conditions in (4.13) will not be precisely satisfied by the initial profiles. For instance, if one picks $L = 10{,}000$ random values of B from a distribution which satisfies (4.13a) then $\langle B^2 \rangle$ of this sample will typically deviate from 1 by a quantity of order $1/\sqrt{L} \sim 1/100$.

The numerical calculations are performed by integrating the nondimensional coupled equations for the temperature and salinity gradients, obtained differentiating (4.1) and setting $\gamma = 1$

$$(T_x)_t = (B_x^2 T_x)_{xx}, \qquad (S_x)_t = (B_x^2 S_x)_{xx}. \tag{4.14a,b}$$

Global conservation of heat and salt is satisfied by requiring that the gradients T_x and S_x vanish at both ends of the domain (i.e., 'no-flux' boundary conditions). We then solve the system in (4.14) on a discrete spatial grid with an explicit Euler forward scheme in time and

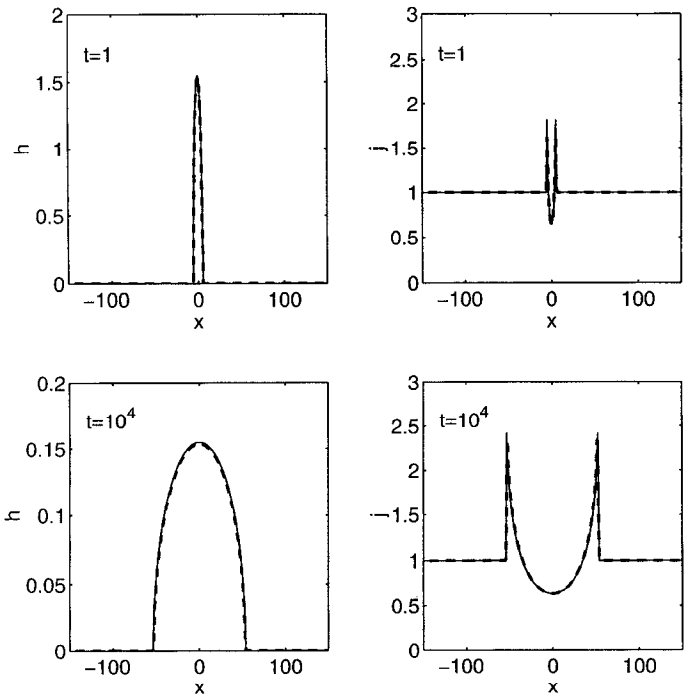

Figure 4. Comparison of numerical (dashed line) and analytic (continuous line) solutions of the coupled thermohaline equations (4.3). The initial conditions for h are given by (4.4) evaluated at $t = 10^{-2}$ with $C = 13.06$. The initial condition for j is simply $j = 1$ everywhere. The peak of h falls exactly on a grid point. The grid spacing is equal to 1. (Left panels) Comparison of the numerical results and the analytic solution (4.4) at $t = 1$ and $t = 10^4$. (Right panels) Comparison of the numerical results and the analytic solution (4.5) at $t = 1$ and $t = 10^4$. The discrete space grid does not completely resolve the two j-spikes (whose maximum value is 3).

finite differences in space. The initial time step, Δt, is short enough to accurately solve the set of ordinary, differential equations obtained by the spatial discretization of (4.14). As the simulation progresses, the buoyancy gradients become weaker and the time step can be increased while preserving numerical accuracy.

We tested this simple numerical scheme by studying the evolution of the initial conditions in (4.4) and comparing the numerics to the analytic solutions in (4.5) and (4.6) (see Fig. 4). Because we cannot resolve a δ-function the initial conditions for h is actually obtained by setting $t = 0.01$ and $C = 13.06$ in (4.5). This corresponds to a narrow single peak, of height $2\sqrt{6}$, resolved with only a few points. The initial conditions for spice is $Q = x$; i.e., $j_0 = 1$. The agreement between the analytic and numerical curves for h is shown in the left panels of Figure 4 at $t = 1$ and $t = 10^4$. In the right panels the corresponding comparison for j improves with time, because the self-similar solution applies only asymptotically as $t \to \infty$. The numerical scheme is not sensitive to the position of the initial buoyancy gradient peak with respect to the underlying grid. In Figure 5 we show a 'worst

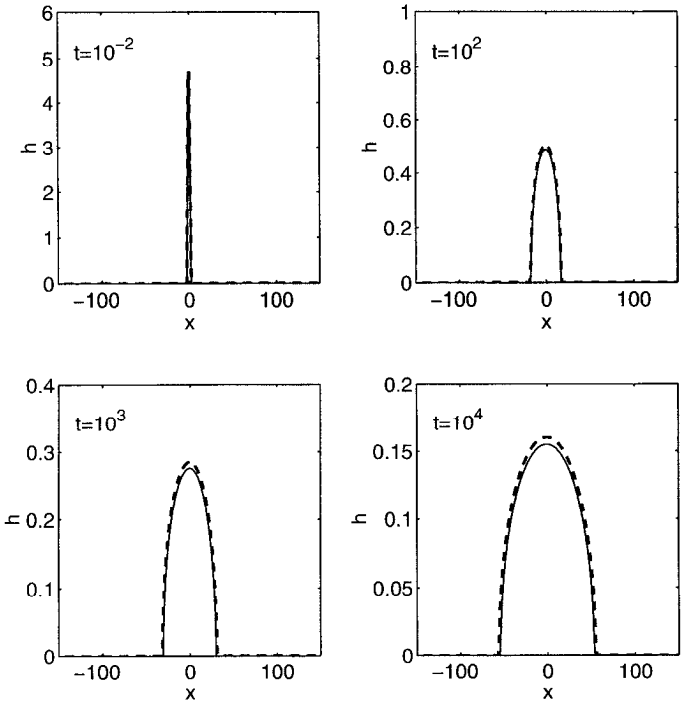

Figure 5. Comparison of numerical (dashed line) and analytic (continuous line) solutions of the nonlinear diffusion equation for h. The initial condition for h is the same as in Figure 4, but in this second example the peak falls half way between two grid points. The following three panels compare the numerical results and the analytic solution (4.4) at $t = 10^2$, $t = 10^3$ and $t = 10^4$.

case' example in which the maximum of the initial peak in h is half-way between two grid points. The initial condition is otherwise the same as in Figure 4. The four panels show the profile at $t = 0.01$, $t = 10^2$, $t = 10^3$ and $t = 10^4$. The agreement between the analytical and numerical curves remains good throughout this time interval. There is a systematic difference (equivalent to a slight discrepancy in the value of C) between the two curves which results from the poor resolution of the initial peak.

The spatial structure of the thermohaline fields is best described in terms of buoyancy $B = T - S$, spice $Q = T + S$ and their gradients h and j. Figures 6 and 7 show profiles of these fields at $t = 100$ and $t = 1000$ after their evolution from random initial conditions with $\alpha = 0$ in (4.11). The simulations are run with $N = 10{,}000$ grid points and Figures 6 and 7 show a typical subsection of 100 points in the center of the domain. The spice gradient profiles in Figures 6 and 7 reveal a number of spikes some of which are 5 times stronger than the maxima of buoyancy gradient. This is the first indication of the development of intermittency in the spice gradient or, equivalently, spicy density-compensated fronts.

