Diffusion versus Nonlocal Models of Stratospheric Mixing, in Theory and Practice

ADAM H. SOBEL

University of Washington, Seattle, Washington

(Manuscript received 19 March 1998, in final form 21 August 1998)

ABSTRACT

In zonally averaged chemical transport models of the stratosphere, quasi-isentropic mixing is represented by diffusion in latitude. However, it is fairly certain that the real mixing is to some extent nonlocal, so that the diffusive representation is not formally justifiable. This issue is explored from a point of view that combines theory and empiricism. Several models of mixing are described and compared. The most general, including as special cases all of the other models considered, is the integral or ''transilient matrix'' model. Some known properties of transilient matrices are discussed in a more formal way than has been done previously, and some new results concerning these matrices and associated equations are derived. Simpler models include the familiar diffusion model, a simple model of nonlocal mixing in which tracer concentrations are everywhere relaxed toward the global average, and a ''leaky barrier'' model in which two regions of nonlocal mixing are separated by a weakly diffusive transport barrier. Solutions to the latter two models, with linear chemistry included to allow nontrivial steady states, are used to derive their ''effective diffusivities.'' These are then used to test how well a diffusion model can mimic the behavior of the nonlocal mixing models over finite regions of parameter space. The diffusion model proves fairly robust, yielding fairly accurate results in situations where no formal argument indicates that it should. These results provide qualified support, from a purely empirical perspective, for the practice of using the diffusion model to represent stratospheric mixing in zonally averaged models. Several important caveats suggest nonetheless that exploration of more theoretically satisfactory representations is warranted.

1. Introduction

The stratosphere is known to be strongly influenced in a number of respects by quasi-isentropic mixing induced by what has come to be known as planetary wave breaking (McIntyre and Palmer 1983, 1984, 1985). In zonally averaged chemical transport models, this mixing is generally represented by a horizontal diffusion term, with diffusivity K_{yy}. This would be theoretically defensible if the stratospheric flow had the character either of homogeneous turbulence (Taylor 1921) or steady, small-amplitude waves (Plumb 1979). Rather than either of these, however, transport by breaking stratospheric planetary waves is better described as chaotic advection [e.g., Pierrehumbert 1991a,b; Pierce and Fairlie 1993; Ngan and Shepherd 1999a,b; on the idea of chaotic advection itself see, e.g., Hama (1962), Aref (1984), Ottino (1989, 1990), and Wiggins (1992)]. In this process, material is transported irreversibly over long distances in coherent, large-scale circulations. Thus it has been argued that diffusion is, in principle, a poor representation of this process (e.g., McIntyre 1991; Pierrehumbert 1991a,b).

On the other hand, one may argue that the lack of theoretical basis for the diffusive representation does not imply that its use must lead to poor model performance, if K_{yy} is allowed to be a function of space and time and is chosen wisely. Even this empirical argument is unsupportable in certain cases. Clear examples are flows in the convective planetary boundary layer (PBL), since these regularly produce upgradient fluxes. In the stratosphere, however, isentropic eddy fluxes of most tracers of interest tend to be downgradient, so the situation is more subtle. One may be aware of the lack of theoretical support for the use of the diffusion model, but be led to use it nonetheless by an empirical approach or a desire for model simplicity. This approach is taken in a number of studies that have estimated K_{yy} from observed or simulated flux–gradient relationships (Newman et al. 1986; Plumb and Mahlman 1987; Newman 1988; Yang et al. 1990).

This paper contains a discussion of this and related issues. The properties and behavior of several different representations of mixing are examined, both from a general theoretical perspective and using specific idealized model examples. The emphasis is on mixing in bounded domains, and on situations in which the largest eddies are comparable in size to the domain itself.

Corresponding author address: Dr. Adam H. Sobel, Department of Atmospheric Sciences, University of Washington, Box 351640, Seattle, WA 98195-1640.
E-mail: sobel@atmos.washington.edu

Section 2 defines the problem to be addressed. Section 3 discusses the conditions under which a diffusion model is a theoretically valid description of tracer transport by fluid motions. Section 4 discusses the integral or ''transilient matrix'' representation, reviewing and augmenting the original work on this subject (Fiedler 1984; Stull 1984; Stull 1993, and references therein). This model is not explicitly used in this study but is valuable as a theoretical backdrop. The other models discussed in later sections can all be viewed as special cases of the transilient matrix; seeing what shared properties are implied by this brings the differences into clearer focus. Section 5a discusses a very simple model of mixing by large eddies, originally due to Budyko (1969), in which tracers are everywhere linearly relaxed toward their domain-wide averages. In section 6, the extent to which the behavior of this model can be mimicked by a diffusion model is explored, if the diffusivity is treated as an empirical parameter and a function of position. A model that bears greater qualitative resemblance to the stratospheric problem, in that it includes a somewhat permeable ''transport barrier,'' is introduced, and the same analysis is repeated for that model.

In the cases studied, the tracer fluxes are uniformly downgradient, so that a positive, finite ''effective diffusivity'' can be defined everywhere simply by the flux/gradient ratio. Because the mixing is not truly diffusive, this ratio is influenced by processes other than the mixing, in particular the nonconservative process that opposes the mixing. This is also the case in the dynamical theory that comes closest to defending the use of the diffusion model for stratospheric flows (Plumb 1979). Different species generally have different diffusivities, even in the small-amplitude case treated by Plumb. We examine how well the diffusion model, with the diffusivity derived using the flux/gradient ratio for one tracer, reproduces the solution for a second tracer subject to the same actual mixing but a different nonconservative process.

We do not address difficulties that are inherent to the zonally averaged framework itself. Some forms of zonal inhomogeneity may be inherently unparameterizable, and consequently some scientific problems may be inaccessible to zonally averaged models. We focus only on the behavior of different models of meridional mixing, given the inherent limitations incurred by averaging over the third spatial dimension. Formally, we assume that the mixing may be represented by a transilient matrix. As explained in section 4 and appendix A, the transilient matrix representation is local in time (a Markov process) and includes only the zonal mean, as opposed to any explicit information on deviations from the zonal mean. The diffusion model can be viewed as a special case of the transilient matrix, and so implies all the same assumptions, but also an additional one, that the mixing is local. We focus only on the differences in model behavior that may result from relaxing the assumption of locality.

2. Definition of the problem

Consider a two-dimensional, incompressible fluid flow $\mathbf{u}(\mathbf{x}, t)$, where $\mathbf{x} = (x, y)$ and $\mathbf{u} = (u, v) = (dx/dt, dy/dt)$. The fluid carries a tracer $q(\mathbf{x}, t)$:

$$\frac{\partial q}{\partial t} + \mathbf{u} \cdot \boldsymbol{\nabla} q = S, \tag{1}$$

where S represents any sources or sinks of the tracer. Equation (1) will be considered here from a purely *kinematic* point of view. That is, $\mathbf{u}(\mathbf{x}, t)$ will be taken to be given. Taking a Reynolds average over x, (1) becomes

$$\frac{\partial \overline{q}}{\partial t} + \overline{v}\frac{\partial \overline{q}}{\partial y} + \frac{\partial}{\partial y}\overline{v'q'} = \overline{S}, \tag{2}$$

where overbars represent the means and primes represent deviations. If the flow takes place in a bounded domain, its two-dimensionality implies that the second term on the lhs vanishes. The third term on the lhs of (2) will be called the eddy flux divergence or kinematic flux divergence.

The problem to be discussed is how to represent the eddy flux divergence by some form of operator on the mean tracer field $\overline{q}(y)$. If the operator acting at a given y depends not only on $\overline{q}(y)$ and its derivatives evaluated at that point, but at other points as well (e.g., through an integral), we say the operator represents *nonlocal* mixing. This is a familiar idea in boundary layer meteorology; for a review of nonlocal mixing models, see Stull (1993).