The profile of j in Figure 7 and the analytic solution in (4.6) (Fig. 3) present some

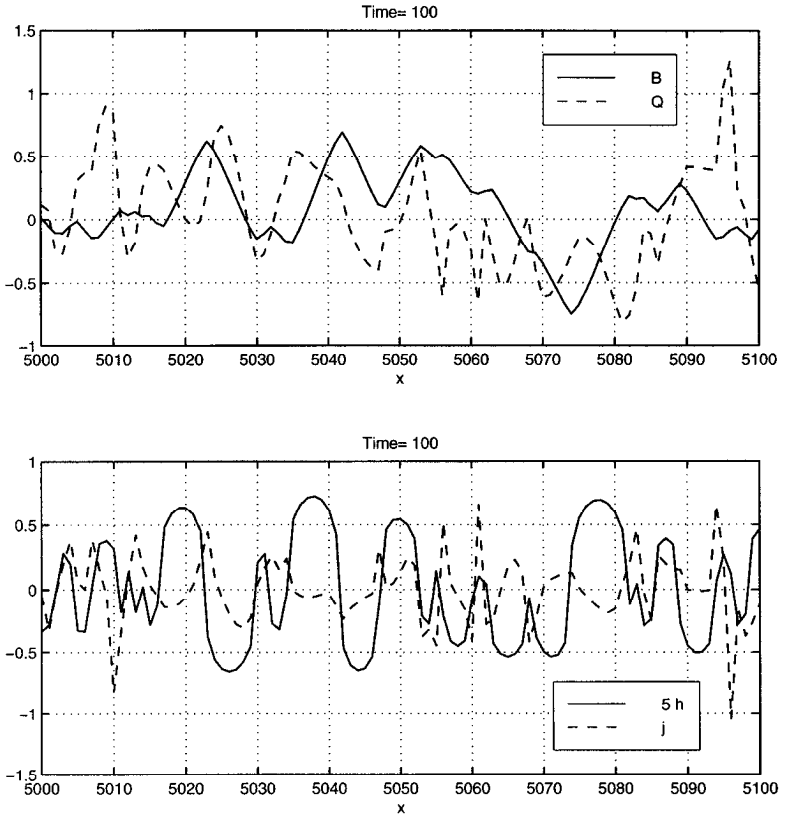

Figure 6. Profiles of B and Q (upper panel) and h and j (lower panel) at $t = 100$ after evolution from random initial conditions with $\alpha = 0$ in (4.11). The simulations are run with $N = 10{,}000$ points and these panels show a typical subsection of 100 points in the center of the domain.

interesting analogies. In both cases the spikes of j are created in the neighborhood of zeros of h where $h_x = b_{xx}$ is large. In the rest of the domain, where h is larger, and h_x smaller, j decays. The intermittency of j is then the result of spike formation in small regions surrounding the zeros of h.

d. Statistical diagnostics. To provide a quantitative interpretation of the simulations we must analyze the numerical results from a statistical point of view. Figure 8 shows the time evolution of the mean square thermohaline fields from random initial conditions with $\alpha = 0$ and $N = 10{,}000$ grid points. The first observation is that the RMS temperature and RMS salinity in the top panel of Figure 8 have a nontrivial transient behavior. The dynamics of the transient can be explained in terms of the relationships that develop between the various terms in the identities

$$\langle B^2 \rangle = \langle T^2 \rangle + \langle S^2 \rangle - 2\langle TS \rangle, \qquad \langle h^2 \rangle = \langle T_x^2 \rangle + \langle S_x^2 \rangle - 2\langle S_x T_x \rangle. \tag{4.15a,b}$$

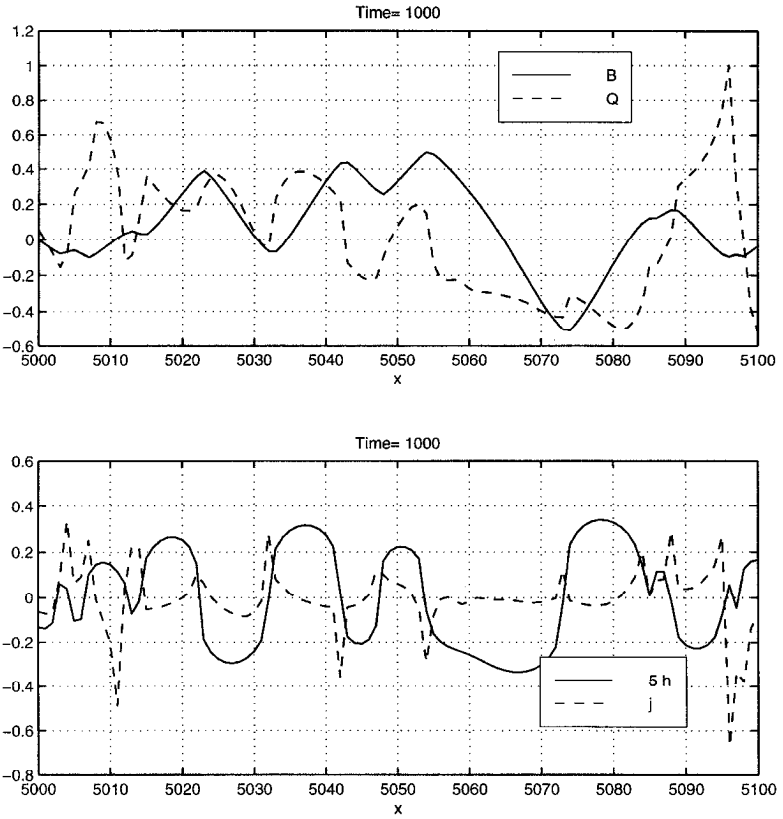

Figure 7. Profiles of B and Q (upper panel) and h and j (lower panel) at $t = 1000$ for the same simulation and subsection shown in Figure 6.

Initially, $\langle TS \rangle = \langle T_x S_x \rangle = 0$. However, because of the nonlinear diffusion, $\langle TS \rangle$ and $\langle T_x S_x \rangle$ both become positive so that the solid and dotted curves in Figure 8 approach each other. This process is weak for $\langle TS \rangle$ – the gap between the solid and dashed curves in the top panel of Figure 8 is merely reduced from its initial value. But for the gradients, the development of positive cross-correlations is so strong that the curves cross at around $t = 0.3$ in the lower panel of Figure 9. After this time RMS temperature and RMS salinity gradients are, on average, stronger than the RMS buoyancy gradient. The accentuation of positive correlation between temperature and salinity is evident in Figure 9 which shows the $T - S$ and $T_x - S_x$ scatter plots. In the gradient plane at $t = 2$ (see Fig. 9d) the thermohaline compensation is evident as an extension of the cloud of points along the 'compensation line' $T_x = S_x$. At the same time, there is a weaker, though nonzero, positive correlation in the $T - S$ plane (see Fig. 9c).

The results discussed so far are for an initial condition in which salinity and temperature contributed equally to the buoyancy field (i.e., $\alpha = 0$ in (4.13)). And, of course, $\langle S^2 \rangle = \langle T^2 \rangle$ at all times as the fluctuations decay.

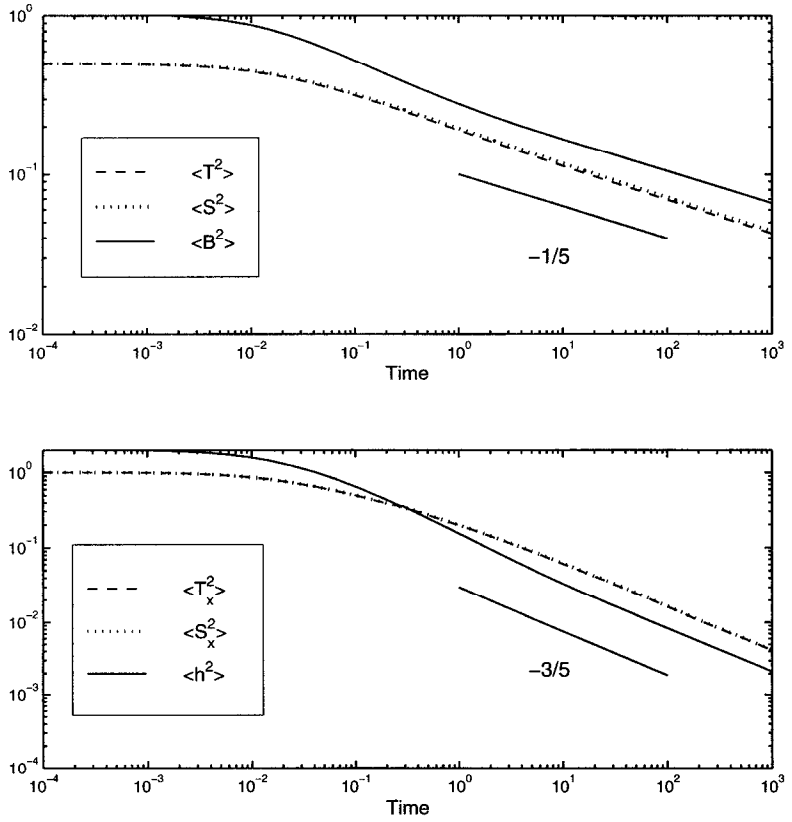

Figure 8. Time evolution of the RMS thermohaline fields (upper panel) and the RMS thermohaline gradients (lower panel). The simulations use $N = 10{,}000$ points. The asymptotic behavior is in good agreement with the power laws (4.10).

We now turn to the evolution of initial conditions with $\alpha \neq 0$. In the top panels of Figure 10 we show a case with $\alpha = \frac{1}{2}$; i.e., two random initial fields with $\langle T^2 \rangle = \frac{3}{4}$ and $\langle S^2 \rangle = \frac{1}{4}$. The bottom panels of Figure 10 show the $T - S$ and $T_x - S_x$ diagrams at $t = 2$. These scatterplots are similar to those in the bottom panels of in Figure 9 (for which $\alpha = 0$). In both cases there is strong compensation between T_x and S_x and a weaker compensation between T and S.

Scatter plots such as those in Figure 10d must be interpreted with a sensitive awareness of the distinction between 'typical' points and 'extreme' points. The gradient compensation in Figure 10d is nearly total for large thermohaline gradients; these are the extreme points which extend out along the compensation line, $T_x = S_x$. In order to quantify the degree of compensation of typical points we use the concept of a 'best-fit ellipse' to describe the tensor of inertia in thermohaline space; in particular, we use the 'orientation,' θ of this ellipse as a diagnostic. The orientation of the best-fit ellipse in the thermohaline gradient

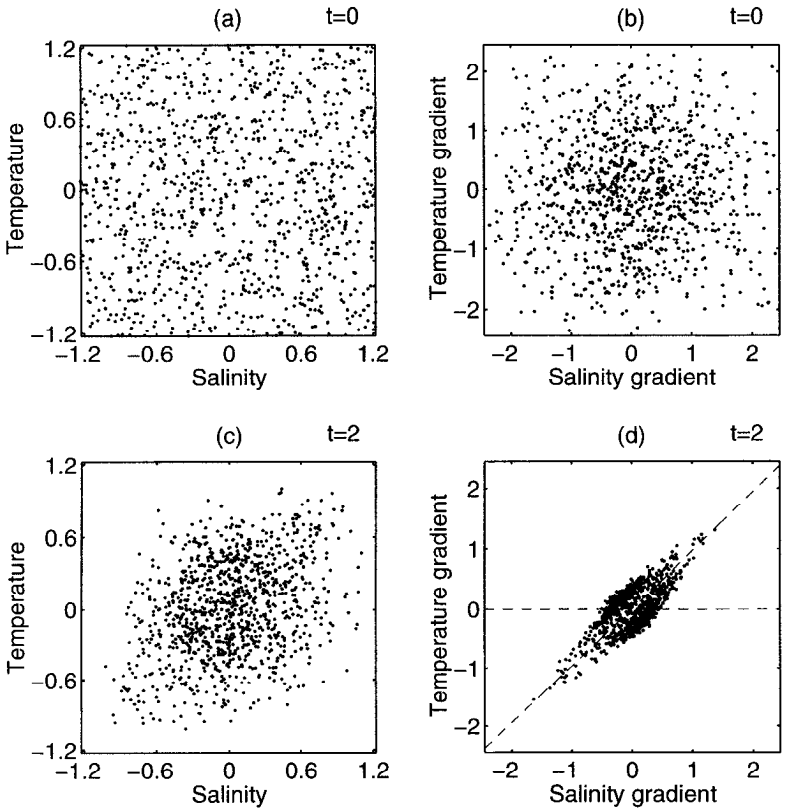

Figure 9. This figure shows the results of a simulation in which 1000 points in the (S, T) and (S_x, T_x) planes are created by picking uncorrelated temperature and salinity from a uniform probability density with variance ½; thus the variance of $B = T - S$ is one. Panels (a) and (b) show scatterplots at time $t = 0$. Panels (c) and (d) scatterplots at $t = 2$.

plane is the angle θ such that the new fields

$$X = T_x \cos\theta - S_x \sin\theta, \qquad Y = T_x \sin\theta + S_x \cos\theta \tag{4.16a}$$

are uncorrelated; explicitly

$$\theta(t) = \frac{1}{2}\,\text{atan}\left[2\frac{\langle T_x S_x\rangle}{\langle S_x^2\rangle - \langle T_x^2\rangle}\right]. \tag{4.16b}$$

The ambiguity in the branch of atan is resolved by ensuring that $0 < \theta < \pi$. (For instance, in Fig. 10b, $\theta = \pi/2$.) The major and minor axes of the best-fit ellipse are defined as the standard deviations of X and Y. We use an analogous procedure to calculate the orientation of the best-fit ellipse in the (S, T) plane.

When $\alpha \neq 0$, the major axis of the best-fit ellipse in the gradient plane does not coincide with the compensation line, i.e., $\theta \neq \pi/4$. If $\alpha > 0$, then $\theta > \pi/4$. For example, in

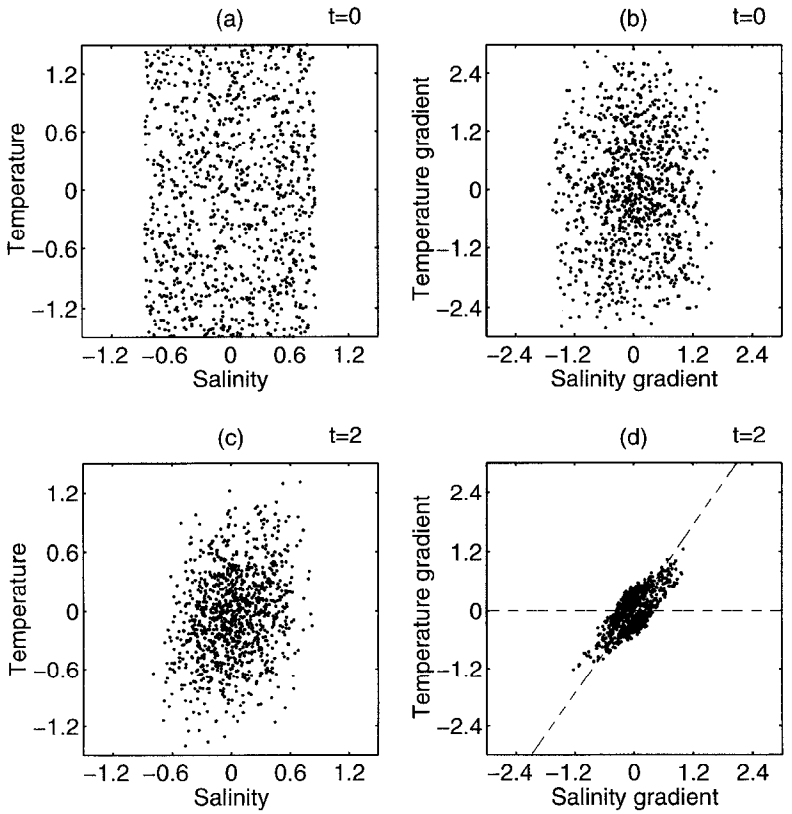

Figure 10. This figure shows the results of a simulation in which 1000 points in the (S, T) and (S_x, T_x) planes are created by picking uncorrelated temperature and salinity from a uniform probability density with a temperature variance of ¾ and a salinity variance of ¼; thus the variance of $B = T - S$ is one. These initial conditions correspond to $\alpha = ½$ in (4.11). Panels (a) and (b) show scatterplots at $t = 0$. Panels (c) and (d) show scatterplots at $t = 2$. The oblique dashed line in panel (d) represents the direction of the major principal axes of the cloud of points: it is somewhat steeper than the compensation line ($\theta = \pi/4$).