If the flow $\mathbf{u}$ were not given, the problem would be simply the classic closure problem in the study of turbulence. As a consequence of the kinematic approach, however, the problem confronted here is a much narrower and more straightforward one. The essential issue here is the choice of the form of the operator, or *transport model* (not to be confused with the larger model in which it may be embedded) rather than the values of any *parameters* that may need to be determined once the choice has been made. Closure would require specifying both, using assumptions about the flow dynamics. Because $\mathbf{u}$ is assumed explicitly known, it can be used directly to evaluate the parameters. As an example of closure, in a PBL problem involving both shear and buoyancy effects (e.g., Stull 1988; Garratt 1992) one might choose the diffusion model (often called ''K theory''), and then specify the diffusivity (the relevant parameter in this model) as a function of a Richardson number to close the problem. These two steps can be viewed as distinct; here we focus only on the first.

3. Limitations of the diffusion model

Following the original work of Taylor (1921), Corrsin (1974) derives conditions that must be met in order for a diffusion equation to be a valid description of the statistical behavior of a stochastic model of particle displacements. Two are particularly important here, since they appear to be strongly violated in the stratospheric case (e.g., McIntyre 1991). These are that ‘‘the transport mechanism length scale must be much smaller than the distance over which the curvature of the mean transported field gradient changes appreciably’’ and ‘‘the transport mechanism length scale must be essentially constant over a distance of a length scale (merely for the concept to be meaningful) and over a distance for which the mean transported field changes appreciably.’’ Note that if the ‘‘transport mechanism length scale’’ is small compared to all other scales in the problem then both requirements will be met. Neither requirement will be met in flows that are strongly *inhomogeneous*; this is close to a definition of that term (e.g., Haidvogel and Held 1980).

That results from such stochastic models are relevant to transport by fully turbulent flows follows from the general idealization of turbulent motions as effectively random. One may legitimately ask whether these results are applicable to planetary-scale, quasi-isentropic stratospheric mixing, which has a character somewhere between transport by waves and by layerwise two-dimensional turbulence. The results of Plumb (1979) are instructive in this regard. Plumb showed that under certain circumstances, small-amplitude waves on a zonally symmetric basic state can lead to meridional diffusion of tracers. The resulting local flux–gradient relationship in these circumstances is a direct consequence of the smallness of the meridional parcel displacements, which follows from the assumption that the waves have small amplitudes. Hence the wave and turbulence theories share essentially the same limitation. This is that there is a key length scale of the fluid motions that must be taken infinitesimally small. The stratosphere appears unlikely to satisfy this condition, regardless of the way in which the length scale is defined.

4. Integral or matrix models

a. Introduction

Here and for the rest of this section, we assume q is an absolutely conserved tracer ($S = 0$); this will be relaxed later. Furthermore, we assume that there is no flux of tracer through the boundaries in y. In x the boundary condition may either be periodic, or impermeable, no-flux walls as in y.

In any 2D incompressible flow, we can write [Batchelor 1949; Roberts 1961; see also section 2 in Thuburn and McIntyre (1997)]

$$q(x, y, t + \Delta t) = \iint A(x, y, t + \Delta t \,|\, x', y', t) q(x', y', t)\, dx'\, dy'. \quad (3)$$

The operator A, a four-dimensional object, maps the entire tracer field at time t to itself at time $t + \Delta t$ and depends only on the flow field $\mathbf{u}$ (though it can be redefined to include explicit diffusion). We can think of A as defining all the parcel displacements that occur over the time between t and $t + \Delta t$ and that begin at locations (x', y'). Equation (3) is linear if q is a passive tracer, and entails no approximation.

While still recognizing that the flow is 2D, we may extend the idea embodied by (3) to the *mean* profile $\overline{q}(y)$, to produce a reduced-dimensional transport model (Fiedler 1984; Stull 1984, 1993):

$$\overline{q}(t + \Delta t) = \int_{y_b}^{y_t} C(y, t + \Delta t \,|\, y', t) \overline{q}(y', t)\, dy' \quad (4)$$

so that $C(y, t + \Delta t \,|\, y', t)$ is now two-dimensional and maps the mean tracer distribution at t to itself at $t + \Delta t$. Here y_b and y_t are the domain boundaries. Unlike (3), (4) is not an exact representation, as discussed below and in appendix A. Discretizing (4) in space yields

$$\mathbf{q}(t + \Delta t) = \mathsf{C}\mathbf{q}(t), \quad (5)$$

where $\mathbf{q}$ is an n-dimensional vector whose elements are the values of $\overline{q}$ at the nodes of a one-dimensional lattice representing the mean profile,

$$\mathbf{q} = (\overline{q}(y_b), \overline{q}(y_b + \Delta y), \overline{q}(y_b + 2\Delta y), \ldots, \overline{q}(y_b + N\Delta y)),$$

and $y_b + N\,\Delta y = y_t$. Here C is an $N \times N$ matrix, which Stull (1984) called the transilient matrix. Because in the applications of interest time and space are both invariably discretized, this form will be discussed hereafter. Unlike the diffusion model, (4) and (5) place no restrictions on the length scale of the transporting eddies, other than those imposed by the domain size, and the spatial discretization in the case of (5).

The formulations (4) and (5), since they apply to $\overline{q}$ rather than q, are *not* exact, in general, but are approximations. The matrix C can be thought of as a matrix of probabilities describing how likely a fluid parcel at any given level is to be transported to any other given level during an interval of time Δt, if it is assumed that each parcel carries a tracer concentration equivalent to the *mean* at its level of origin. This last assumption is the cause of the inexactness. Ebert et al. (1989) called this effect ‘‘convective structure memory,’’ and provided a heuristic description of its cause. The practical implication of this effect is that the transilient matrix formulation is not robust to large variations in the time step Δt, a free parameter. These issues are discussed further in appendix A, where (5) is derived directly from a discretized version of (3), and convective structure memory is seen to arise through neglect of certain terms. It also makes clear that, despite its approximate nature, the transilient matrix is the most general transport model that is local in time and that incorporates information about only the mean tracer profile. Rather than starting from first principles such as are embodied by (3), the derivation of Fiedler (1984) used the ‘‘spectral diffusivity’’ parameterization of Bercowicz and Prahm

(1979), while Stull hypothesized (5) directly on intuitive physical grounds. The independent work of Larson (1999) provides a derivation that is more detailed and general than the one given here, including a more thorough treatment of boundary conditions. Above and *for the remainder of this paper, we assume no-flux boundary conditions.*

Prather (1996) describes modes that are eigenvectors of a matrix that includes both transport and linearized chemistry. The transilient matrix, which includes transport only, is a special case of Prather's matrix.

b. Mathematical properties

All other transport models discussed in this paper, once discretized, are special cases of the transilient matrix, and the ''chemical transport'' models discussed in section 6 are likewise special cases of the associated equation [(9)]. A brief discussion of the properties of transilient matrices serves to highlight the properties that these models (as well as many other possible ones) have in common.

We begin by reviewing some known properties of transilient matrices (Stull 1984, 1993). First, when the matrix $\mathbf{C}$ is known (as it is in the kinematic approach), (5) is a linear equation. Second, the form of this equation shows that the tracer distribution at each time step depends only on the tracer distribution at the previous time step and not on the distribution at any earlier times. Mathematically, (5) describes a Markov process; or we might say that it is local in time. Third, since no negative tracer mixing ratios can develop from an initial distribution that is all positive, all elements must be nonnegative. Fourth, by conservation of fluid mass and of the tracer, the rows and columns of $\mathbf{C}$ must all sum to unity:

$$\sum_j C_{jj'} = \sum_{j'} C_{jj'} = 1,$$

where $C_{jj'}$ is the element of $\mathbf{C}$ denoting the amount of fluid from bin j' that is transported to bin j. Formally, $\mathbf{C}$ is doubly stochastic.