Figure 10d the major axis of the best-fit ellipse (the oblique dashed line) is steeper than the compensation line. The detailed structure of the $T_x - S_x$ diagram in Figure 10d can be better appreciated in the rotated and anisotropic coordinates h and j of Figure 11. The best fit ellipse (which does not fit very well) in Figure 11 has its major axis partially tilted with respect to the compensation line. (The anisotropic scaling of the coordinate axes in Figure 11 exaggerates the inclination of the ellipse axis to the compensation line.)

To complete the statistical description we will use the correlation coefficients

$$R(t) \equiv \langle TS \rangle / \sqrt{\langle T^2 \rangle \langle S^2 \rangle}, \qquad C(t) \equiv \langle T_x S_x \rangle / \sqrt{\langle T_x^2 \rangle \langle S_x^2 \rangle}, \tag{4.17a,b}$$

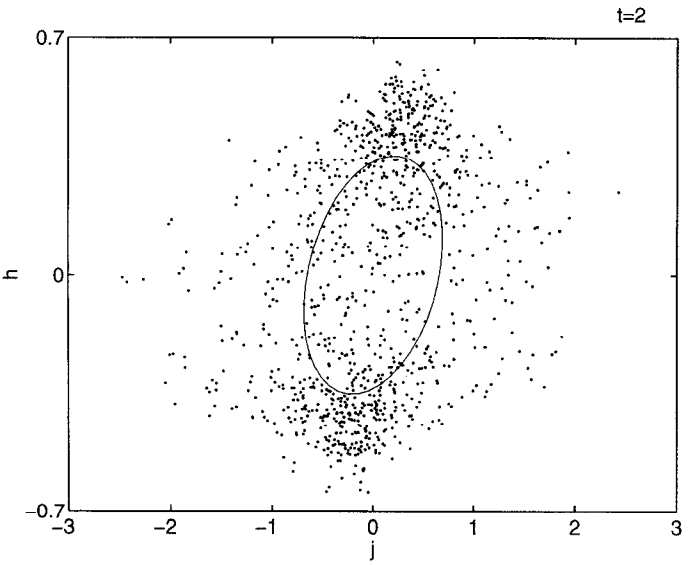

Figure 11. A detailed view of the $S_x - T_x$ diagram of Figure 10d. The axis have been rotated and scaled anisotropically. The best-fit ellipse is tilted with respect to the compensation line (which is the horizontal axis in this figure). The anisotropic scaling of the coordinates reveals the compensation between typical points in the middle of the cloud.

as diagnostics. Figures 12 and 13 show the evolution of $R(t)$ and $C(t)$ and the orientation of the corresponding best-fit ellipse for four simulations with different values of α. In Figures 12a and 13a the correlation coefficients are initially very close to zero. For all four values of α there is a fast increase (notice that the time axis is logarithmic) followed by an approach to what might be limiting values as $t \to \infty$. However, the asymptotic behavior of $C(t)$ as $t \to \infty$ is not so trivial; from Figure 12a there is a very slight increase in $C(t)$ during the final decade $10^2 < t < 10^3$. (This is more evident in the simulations with the larger values of α.) As time proceeds, the typical points collapse approaching a limiting value for both $R(t)$ and $C(t)$, while the extreme points (the spikes in spice gradient) sit on the compensation line without decaying. At large times the effect of the extreme points becomes important in the determination of the best fit ellipse. This effect happens first for larger values of α, because these simulations reach the asymptotic regime at earlier times.

Figures 12b and 13b show the evolution of the orientation, $\theta(t)$, in the (S, T) and (S_x, T_x) planes respectively. When α is nonzero and positive, $\theta(0) = \pi/2$. Then, as t increases $\theta(t)$ decreases and approaches a steady value which is a function of α. In the three cases with nonzero α in Figures 12b and 13b, the axis of the best fit ellipse is steeper than the compensation line (e.g., as in Fig. 10d). However, the rotation of the best-fit ellipse toward the compensation line indicates that there is density compensation developing amongst the typical points in the simulation.

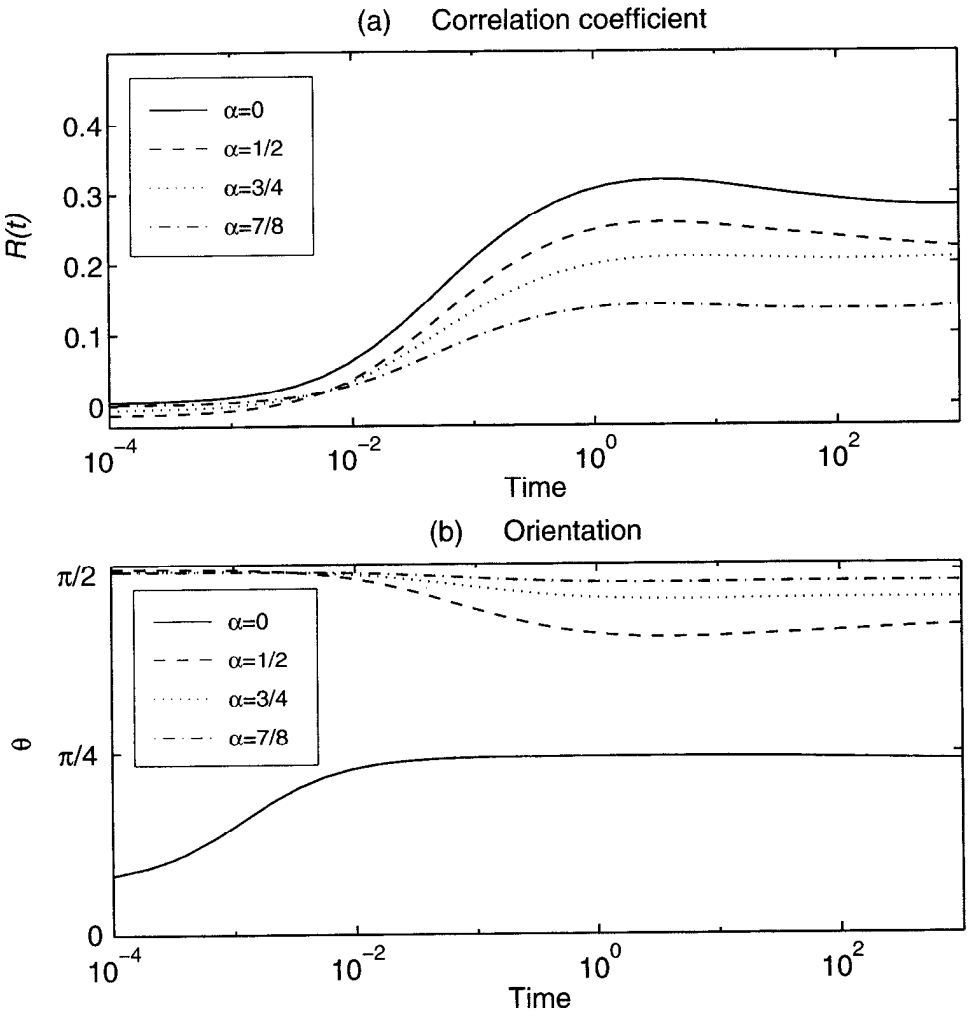

Figure 12. (a) Evolution of the correlation coefficient $R(t)$ defined in (4.17a). (b) Evolution of the orientation of the best-fit ellipse in the (S, T) plane. The initial conditions is 10,000 (S, T) points selected from uniform probability densities with variances $(1 + \alpha)/2$ and $(1 - \alpha)/2$ respectively. For $\alpha = 0$ the orientation is always close to $\pi/4$ except for an initial transient due to random asymmetries in the initial conditions. For $\alpha \neq 0$, $\theta(0) = \pi/2$; then as the time increases, $\theta(t)$ decreases and approaches a steady value which is a function of α. If $\alpha \neq 0$, the orientation is steeper for the fields than for the gradients.