A fifth property involves the flux due to mixing by a transilient matrix. Since there is no flux through the boundaries of the domain, the flux at a point whose index j is equal to some particular value k may be written as (Ebert et al. 1989)

$$F_k = \frac{\Delta y}{\Delta t}\sum_{j=1}^{k}\sum_{j'=k+1}^{N}(C_{j'j}\overline{q}_j - C_{jj'}\overline{q}_{j'}), \tag{6}$$

where the sign convention is such that a positive flux is in the direction of increasing j. In the case of a symmetric transilient matrix, (6) reduces to

$$F_k = \frac{\Delta y}{\Delta t}\sum_{j=1}^{k}\sum_{j'=k+1}^{N}C_{jj'}(\overline{q}_j - \overline{q}_{j'}). \tag{7}$$

Remember that all the matrix elements $C_{jj'}$ are nonnegative, and notice in the above equation that the index j' of the second sum is always greater than the index j of the first. Hence, this equation shows that for mixing by a symmetric transilient matrix, *if the gradient of the tracer is everywhere of one sign, the flux is everywhere downgradient* no matter what the other details of the transilient matrix. This is fairly directly implied by (6) but was not noted by Ebert et al. (1989). Equation (5) is not, in general, equivalent to a diffusion equation, but is nonlocal. This reinforces our knowledge that the presence of downgradient flux is not generally indicative of a causal relationship between the flux and the local gradient. If the tracer gradient changes sign, then (7) can yield upgradient fluxes, though this condition is only necessary, not sufficient. Whether upgradient fluxes occur depends on the details of both the tracer profile and the transilient matrix.

We now state some additional properties of transilient matrices in the formal language of stochastic processes. Necessary definitions and background results can be found in Taylor and Karlin (1994). It is first useful, and physically plausible, to make two further assumptions about all transilient matrices that represent real fluid flows. The first is that they are *irreducible*, meaning that there are no absolutely impermeable transport barriers. In other words, given a sufficiently long time, each region of the fluid *communicates* with each other region. The second is that there is no unmixing (i.e., the total variance of the tracer profile cannot increase due to mixing by the transilient matrix), so that the process is *aperiodic*. Together, these imply that the matrices are *regular*, meaning that for sufficiently large n, all elements of the matrix $\mathbf{C}^n$ are strictly positive. It follows immediately from this and the double stochasticity of $\mathbf{C}$ that the *limiting distribution* of the process defined by $\mathbf{C}$ is the uniform distribution, that is, a vector with all elements equal. This implies that each eigenvalue λ_k of $\mathbf{C}$ satisfies

$$0 \le \lambda_k < 1 \tag{8}$$

except the one corresponding to the uniform distribution, for which $\lambda_k = 1$. This formalizes the intuitive notion that mixing, acting in the absence of other processes, must lead eventually to a state in which all material properties are homogenized.

c. Interaction of large eddy transport and linear nonconservative processes

We desire models that can be used to study the interaction of nonconservative processes with nonlocal mixing. The transilient matrix equation [(5)] may be augmented by inclusion of a term representing the effect of photochemistry on a chemical tracer, or radiation on a dynamical tracer, such as potential vorticity or temperature, which is affected by diabatic heating. The simplest form of such a term that still allows some interesting dynamics is that of a Newtonian relaxation. That is, we can study the system

$$\mathbf{q}(t + \Delta t) = \mathbf{C}(t, \Delta t)\mathbf{q}(t) + \frac{\Delta t}{\tau_c}(\mathbf{q}_c - \mathbf{q}(t)), \qquad (9)$$

where $\mathbf{C}$ is the transilient matrix, $\mathbf{q}_c$ is a given profile representing the equilibrium tracer distribution in the absence of transport, and τ_c is a timescale for the nonconservative process. Hereafter this process will be referred to as chemistry for brevity. If it does represent chemistry, the linearization is justified for sufficiently small perturbations from an equilibrium state, as in Prather (1994, 1996). The factor Δt in the second term on the rhs enters because the transilient matrix represents an amount of mixing occurring in a finite time Δt. Equation (9) is derived in appendix B. As made clear in that derivation, this equation is valid only if Δt is much smaller than both τ_c and τ_m. The latter is the ''mixing timescale,'' defined more carefully below; it may be thought of as the characteristic time over which the domain approaches homogeneity under the action of mixing alone.

It is interesting to consider the steady-state case of (9). In this case we have

$$\mathbf{C}\mathbf{q} - \left(1 + \frac{\Delta t}{\tau_c}\right)\mathbf{q} = -\frac{\Delta t}{\tau_e}\mathbf{q}_c. \qquad (10)$$

The continuous version of (10) may be written in the form

$$\int_0^{N\Delta y} C(y|y')\overline{q}(y')\, dy' - \lambda\overline{q}(y) = f(y), \qquad (11)$$

which is a Fredholm equation of the second kind. The reduction from continuous to discrete form raises mathematical issues in principle, but it will be assumed here that all functions and operators are smooth enough that the two are interchangeable. In practice, issues of resolution must be treated with care. It is not guaranteed a priori that equations of this type have a unique nonzero solution. The solvability of (10), however, follows from (8), as shown in appendix C.

5. Relaxation to the average

In this section, we describe a simple though nonlocal model, which we call ''relaxation to the average'' (RA). In section 5a this model is derived as an approximation, valid in the long time limit, to mixing by *any* transilient matrix. In section 5b a derivation from a highly idealized kinematic flow field indicates the sort of flow for which this model is a good approximation at relatively short times as well. It was originally introduced by Budyko (1969) to represent heat transport in a one-dimensional energy balance climate model, and has more recently been used by Raymond (1994) to represent mixing by deep tropical convection. The transport length scale associated with this model is simply the domain size, as illustrated in section 5b.

a. Derivation from transilient matrix model

Rather than using the transilient matrix as in (5), we can use the explicit form, that is,

$$\Delta\mathbf{q} = \mathbf{q}(t + \Delta t) - \mathbf{q}(t) = (\mathbf{C} - \mathbf{I})\mathbf{q}, \qquad (12)$$

with $\mathbf{I}$ the identity matrix. We also redefine $\mathbf{q}$ by subtracting the global mean from it. This incurs no loss of generality, since the transilient matrix has no effect on the global mean. The eigenvectors of $\mathbf{C} - \mathbf{I}$ are the same as those of $\mathbf{C}$, but the eigenvalues are reduced by 1, that is,

$$\Delta a_k = a_k(t + \Delta t) - a_k(t) = \epsilon_k a_k,$$

where $\epsilon_k = \lambda_k - 1$; a_k is the coefficient of the eigenvector $\mathbf{q}_k$ in the expansion of the tracer profile,

$$\mathbf{q} = \sum_k a_k\mathbf{q}_k; \qquad (13)$$

and $-\epsilon_k$ is the inverse of the decay timescale for each eigenvector, scaled by Δt.

If we retain only one eigenvector, whose eigenvalue $\epsilon_k = \epsilon$, in the expansion of $\overline{q}$, we have

$$\Delta\mathbf{q} = \epsilon\mathbf{q}.$$

Since the global mean has been removed in the above discussion, the above is nothing more than a discrete version of

$$\frac{\partial\overline{q}}{\partial t} \approx \tau_m^{-1}(\hat{q} - \overline{q}), \qquad (14)$$

where

$$\hat{q} = \frac{\displaystyle\int_0^L \overline{q}(y)\, dy}{\displaystyle\int_0^L dy},$$

with L the size of the domain. That is, (14) represents relaxation of the zonal-mean tracer field everywhere toward its global average, with a single timescale τ_m, where

$$\tau_m \equiv \lim_{\Delta t\to 0} \frac{\Delta t}{\epsilon}.$$

The limiting process represents the choice to view the system only on timescales long compared to the eddy turnover time t^*, analogously to Taylor's (1921) derivation of the eddy diffusion model. Equation (14) is what we mean by the RA model.