Comparing Figure 12 with Figure 13, we see that all of the correlations are stronger for the gradients, (S_x, T_x), than for the fields, (S, T). (Notice the different vertical scales in Figs. 12a and 13a.) This happens because in the model, diffusion is proportional to the buoyancy gradient: buoyancy gradients are diffused everywhere (diffusion is weak only

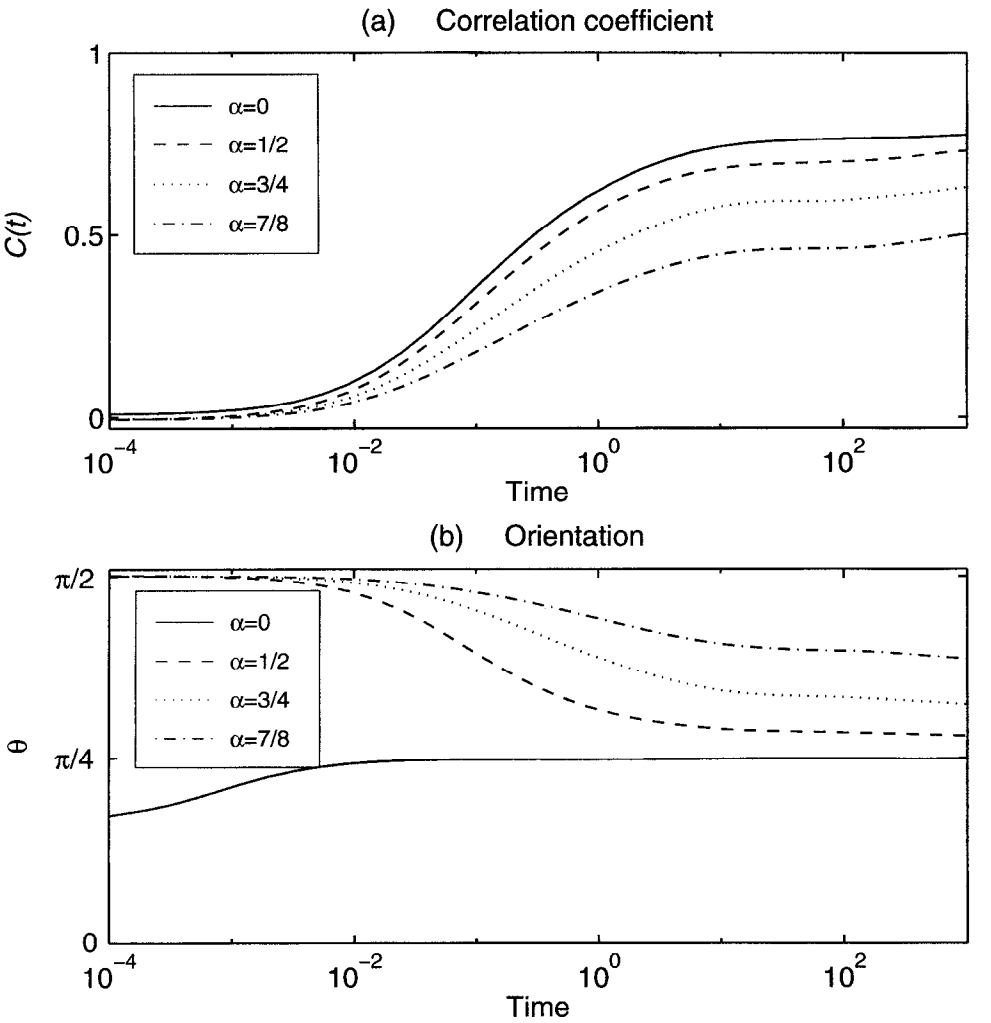

Figure 13. (a) Evolution of the correlation coefficient $C(t)$ defined in (4.17b) (b) Evolution of the orientation of the best-fit ellipse in the $T_x - S_x$ plane. The initial condition is the same as that of Figure 12.

where there are no such gradients) while buoyancy itself is diffused only where its gradient is large.

e. A statistical theory of thermohaline correlations. In this section we develop a statistical theory of the thermohaline correlations discussed in the previous section. A complete statistical description of $h(x, t)$ and $j(x, t)$ is contained in a hierarchy of equations for all

moments of the two fields, whose evolution is governed by

$$h_t = (h^3)_{xx}, \qquad j_t = (h^2 j)_{xx}. \tag{4.18a,b}$$

This hierarchy becomes much simpler if we assume that all our fields are homogeneous. Homogeneity means that the statistical properties of any function of the fields do not vary with absolute position. By choosing initial conditions with zero mean we can restrict our discussion to fields h and j with zero mean at all times; i.e.,

$$\langle h \rangle \equiv \frac{1}{L}\int_0^L h \, dx = 0, \qquad \langle j \rangle \equiv \frac{1}{L}\int_0^L j \, dx = 0. \tag{4.19a,b}$$

The equations for the second moments (the two-point correlation functions) are written by introducing $\tilde{h} \equiv h(x + \xi, t)$ and $\tilde{j} \equiv j(x + \xi, t)$. Under the assumption that both fields are homogeneous, so that the correlation functions depend only on the space lag ξ and the time t, it follows from (4.18a,b) that

$$\begin{aligned} \langle h\tilde{h} \rangle_t &= 2\langle h^3 \tilde{h} \rangle_{\xi\xi}, \\ \langle h\tilde{j} \rangle_t &= \langle h^3 \tilde{j} \rangle_{\xi\xi} + \langle h\tilde{h}^2 \tilde{j} \rangle_{\xi\xi}, \\ \langle j\tilde{j} \rangle_t &= 2\langle h^2 j\tilde{j} \rangle_{\xi\xi}. \end{aligned} \tag{4.20a,b,c}$$

These equations are not a closed set. In order to solve for the second-order moments we first have to solve a similar set of equations for the fourth-order moments, which in turn require the sixth-order moments and so on.

A straightforward way to overcome the closure problem is to make a quasi-normality hypothesis as discussed in the text by McComb (1990). This hypothesis is used to close the moment hierarchy by expressing the quadruple moments on the RHS of (4.20) in terms of products of second moments. If we denote the quadruple correlation symbolically by $\langle h_1 h_2 h_3 h_4 \rangle$ then for a Gaussian field we have

$$\langle h_1 h_2 h_3 h_4 \rangle = \langle h_1 h_2 \rangle \langle h_3 h_4 \rangle + \langle h_1 h_3 \rangle \langle h_2 h_4 \rangle + \langle h_1 h_4 \rangle \langle h_2 h_3 \rangle. \tag{4.21}$$

The application of (4.21) to each of the fourth-order moments on the RHS of (4.20) gives a closed system of three equations for the second order correlation functions

$$\begin{aligned} \langle h\tilde{h} \rangle_t &= 6\langle h^2 \rangle \langle h\tilde{h} \rangle_{\xi\xi}, \\ \langle h\tilde{j} \rangle_t &= 4\langle h^2 \rangle \langle h\tilde{j} \rangle_{\xi\xi} + 2\langle hj \rangle \langle h\tilde{h} \rangle_{\xi\xi}, \\ \langle j\tilde{j} \rangle_t &= 2\langle h^2 \rangle \langle j\tilde{j} \rangle_{\xi\xi} + 4\langle hj \rangle \langle h\tilde{j} \rangle_{\xi\xi}. \end{aligned} \tag{4.22a,b,c}$$

In order to solve the (4.22) we need appropriate initial conditions. The random initial distributions considered in the previous sections correspond to δ-function autocorrelations

for temperature and salinity and therefore initial δ'' autocorrelations for h and j

$$(\langle h\tilde{h}\rangle, \langle h\tilde{j}\rangle, \langle j\tilde{j}\rangle) = -(1, \alpha, 1)\delta''(\xi) \qquad \text{when} \qquad t = 0. \tag{4.23}$$

The equations above are a good approximation only if the relationship (4.21) is approximately true. The additional terms that appear in the RHS of (4.22) for non-Gaussian fields are called "cumulants." It is then more appropriate to refer to the closure approximation in (4.22) as the 'cumulant discard approximation.'