The transilient matrix corresponding to this model has each element along the diagonal equal to $1 - (1 - 1/N)\Delta t/\tau_m$ while every off-diagonal matrix element is equal to $\Delta t/N\tau_m$. Even if this is not the true transilient matrix of the flow being studied, one can choose τ_m so as to represent the decay rate of the longest-lived components of the tracer structure; these will then nonethe-

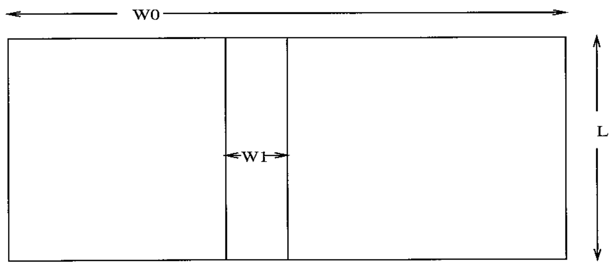

FIG. 1. Schematic for heuristic derivation of RA model.

less be represented accurately. One will then simply be underdamping the other eigenvectors. If the true transilient matrix is close to this form, the RA model will be a good approximation not only at long times, but at all times for which the transilient matrix formalism itself is valid.

It is worth noting that a diffusion model can be approximated by (14) as well as any other, in the sense that in the long time limit, if no nonconservative process opposes the mixing, the evolution of the tracer profile will be well described by that of the most slowly decaying eigenvector of the transilient matrix that the diffusion model describes (a tridiagonal one). This provides an indication that in some situations local and nonlocal mixing may be difficult (or even impossible) to distinguish.

The homogenization timescale τ_m need not be the same as the eddy turnover time t^*. The former is the timescale over which the entire domain approaches homogeneity due to mixing, while the latter is roughly the time for a parcel caught up in a large eddy to traverse the domain. The latter can be shorter than the former if the eddies are small in the unresolved dimension, that is, narrow, widely spaced plumes or filaments. Hence, it is not necessarily inconsistent to use the RA model to represent the transport of species that have sources and sinks whose timescales are comparable to τ_m. The derivation of the RA model below from an idealized kinematic flow makes this point clear, as well as illustrating in a heuristic way the sort of flow for which the RA model is a good approximation at all times long enough so that the transilient matrix formalism is valid.

b. Derivation from idealized kinematic flow

Consider a two-dimensional, rectangular domain, of length L in the y direction and width W_0 in the x direction. Some tracer is distributed throughout the domain in an x-independent basic state $\overline{q}(y)$.

Every time interval Δt, an ‘‘eddy’’ that spans the domain in y, entirely homogenizes a region of width W_1 (which is much smaller than W_0) and length L, causing net transport of tracers in the y dimension. This situation is sketched schematically in Fig. 1. Immediately thereafter, some strong mixing process that acts purely in the x direction immediately renders the fluid homogeneous again in the x direction. Hence,

$$\overline{q}(y, t + \Delta t) = \left(1 - \frac{W_1}{W_0}\right)\overline{q}(y, t) + \frac{W_1}{W_0}\hat{q}.$$

Now we rearrange, divide by Δt, and take the limit as Δt vanishes, but requiring that W_1 vanish in a directly proportional manner, that is,

$$\lim_{\Delta t \to 0} \frac{W_1}{W_0 \Delta t} \equiv \tau_m^{-1}. \tag{15}$$

This leads immediately to (14).

In the limit as τ vanishes, the approximation becomes the familiar ‘‘mixed layer’’ model, in which all properties are considered entirely uniform throughout the domain. This is a valid approximation when all other processes are very slow compared to the mixing.

Parenthetically, a more general version of the RA model may be constructed by allowing the ‘‘eddy width’’ W_1 to be a function of y. Steps analogous to those taken above then lead to

$$\frac{\partial \overline{q}}{\partial t} = \tau_m^{-1}(y)(\tilde{q} - \overline{q}),$$

where, at any given y, we have redefined

$$\lim_{\Delta t \to 0} \frac{W_1(y)}{W_0 \Delta t} = \tau_m^{-1}(y)$$

and

$$\tilde{q} = \frac{\displaystyle\int_0^L \tau_m^{-1}(y)\overline{q}(y, t)\, dy}{\displaystyle\int_0^L \tau_m^{-1}(y)\, dy}.$$

The properties of this generalization will not be explored further in this study.

6. Effective diffusivities computed from simple models of nonlocal mixing

Having a simple model of nonlocal mixing, we are now prepared to examine through simple, semianalytical examples the extent to which a diffusion equation can accurately model mixing that is actually nonlocal. The diffusivity is treated as an empirical function, computed from the ‘‘observed’’ flux/gradient ratio. This represents a best-case scenario in the sense that here, unlike in reality, the fluxes and gradients are known with no uncertainty.

In general, but not in the examples studied below, there are situations in which the diffusion model will obviously fail to mimic nonlocal mixing, no matter what the value of the diffusivity (assuming it to be positive and finite). One example is any situation in which the nonlocal mixing is such as to lead to upgradient fluxes. Likewise, in an initial value problem, we know that the diffusion model will cause an initial delta function in tracer concentration to spread continuously in the form of a Gaussian, or if the diffusivity is nonuniform, a convolution of Gaussians. Nonlocal mixing implies that the same delta function will spread into some other shape, in general arbitrary (beyond the constraints of mass conservation, etc.), and that it will acquire finite width in an infinitesimally small time, though ''infinitesimally small'' really means on the order of the eddy turnover time, which was assumed much smaller than some other timescale in the problem when the transport model was derived. In this case again, the diffusion model must fail to mimic the nonlocal mixing accurately.

More subtle are cases in which fluxes are everywhere downgradient and in which the system is in a nontrivial steady state. We focus below on two simple examples of this sort.

a. K derived from RA model

Consider the equation

$$\frac{\partial q}{\partial t} = \frac{\hat{q} - q}{\tau_m} + \frac{q_c - q}{\tau_c}, \tag{16}$$

where the overbar has been dropped so that q is now the mean profile $q(y)$. The domain is $0 < y < L$. Since the relaxation to the average is an integral operator, which in discrete form is a transilient matrix, this equation (again neglecting the distinction between continuous and discrete) is a special case of (9). The steady-state solution is

$$q = \tau_t\left(\frac{\hat{q}}{\tau_m} + \frac{q_c}{\tau_c}\right), \tag{17}$$

where

$$\tau_t \equiv \left(\frac{1}{\tau_m} + \frac{1}{\tau_c}\right)^{-1}.$$

Integrating (17) shows that $\hat{q} = \hat{q}_c$. This means that the interaction of mixing with linear chemistry cannot change the global tracer burden, although if τ_c were allowed to be a function of space this would no longer be true.

We can define the effective flux F due to the RA operator by noting that its divergence must balance the chemical source:

$$\frac{\partial F}{\partial y} = \frac{q_c - q}{\tau_c}. \tag{18}$$

Inserting the solution for q and integrating with the assumption of no flux through the boundaries yields

$$F = \frac{1 - \dfrac{\tau_t}{\tau_c}}{\tau_c}\int_0^y q_c(y')\,dy' - \frac{\tau_t\hat{q}_c y}{\tau_c\tau_m}. \tag{19}$$

Differentiating the solution for q yields

$$\frac{\partial q}{\partial y} = \frac{\tau_t}{\tau_c}\frac{\partial q_c}{\partial y}.$$

The negative of the flux divided by the gradient then defines the effective diffusivity, which is

$$K_e(y) = \frac{\hat{q}_c y - \displaystyle\int_0^y q_c(y')\,dy'}{\tau_m \dfrac{\partial q_c}{\partial y}}. \tag{20}$$

An interesting feature of (20) is that τ_c does not appear in it, so that species whose chemical equilibrium states q_c have the same shape, but whose chemical lifetimes are different, will have the same K_e. In this respect the diffusion model proves surprisingly robust.

However, the shape of q_c affects K_e, mitigating this. As an example, let $\tau_m = \tau_c = \tau$, and consider two different species, for which $q_c(y) = \alpha y$ and $q_c(y) = \beta y^2$ with α, β constants. The effective diffusivity for the first is

$$K_e(y) = \frac{yL - y^2}{2\tau} \tag{21}$$

while that for the second is

$$K_e(y) = \frac{L^2 - y^2}{2\tau}, \tag{22}$$

which obviously do not have the same functional form. In fact, (22) does not vanish at $y = 0$; since $\partial q/\partial y$ does vanish there for the tracer for which (22) was obtained, the solution still satisfies the no-flux boundary condition. However, it would not do so if the diffusivity obtained for the second tracer was used to transport the first.

b. Leaky barrier model

We now consider a model that, though still highly idealized, comes closer to representing the situation of interest in the stratosphere. The model contains regions of strong nonlocal mixing, separated from each other by a leaky transport barrier; hence, it will be referred to as the ''leaky barrier model.'' The nonlocal mixing regions may be thought of as, for example, the polar vortex interior and midlatitude surf zone, or midlatitude surf zone and Tropics, to a high degree of idealization.

This model is defined on a one-dimensional domain with impermeable boundaries at $y = L$ and $y = -L$. The governing equation is

$$\frac{\partial q}{\partial t} = \frac{\hat{q} - q}{\tau_m} + \frac{q_c - q}{\tau_c} + k\frac{\partial^2 q}{\partial y^2}, \tag{23}$$

with k a constant. Besides the addition of the diffusion term, (23) differs from (9) in that now we redefine

$$\hat{q} = \frac{1}{L}\int_{-L}^{0} q\,dy; \qquad -L < y < 0$$

$$\hat{q} = \frac{1}{L}\int_{0}^{L} q\,dy; \qquad 0 < y < L, \tag{24}$$

so that we have essentially two adjacent but nonoverlapping RA models, connected only by the diffusion term at the transport barrier, $y = 0$. The diffusivity k will be taken to be small so that its effect away from $y = 0$ is negligible compared to the nonlocal mixing. In cases where wave breaking strips material from the edges of a transport barrier of finite width but leaves an inner belt of strong tracer gradients ''unbroken,'' the transport across that belt will presumably be controlled by any small-scale diffusion or diffusion-like mixing process that is present (Flierl and Dewar 1985; Sobel and Plumb 1999). Hence, despite the simplicity and ad hoc nature of the present model, it does have some qualitative relation to quasi-isentropic stratospheric mixing, since in many situations this may be described by regions of fast, large-scale mixing (surf zones) separated by thin, weakly diffusive transport barriers.

In the case studied below we will set

$$q_c = \frac{q_0 y}{L}.$$

We will also consider only the steady-state behavior of (23).

Because of the exact antisymmetry of the problem, it is only necessary to consider half of the domain, which we choose to be $y > 0$. The antisymmetry also dictates that the first boundary condition on the solution is that $q = 0$ at $y = 0$. At $y = L$ we can allow no diffusive flux through the boundary, which implies the second boundary condition $\partial q/\partial y = 0$ at $y = L$.

With these boundary conditions, if $\hat{q}$ were known, (23) would have a particularly straightforward analytical solution. However, $\hat{q}$ is in fact part of the solution. Unlike in the previously considered model, it can differ from $\hat{q}_c$ since material may leak diffusively from one half-domain to the other. This makes matters slightly more complex, but an analytical solution is still possible. Strictly, (23) is an integro-differential equation [and again, a special case of (9) if the distinction between continuous and discrete is ignored]. However, it is a particularly simple one, which may be solved by taking $\hat{q}$ to be a constant, finding the solution in terms of that constant, and integrating the latter from $y = 0$ to $y = L$ to obtain an expression containing only $\hat{q}$ and known parameters of the problem, which then may be straightforwardly solved for $\hat{q}$. This may then be inserted back into the solution for q so that the latter is fully determined.

The result is

$$q = C_1 e^{-y/l} + C_2 e^{y/l} + C_3 y + \frac{\tau_t}{\tau_m}\hat{q}, \tag{25}$$

where we have defined $l = \sqrt{k\tau_t}$, and

$$C_1 = \left(\frac{l q_0 \tau_t}{L\tau_c} e^{-L/l} - \frac{\tau_t}{\tau_m}\hat{q}\right)(1 + e^{-2L/l})^{-1}$$

$$C_2 = -\left(C_1 + \frac{\tau_t}{\tau_m}\hat{q}\right)$$

$$C_3 = \frac{q_0 \tau_t}{L\tau_c}, \tag{26}$$

and the solution for $\hat{q}$, obtained as described above, is

$$\hat{q} = \frac{q_0\tau_t}{\tau_c}\frac{\frac{1}{2} + \left(\frac{l}{L}\right)^2[\mathrm{sech}(L/l) - 1]}{1 - \frac{\tau_t}{\tau_m} + \frac{l\tau_t}{L\tau_m}\left(\frac{2}{1 + e^{-2L/l}} - 1\right)}.$$

In the solutions to be presented, we have nondimensionalized such that $L = q_0 = \tau_m = 1$. The most relevant cases for our purpose are those for which τ_c is either larger (possibly much larger) than or comparable to τ_m, corresponding to either long-lived or ''intermediate-timescale'' tracers. Figure 2 shows solutions for $\tau_c = 1$, 10, and 100, with $k = 10^{-3}$. The quantities displayed are q, the flux F, gradient $\partial q/\partial y$, and resultant effective diffusivity K_e. Note in all solutions for q the boundary layer of width l near $y = 0$. In this region, the actual diffusion is important due to the large gradient. As might be expected, for $\tau_c = 1$, q retains a significant slope in the interior, while for the other cases it becomes quite nearly flat. Also, the larger τ_c, the lower the total burden in $y > 0$ (i.e., $L\hat{q}$), as the chemical source then has greater difficulty compensating for the rapid transport by the nonlocal mixing of tracer toward $y = 0$ where it can be diffused into the region $y < 0$.

Examining the effective diffusivities, one notices that to a reasonable degree of approximation they share a common shape. They do, however, differ in their subtle variations on that shape (particularly between the $\tau_c = 1$ case and the others) and in magnitude in the interior, unlike the simpler model considered above in which K_e was independent of τ_c.

Figure 3 illustrates the dependence of $\hat{q}$ on τ_c and k. As might be expected, for either of these parameters sufficiently large the solution reduces to $q = \hat{q} = 0$, while if both are sufficiently small we have $\hat{q} = ½$.