A comparison of the size of the diffusion coefficients in Eqs. (4.22) enables us to anticipate some differences in the evolution of h versus that of j. Let us consider first an initial condition with $\alpha = 0$. The correlation function $\langle h\tilde{j}\rangle$ remains zero at all times. The equations for $\langle h\tilde{h}\rangle$ and $\langle j\tilde{j}\rangle$ differ only because of the coefficient of the nonlinear diffusivity $\langle h^2\rangle$: it is three times bigger in (4.22a) than in (4.22c). The slower spreading of the spice gradient correlation function is a statistical indication of the spikes observed in j.

The solution of (4.22) and (4.23) is

$$\begin{aligned}
\langle h\tilde{h}\rangle &= \frac{1}{12\sqrt{3\pi\tau^3}}\left(1 - \frac{1}{6}\frac{\xi^2}{\tau}\right)\exp(-\xi^2/12\tau),\\
\langle h\tilde{j}\rangle &= \frac{\alpha}{12\sqrt{3\pi\tau^3}}\left(1 - \frac{1}{6}\frac{\xi^2}{\tau}\right)\exp(-\xi^2/12\tau),\\
\langle j\tilde{j}\rangle &= \frac{\alpha^2}{12\sqrt{3\pi\tau^3}}\left(1 - \frac{1}{6}\frac{\xi^2}{\tau}\right)\exp(-\xi^2/12\tau) + \frac{1-\alpha^2}{4\sqrt{\pi\tau^3}}\left(1 - \frac{1}{2}\frac{\xi^2}{\tau}\right)\exp(-\xi^2/4\tau),
\end{aligned} \tag{4.24a,b,c}$$

where $\tau \equiv (5t/12\sqrt{3\pi})^{2/5}$. The autocorrelation functions $\langle h\tilde{h}\rangle$ and $\langle j\tilde{j}\rangle$ in (4.24), derived from the cumulant discard approximation, pass an important consistency test; viz., the spectra (that is, the Fourier transforms of (4.24)) are positive definite. In the far more difficult problem of fully developed turbulence the cumulant discard approximation is obviously wrong because it results in autocorrelation functions with some negative coefficients in the Fourier transform.

We can now integrate the gradient correlation functions in (4.24) to obtain the correlation functions of buoyancy B and spice Q

$$\begin{aligned}
\langle B\tilde{B}\rangle &= \frac{1}{\sqrt{12\pi\tau}}\exp(-\xi^2/12\tau),\\
\langle B\tilde{Q}\rangle &= \frac{\alpha}{\sqrt{12\pi\tau}}\exp(-\xi^2/12\tau),\\
\langle Q\tilde{Q}\rangle &= \frac{\alpha^2}{\sqrt{12\pi\tau}}\exp(-\xi^2/12\tau) + \frac{1-\alpha^2}{\sqrt{4\pi\tau}}\exp(-\xi^2/4\tau).
\end{aligned} \tag{4.25a,b,c}$$

Figure 14 compares the correlation functions in (4.25) with simulation for $\alpha = \frac{1}{2}$; the

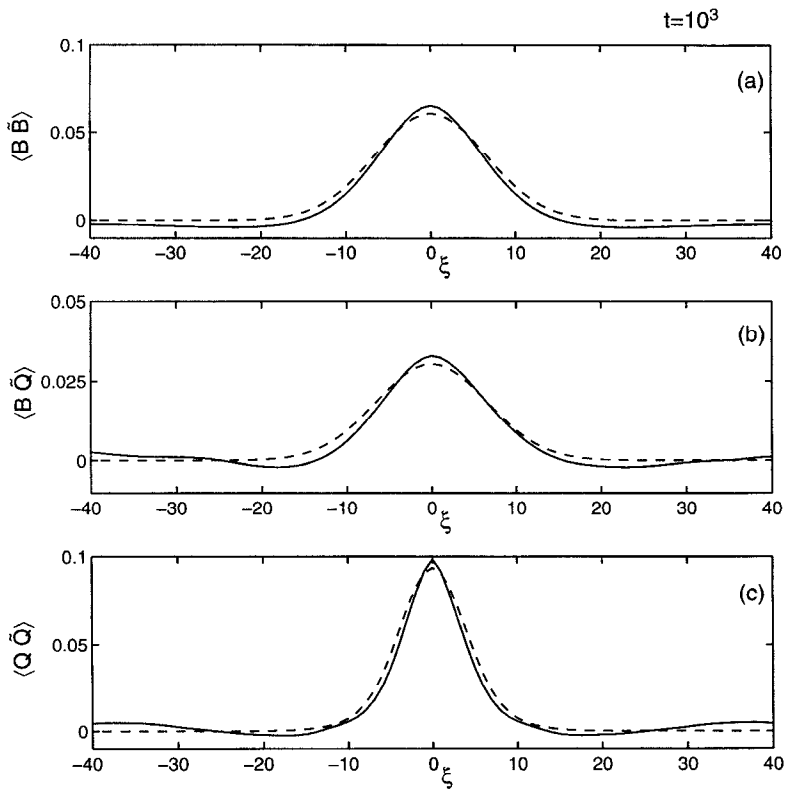

Figure 14. This figure shows the correlation functions of buoyancy B and spice Q. The dashed curves are the analytic expressions in (4.25) evaluated at $t = 1040$. The continuous curves are obtained by averaging the correlation functions obtained from 10 subsections of a simulation run with 10,000 points and $\alpha = \frac{1}{2}$ at $t = 10^3$.

agreement is satisfactory in all three cases. Figure 15 compares the gradient correlation functions in (4.24) with the same simulation. In the top two panels of Figure 15 there is reasonable agreement between theory and simulation (though not as good as that in Fig. 14). However in Figure 15c, which shows the autocorrelation function of the spice gradient, the results are less satisfactory. First, notice that the vertical scale of Figure 15c differs by a factor of 5 from that of Figures 14a and 14b. Thus, in both the theory and the simulation, the spice gradient autocorrelation is much larger than the buoyancy gradient. But the simulation shows a large 'pointy peak' in the $\langle j\tilde{j}\rangle$ correlation function at zero lag. The simulated $\langle j\tilde{j}\rangle$ correlation function is decaying more slowly than the theoretical estimate in (4.24c). The 'pointy peak' in Figure 15c is the signature of spikes in spice gradient. As time proceeds these spikes produce a progressive departure of the statistics of j from the assumed Gaussian distribution. This is shown in Figure 16 where we plot the kurtosis of buoyancy, spice and their gradients. We define the kurtosis of a field, f, with zero average as $\langle f^4\rangle/3\langle f^2\rangle^2$. For a Gaussian distribution this ratio is identically one. The kurtosis