Next we consider the predictive power of the empirically derived effective diffusivities. That is, we examine solutions to the steady-state equation

$$\frac{q_c - q}{\tau_c} + \frac{\partial}{\partial y}\left(K_e\frac{\partial q}{\partial y}\right) = 0. \tag{27}$$

One assumes, and it is verified below, when K_e is derived for a particular value of τ_c, and then (27) is solved using that K_e and the same value of τ_c, that the resulting

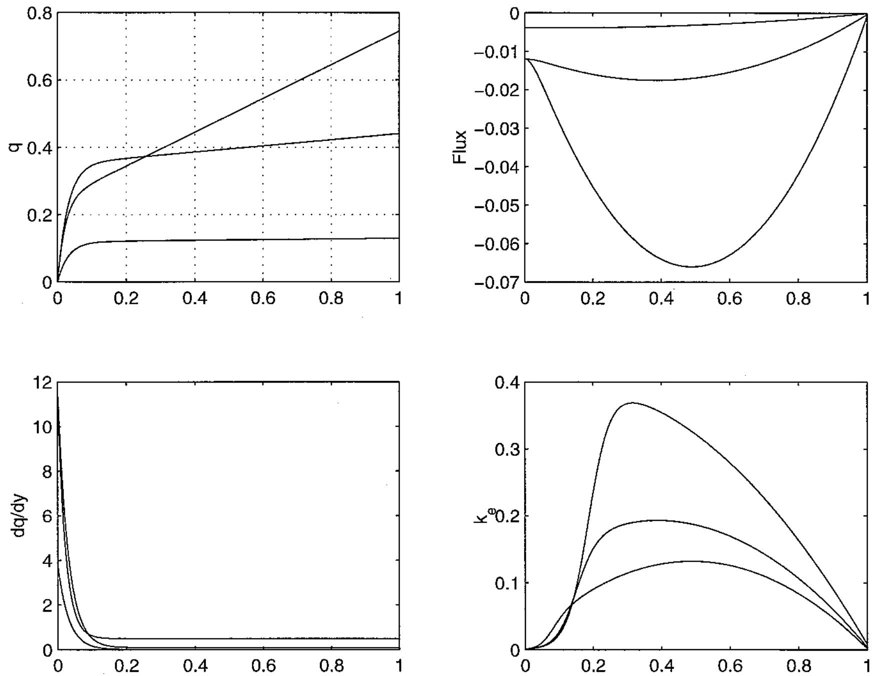

FIG. 2. Solutions of leaky barrier model, for $\tau_c = 1$, 10, and 100. (upper left) Tracer q (ordinate) vs y (abscissa); (upper right) flux F; (lower left) gradient $\partial q/\partial y$; (lower right) effective diffusivity K_e. The curves may be identified by noting that higher τ_c corresponds in the interior to smaller q, larger F, and larger K_e.

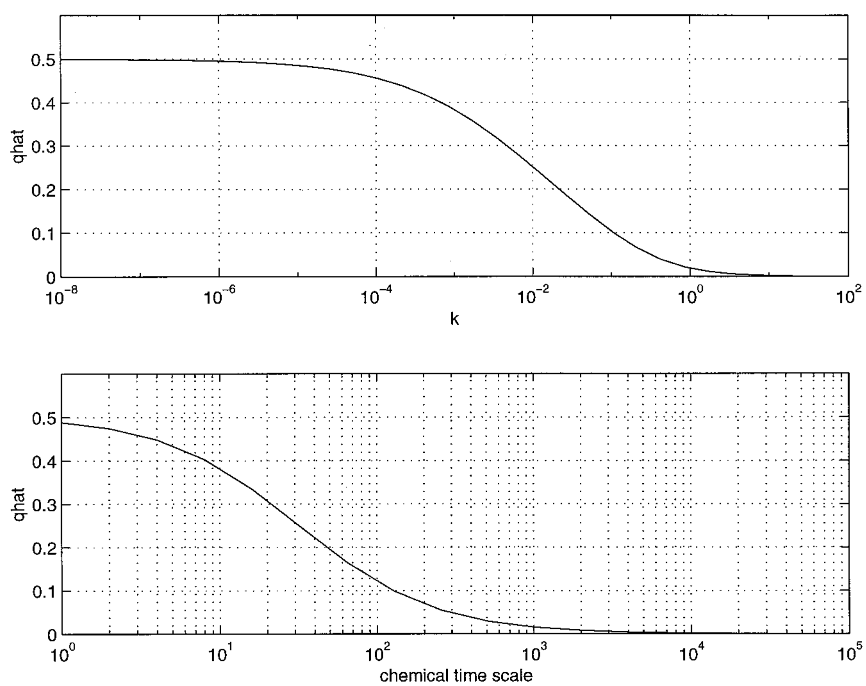

FIG. 3. Dependence of $\hat{q}$ on parameters. (upper) $\hat{q}$ as a function of k, for $\tau_c = 10$; (lower) $\hat{q}$ as a function of τ_c, for $k = 10^{-3}$.

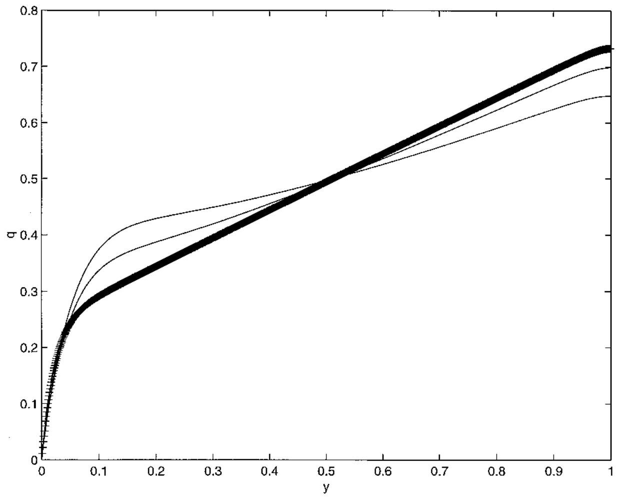

FIG. 4. Solutions of the effective diffusion equation, for $\tau_c = 1$, but with K_e derived from the solutions to the leaky barrier model with $\tau_c = 1$ (pluses/heavy curve), 10, and 100.

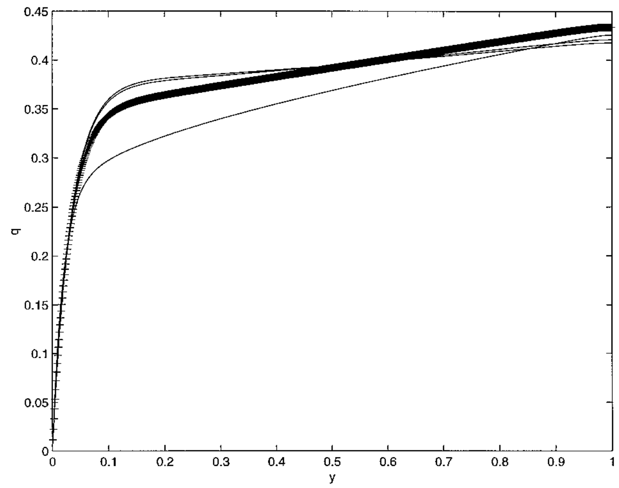

FIG. 5. Solutions of the effective diffusion equation, for $\tau_c = 10$, but with K_e derived from the solutions to the leaky barrier model with $\tau_c = 1$, 10 (pluses/heavy curve), 100, and 1000.

solution mimics exactly the corresponding solution of (23). Of more interest is the solution of (27) when τ_c is different from that for which K_e was derived. The extent to which a single profile K_e can reproduce the solution for various different values of τ_c is the extent to which the diffusion model can be made to *consistently* mimic the behavior of the model studied here, which includes a simple form of nonlocal mixing, throughout a finite region of the latter's parameter space. If K_e must be rederived from the solution to (23) every time a reasonably accurate solution of (27) is desired for some value of τ_c (again taking solutions of the former equation to represent ''ground truth''), then the latter equation has no practical value. Since there is not any built-in relationship between (total) flux and gradient in the real system, the relationship that actually occurs is interesting only if it proves, empirically, to have predictive power in the sense just discussed.