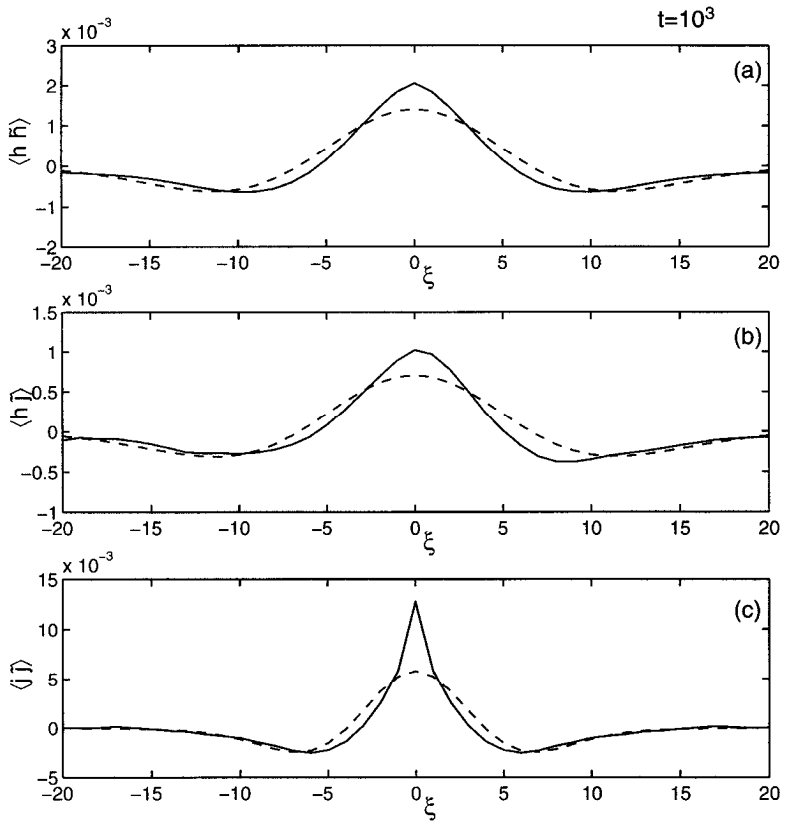

Figure 15. This figure shows the correlation functions of buoyancy gradient h and spice gradient j. The dashed curves are the analytic expressions in (4.24) evaluated at $t = 1040$. The continuous curves are obtained by averaging the correlation functions obtained from 10 subsections of a simulation run with 10,000 points and $\alpha = \frac{1}{2}$ at $t = 10^3$. Notice the different vertical axis scales in the various panels.

of j grows in time, because the spikes in j dominate the statistics of high order moments. For the same reason, the cumulants neglected in the RHS of (4.22c) grow and the cumulant discard approximation ceases to be valid.

Now consider the correltaion functions $R(t)$ and $C(t)$ in (4.17). Using the solutions in (4.24) and (4.25) (evaluated at zero lag) we obtain the cumulant discard approximations

$$R_{CD} = \frac{\sqrt{3} - 1}{\sqrt{4 + 2\sqrt{3}[(1 + \alpha^2)/(1 - \alpha^2)]}},$$
$$C_{CD} = \frac{3\sqrt{3} - 1}{\sqrt{28 + 6\sqrt{3}[(1 + \alpha^2)/(1 - \alpha^2)]}}. \qquad (4.26a,b)$$

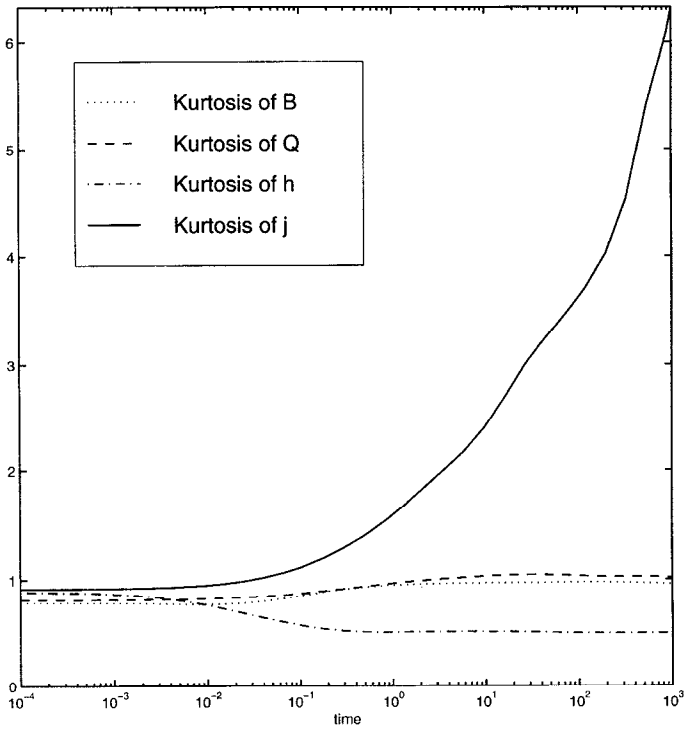

Figure 16. Evolution of the kurtosis of buoyancy B, spice Q, buoyancy gradient h and spice gradient j. The kurtoses of B and Q approach 1, showing that these two fields have a quasi-normal statistics as $t \to \infty$. The kurtosis of j increases indefinitely because extreme events (spikes) dominate the statistics as $t \to \infty$.

In Figure 17 we show a comparison between the approximations in (4.26) and the results of simulation. An encouraging aspect of the comparison is that the points from the simulations seem generally to follow the theoretical curve with a constant offset. Notice, however, that as t increases, the offset between theory and simulation in the bottom panel of Figure 17 increases slowly but systematically; this growing discrepancy is related to the slight rise in $C(t)$ which occurs in the final decade, $10^2 < t < 10^3$, of Figure 15a. Once again, this is the long term effect of the formation of spikes in the spice gradient.

5. Discussion and conclusion

The main result of this paper is that nonlinear diffusion parameterizations, which use the buoyancy gradient as the 'driving field,' produce strong density compensation between the thermal and haline components. One interesting aspect of this process in our model is that the compensation is stronger between the gradients, T_x and S_x, than between the fields T and S. This result can be used to test the model against oceanic mixed layer measurements. We expect that this gradient compensation is a general consequence of all nonlinear diffusion

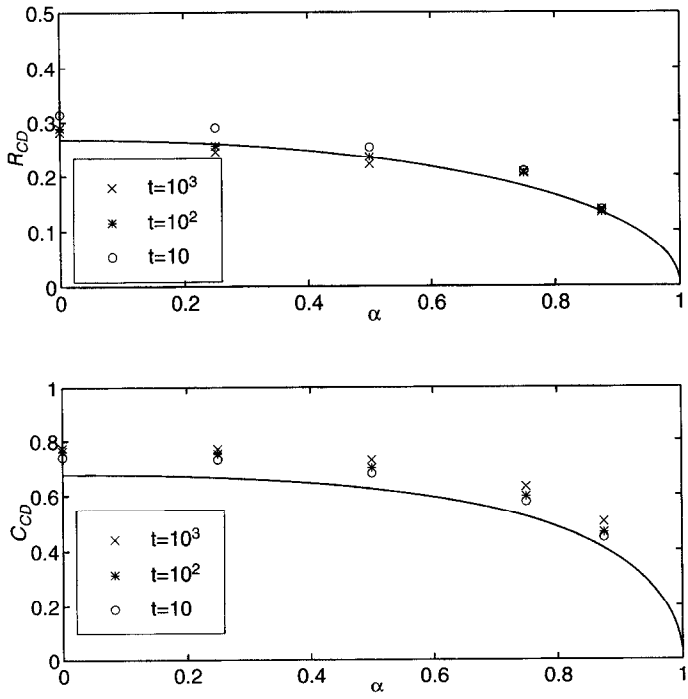

Figure 17. A comparison between the analytic expressions in (4.26) (the continuous curves) and the results of numerical simulation at three different times. All of the simulations have 10,000 points. The agreement between simulation and analysis is consistently better in the upper panel than in the lower panel. In the lower panel, as time increases, the results of the simulation move systematically away from the analytic expression. Notice also the different vertical axis scales between the upper and lower panels.

parameterizations in which the diffusivity increases with the horizontal buoyancy gradient. We also expect that the creation of small-scale structure in spice, or any passive combination of temperature and salinity, is in general property of this class of models.