Figure 4 shows solutions of (27) when $\tau_c = 1$, but K_e has been derived from experiments with $\tau_c = 1$, 10, and 100. Figure 5 shows solutions when $\tau_c = 10$, and K_e has been derived from experiments with $\tau_c = 1$, 10, 100, and 1000. Equation (27) has been solved by matrix inversion on a staggered finite-difference grid with 1000 grid points. In both figures the heavy line is the solution when K_e has been derived for the correct value of τ_c [which is identical to the corresponding solution to (23)]. Though there are some significant differences, particularly in the former figure, in general the solutions are reasonably close to one another, again indicating reasonably good predictive power of the diffusion model when K_e is diagnosed empirically as we have done.

One fact that mitigates this positive assessment is that all the profiles K_e are constrained by the construction of the problem to be equal at $y = 0$, and indeed to be close to one another throughout the diffusive boundary layer near $y = 0$ since the true physics is in fact mainly diffusive there, with diffusivity k. This minimum value of K_e is an important parameter of the problem; it is the ''leakiness'' of the transport barrier. Figure 3 shows that throughout a range where k may be considered ''small,'' in that the associated length scale l is a small fraction of the domain size L, there is nonetheless significant variation with k in $\hat{q}$, which implies corresponding variation in the net steady-state flux into the region $y < 0$. However, the variation is roughly logarithmic in k in that region, so that an order-of-magnitude estimate of k is sufficient to determine $\hat{q}$ to better than a factor of 2.

On the other hand again, despite the extent to which we have constrained the diffusion model to perform well by fixing K_e at the boundary, there is some genuine physical meaning in the reasonably good success of that model in this case. This is, simply, that any form of mixing that is sufficiently strong to homogenize the interior is difficult to distinguish from any other. Homogeneity is the end result of all mixing that is strong enough that it dominates other processes (this is intuitive, but we showed it formally for any mixing that can be represented by a transilient matrix), and the process of homogenization is, essentially by definition, one in which information is lost. Hence, if K_e is sufficiently large in the interior to homogenize it, the solution will be reasonably insensitive to modest variations in K_e such as may be seen in Fig. 2. In Fig. 4, the above argument does not apply since the interior is not well mixed; it is noteworthy that $\tau_c = 1$ shows greater inconsistencies with the larger values of τ_c than the larger values do with each other. That is, the K_e evaluated for $\tau_c = 1$ gives relatively poor results when used in (27) with $\tau_c = 10$ and vice versa, while K_e evaluated for $\tau_c = 100$ or 1000 yields relatively good results when used in (27) with $\tau_c = 10$.

7. Summary

Some properties of various transport models have been described. New material on transilient matrices (Stull 1984, 1993; Fiedler 1984) includes their derivation from a (discretized) advection equation, an attendant formal explanation of the effect known as convective structure memory (Ebert et al. 1989), statements of known properties of transilient matrices in more formal language than in previous work, and the observation, following the formula of Ebert et al. (1989), that the flux due to nonlocal mixing by a symmetric transilient matrix is always downgradient if the tracer gradient has a single sign. Simple models of nonlocal mixing combined with linearized chemistry have been developed and used as test beds to determine how well a diffusion model, with an empirically computed effective diffusivity, can represent their behavior over finite regions of parameter space. Because fluxes are downgradient in the nonlocal models (as well as in the stratosphere, generally), an effective diffusivity can always be derived from the solution for a single tracer that is positive and finite, and reproduces the solution for that tracer, though it contains no new information in that case. A practical limitation of the diffusion model has been explored by using the diffusivity derived for one species to model another.

The results provide some qualified support for the practice of using empirically derived diffusivity profiles to represent mixing that is at least partly nonlocal, as is the practice in some zonally averaged stratospheric models. This is in accord with previous boundary layer studies, which also showed that the diffusion model works reasonably well in situations where theory suggests it is inappropriate, if the diffusivity is treated as an empirical function and is optimized for the problem at hand (Lamb et al. 1975; Lamb and Durran 1978).

The reason for this may be partly the many degrees of freedom, or one might say tunability, that are implied by treating the diffusivity as an empirical function. However, another reason appears to be simply that mixing destroys information, in the process eliminating evidence about precisely how the destruction occurred. This can be illustrated by adapting Pierrehumbert's (1991a) analogy to a deck of cards. After the deck is shuffled, even if the shuffling is somewhat incomplete, it can be quite difficult to determine with certainty whether it was simply shuffled all together (nonlocal mixing) or broken into much smaller subdecks that were first shuffled individually, then combined a few at a time into larger subdecks that in turn were shuffled, etc. (local mixing). The same phenomenon is operative here.

Nonetheless it should not be forgotten that local and nonlocal mixing are different processes, and it should not be assumed that they cannot lead to different results. The present study simply shows that in some idealized stratosphere-like examples, the differences are subtle. There is some resonance with other problems in large-scale atmospheric dynamics; Pavan and Held (1996) have noted that diffusive theories for the effects of baroclinic eddies on tropospheric climate can in principle yield predictions indistinguishable from ''adjustment'' theories, which generally imply (whether incidentally or by design) a degree of nonlocality. With this in mind, several important caveats should be noted.

First, within the context of the simple models examined here, only a small part of the possible parameter space has been explored. We have not explored the sensitivity of the leaky barrier model to either varying shapes of the equilibrium profile $q_c(y)$, or to spatially variable τ_c. Varying these parameters might lead to different behavior, but the possible number of variations is so great that doing so appears premature in the absence of any particular scenario of clear interest. Additionally, these variations will generally render the leaky barrier model unsolvable by analytical means.

Second, only two specific nonlocal models have been used to test the diffusion model. One can legitimately object that the diffusion model might fail more dramatically if the same sort of tests were performed using some other nonlocal model. However, the two models chosen at least demonstrate two qualitatively different effects that one might have expected to test the limitations of the diffusion model: nonlocality per se (RA model) and a sharp transition from large-scale to small-scale mixing (leaky barrier model).

More importantly, we have only considered the optimal case in which the true fluxes and gradients of at least one species are known exactly, so that the effective diffusivity for that species can be computed with perfect accuracy. In reality, computing fluxes and gradients from stratospheric data is not a trivial step. The zonally averaged formalism can be expected to cause zonally asymmetric transport barriers to be misrepresented. The large amplitudes and long timescales of planetary waves make the accurate determination of the *irreversible* part of the eddy flux difficult (unless the time-averaging period is very long, i.e., a season or longer), and cause the sharp gradients near transport barriers to be blurred by zonal averaging (e.g., Nakamura 1995). The effective diffusivity implied by the resulting flux/gradient ratio is unlikely to lead to very accurate model representations of transport barriers. The present results demonstrate that the minimum value of the diffusivity is an important parameter, determining the leakiness of the transport barrier. Hence, though unaddressed here, the practical issue of how one can best compute the appropriate minima of K_{yy} at the poleward and equatorward edges of the stratospheric surf zones so as to maintain appropriate isolation of the different air masses is a significant one.

Acknowledgments. A large fraction of this work was done as part of the author's Ph.D. thesis at MIT under the supervision of R. Alan Plumb. I am grateful to Dr. Plumb for his encouragement, insights, and constructive criticisms, and in particular for very detailed comments on the first draft of this manuscript, leading to a number of significant improvements. I also thank Glenn Flierl and Dale Durran for helpful advice and discussions, and three anonymous reviewers for comments that led to a number of improvements. This work was completed at the University of Washington, where the author is supported by a NOAA Postdoctoral Fellowship in Climate and Global Change, administered by the University Corporation for Atmospheric Research. At MIT, this work was partly supported by a NASA Graduate Student Fellowship in Earth System Sciences, and partly by NASA Grant NAGW-1727.