Our model uses vertically averaged thermohaline fields and thus it is most appropriately applied to shallow systems in which strong vertical mixing manages to arrest the restratification which is driven by isopycnal slumping; the ocean mixed layer is an example. The development of models for the three-dimensional problem, and the analysis of systems with continuously acting thermohaline forcing are both obvious directions for future research.

The other aspect of our idealized model which should be emphasized is that in the initial condition both temperature and salinity fluctuations are created at the smallest resolved scale. On the other hand, in the mixed layer of the ocean, there is probably an interesting difference between the horizontal scales at which temperature and salinity variance are forced by meteorological inputs; it seems likely that salinity variance is forced by rainfall at smaller length-scale than temperature variance (e.g. Stommel, 1993). Models such as the

one formulated in this paper help us to think about the likely consequences of such statistical asymmetries.

Acknowledgments. RF acknowledges support from the National Science Foundation under award number OCE95-29752. WRY acknowledges support from the National Science Foundation under award number OCE96-16017. We thank D. L. Rudnick for comments on this work.

APPENDIX

We look for asymptotic self-similar solutions of the equation for the spice gradient $j(x, t)$

$$j_t = (h^2 j)_{xx}, \tag{A.1a}$$

for a prescribed buoyancy gradient profile $h(x, t)$ given by

$$h = C\sigma t^{-1/4}(2/\pi)\sqrt{1-\eta^2}\,\mathrm{H}(1-\eta^2), \qquad \eta \equiv \sigma x/t^{1/4}, \tag{A.2}$$

where $\mathrm{H}(x)$ is the Heaviside step function and $\sigma \equiv (\pi^2/48C^2)^{1/4}$. We impose a constant concentration $j(x, 0) \equiv j_0$ as initial condition. At the free boundaries the spice gradient is held constant $j(|x| \to \infty, t) = j_0$. We assume that at large times there is a self-similar solution $j \equiv j(\eta)$. In the interval $|\eta| < 1$ i.e., $|x| < t^{1/4}/\sigma$, the equation for j, in terms of η, is simply

$$(1-\eta^2)j_{\eta\eta} - \eta j_\eta - 2j = 0. \tag{A.3}$$

This equation has even solutions of the form

$$j(\eta) = J\frac{\cosh(\sqrt{2}\arcsin\eta)}{\cosh(\pi/\sqrt{2})}. \tag{A.4}$$

A plot of this function is given in the bottom panel of Figure 3 (in the present discussion $\gamma = 1$). The constant J is the value of j at $\eta = \pm 1$. It can be determined using the fact that the integral of j over the interval $-1 < \eta < 1$ is constant in time: it is twice the value of spice Q in $\eta = 1$, which is constant in time. The integral at time $t = 0$ is equal to $2j_0$ and yields $J = 3j_0$.

REFERENCES

Chen, L. and W. R. Young. 1995. Density compensated thermohaline gradients and diapycnal fluxes in the mixed layer. J. Phys. Oceanogr., *25,* 3064–3075.

Colin de Verdiere, J. G. Harvey and M. Arhan. 1986. Stirring and mixing of thermohaline anomalies. J. Mar. Res., *44,* 93–118.

Erdogan, M. E. and P. C. Chatwin. 1967. The effects of curvature and buoyancy on the laminar dispersion of solute in a horizontal tube. J. Fluid Mech., *29,* 465–484.

Flament, P., L. Armi and L. Washburn. 1985. The evolving structure of an upwelling filament. J. Geophys. Res., *90,* 11765–11778.

Godfrey, J. S. 1980. A numerical model of the James River Estuary, Virginia, U.S.A. Estuar. Coast. Mar. Sci., *11,* 295–310.

Landau, L. D. and E. M. Lifshitz. 1993. Fluid Mechanics, 2nd ed., Pergamon Press, 539 + xiii pp.
Linden, P. F. and J. E. Simpson. 1988. Modulated mixing and frontogenesis in shallow seas and estuaries. Cont. Shelf Res., *8,* 1107–1127.
McComb, W. D. 1990. The Physics of Fluid Turbulence, Clarendon Press, 572 pp.
Munk, W. H. 1981. Internal waves and small-scale processes, *in* Evolution of Physical Oceanography, B. A. Warren and C. Wunsch, eds., Chapter 9, MIT Press, Cambridge, Massachusetts.
Niiler, P. P. 1984. The three dimensional circulation near the eastern North-Pacific Subtropical Front. J. Phys. Oceanogr., *14,* 217–230.
Pavan, V. and I. Held. 1996. The diffusive approximation for eddy fluxes in baroclinically unstable jets. J. Atmos. Sci., *53,* 1262–1272.
Roden, G. I. 1977. Oceanic subarctic fronts of the central Pacific: Structure of and response to atmospheric forcing. J. Phys. Oceanogr., *7,* 761–778.
—— 1980. On the subtropical frontal zone north of Hawaii during winter. J. Phys. Oceanogr., *10,* 342–362.
—— 1986. Thermohaline fronts and baroclinic flows in the Argentine Basin during the austral Spring of 1984. J. Geophys. Res., *91,* 5075–5093.
Roemmich, D., M. Y. Morris, W. R. Young and J. R. Donguy. 1994. Fresh equatorial jets. J. Phys. Oceanogr., *24,* 540–558.
Rudnick, D. L. 1996. Intensive surveys of the Azores Front, 1, Tracers and dynamics. J. Geophys. Res., *101,* 923–939.
Samelson, R. M. and C. A. Paulson. 1988. Towed thermistor chain observations of fronts in the Subtropical North Pacific. J. Geophys. Res., *93,* 2237–2246.
Simpson, J. E. and P. F. Linden. 1989. Frontogenesis in a fluid with horizontal density gradients. J. Fluid Mech., *202,* 1–16.
Stommel, H. 1961. Thermohaline convection with two stable regimes of flow. Tellus, *13*(2), 224–230.
—— 1962. On the cause of the temperature-salinity curve in the ocean. Proc. Natl. Acad. Sci., *48,* 764–766.
—— 1993. A conjectural mechanism for determining the thermohaline structure of the oceanic mixed layer. J. Phys. Oceanogr., *23,* 142–148.
Talley, L. D. 1998. Temperature, salinity and buoyancy restratification in the mixed layer. J. Mar. Res., (submitted).
Tandon, A. and C. J. R. Garrett. 1994. Mixed layer restratification due to a horizontal density gradient. J. Phys. Oceanogr., *24,* 1419–1424.
Taylor, G. I. 1953. Dispersion of soluble matter in solvent flowing slowly down a tube. Proc. Roy. Soc. London Ser. A., *219,* 186–203.
Veronis, G. 1972. On properties of seawater defined by temperature, salinity and pressure. J. Mar. Res., *30,* 227–255.
Visbeck, M., J. Marshall, T. Haine and M. Spall. 1997. Specification of eddy transport coefficients in coarse-resolution ocean circulation models. J. Phys. Oceanogr., *27,* 381–402.
Welander, P. 1971. The thermocline problem. Phil. Trans. Roy. Soc. Lond. A., *270,* 415–421.
—— 1989. A new type of double-diffusive instability? Tellus, *41A,* 66–72.
Young, W. R. 1994. The subinertial mixed layer approximation. J. Phys. Oceanogr., *24,* 1812–1826.
Young, W. R. and S. Jones. 1991. Shear dispersion. Phys. Fluids, *A3,* 1087–1101.
Yuan, X. and L. D. Talley. 1992. Shallow salinity minimum in the North Pacific. J. Phys. Oceanogr., *22,* 1302–1316.

Received: *24 July 1997;* revised: *23 September 1997.*