APPENDIX A

Derivation of Transilient Matrix

Consider the Eulerian advection equation [(1)], in a bounded domain with no flow, hence no flux of tracer, through the boundaries. Consider a discretization of the tracer field in a Cartesian coordinate space, so that q_{ij} is the value of the tracer at $x = i\Delta x$, $y = j\Delta y$, with Δx and Δy the grid spacings in x and y. For convenience take $\Delta x = \Delta y$, and let the domain be square so that its length in each dimension is $N\Delta x$. These geometric simplifications are easily relaxed. Also, let the grid spacing be sufficiently small so that each grid box may be considered homogeneous; there is negligible subgrid-scale variability in q.

Since the advection of a passive tracer by a known flow field is a linear process, we can write

$$q_{ij} = \sum_{i'} \sum_{j'} A_{iji'j'} q_{i'j'},$$

where primes represent values at the previous time step, and sums are implicitly taken over all values of their indices, or in more compact notation as

$$\mathbf{q}(t + \Delta t) = \mathbf{A}(t)\mathbf{q}(t), \tag{A1}$$

where $\mathbf{q}$ is the entire two-dimensional matrix of q_{ij}'s and $\mathbf{A}$ the four-dimensional matrix of $A_{iji'j'}$'s. The elements of $\mathbf{A}(t)$ depend on $\mathbf{u}$ at the previous time step and perhaps at the current time step but are independent of $\mathbf{q}$, by linearity. The matrix $\mathbf{A}$ can include diffusion as well since it is a linear process.

By successive multiplications from the left one can express the tracer field at time $t + m\Delta t$, where m is any integer, in terms of the field at time t by

$$\mathbf{q}(t + m\Delta t) = \mathbf{A}(t + [m-1]\Delta t)\mathbf{A}(t + [m-2]\Delta t) \times \cdots \mathbf{A}(t)\mathbf{q}(t). \tag{A2}$$

We can write (A2) more compactly as

$$\mathbf{q}(t + m\Delta t) = \mathbf{B}(t, m\Delta t)\mathbf{q}(t) \tag{A3}$$

if we define

$$\mathbf{B}(t, m\Delta t) \equiv \mathbf{A}(t + [m-1]\Delta t)\mathbf{A}(t + [m-2]\Delta t) \times \cdots \mathbf{A}(t)$$

so that $\mathbf{B}(t, m\Delta t)$ maps the tracer field at time t directly to itself m time steps later. In order to justify the exclusion of the source and sink terms from this matrix, we must assume that they are small, that is, that the tracer q is quasi-conserved over times $m\Delta t$.

Now define a standard average over x:

$$\overline{q}_j = \frac{1}{N} \sum_i q_{ij}$$

and perturbation

$$\Delta q_{ij} = q_{ij} - \overline{q}_j.$$

Then, letting primes denote indices at time t and quantities without primes denote indices at $t + m\Delta t$, take the x average of Eq. (A3). In element form this is

$$\overline{q}_j = \frac{1}{N} \sum_i \sum_{i'} \sum_{j'} B_{iji'j'}(\overline{q}_{j'} + \Delta q_{i'j'}). \tag{A4}$$

If we break the rhs up into two separate terms, we get

$$\overline{q}_j = \frac{1}{N} \sum_i \sum_{i'} \sum_{j'} B_{iji'j'}\overline{q}_{j'} + \frac{1}{N} \sum_i \sum_{i'} \sum_{j'} B_{iji'j'}\Delta q_{i'j'}. \tag{A5}$$

Now if we define

$$C_{jj'} \equiv \frac{1}{N} \sum_i \sum_{i'} B_{iji'j'}, \tag{A6}$$

then (A5) becomes

$$\overline{q}_j = \sum_{j'} C_{jj'}\overline{q}_{j'} + \frac{1}{N} \sum_i \sum_{i'} \sum_{j'} B_{iji'j'}\Delta q_{i'j'}, \tag{A7}$$

which is simply (5) with an additional term on the rhs, which will be called the error term below.

If the full 2D advection–diffusion operator conserves mass, the transilient matrix defined by (A6) does also, since the full operator operating on a tracer distribution that is initially uniform in the x direction is identical to the transilient matrix operating on such a distribution. Additionally, we can guarantee that the transilient matrix will have no negative elements, since the matrix $\mathbf{A}$ cannot have any; if it did it could produce negative values where there were none before, which advection cannot do. However, it should be pointed out that while the true physical advection process is both linear and positive definite, most real numerical advection schemes do not have both of these properties (Thuburn and McIntyre 1997).

If the error term is significant, it will be impossible to correctly represent the mean profile of q at $t + m\Delta t$ purely in terms of the mean profile at t, since the error term depends unavoidably on variations of q with respect to i'. Mathematically, the transilient matrix model may be thought of as an approximation to the representation in terms of the exact Lagrangian conditional probability density (e.g., Batchelor 1949; Lamb et al. 1975) in which all information about tracer inhomogeneities in the x direction (in our two-dimensional analysis) is ignored. In physical terms, the transilient matrix formulation implies either that an air parcel traveling from one level to another during a given fixed time interval carries with it a tracer concentration equal to the mean concentration at its level of origin, or that the deviations of transported parcels from the mean concentrations at their levels of origin cancel in the sums that define the error term.

The above discussion shows that (5) is the most general model possible of eddy transport that uses only the mean profile of the tracer and that is local in time (the Markov property). Any valid transport model that represents the eddy flux divergence by an operator on the mean tracer field can be written, once discretized in space and time, in the form of a transilient matrix. A diffusion model, unless the time step is taken to be long enough to represent a nominally finite time [i.e., significantly longer than $(\Delta y)^2/k$ with k the diffusivity and y the grid spacing], will appear as a tridiagonal matrix.

The fact that the error term need not vanish in general can cause the behavior of a transilient matrix model to be dependent on the time step, which is a free parameter. That is, one may find that

$$\mathbf{C}(n\Delta t) \neq [\mathbf{C}(\Delta t)]^n, \tag{A8}$$

where n is some number, and the transilient matrix is written as dependent on the time step only, rather than on the time as well, since the turbulence is assumed stationary. Ebert et al. (1989; see also Stull 1993) referred to this effect as convective structure memory and gave a heuristic discussion of its cause. They found, in the case of a particular numerically simulated turbulent flow for which they computed transilient matrices with varying values of the time step, that this difficulty was minimized when Δt was chosen to be near the ‘‘convective timescale,’’ essentially the turnover time of the largest eddies.

It is plausible, but not obviously true, that the turnover time of the largest eddies is the choice of time step that yields the greatest self-consistency of the transport model (5) for *any* flow for which one might wish to use that model, as implied by Stull (1993). A more general analysis of this issue would be desirable.

APPENDIX B

Derivation of 1D Chemical Transport Model with the Transilient Matrix

Assume, adding tendencies in the spirit of some numerical modeling approaches, that the tracer profile $\mathbf{q}$ is first mixed instantaneously by the transilient matrix, yielding a profile $\mathbf{q}^*(t + \Delta t)$, and then that local chemistry acts for time Δt. Then we have

$$\mathbf{q}^*(t + \Delta t) = \mathbf{C}\mathbf{q}(t);$$

$$\mathbf{q}(t + \Delta t) = \mathbf{q}_c - [\mathbf{q}_c - \mathbf{q}^*(t + \Delta t)]e^{-\Delta t/\tau_c},$$

so that in total,

$$\mathbf{q}(t + \Delta t) = \mathbf{C}\mathbf{q}(t)e^{-\Delta t/\tau_c} + \mathbf{q}_c[1 - e^{-\Delta t/\tau_c}].$$

In the limit $\Delta t/\tau_c \ll 1$,

$$\mathbf{q}(t + \Delta t) \approx \mathbf{C}\mathbf{q}(t) + \frac{\Delta t}{\tau_c}[\mathbf{q}_c - \mathbf{C}\mathbf{q}(t)].$$

Now assume also that $\Delta t/\tau_m \ll 1$, where τ_m is the mixing timescale. Then $\mathbf{C} \approx \mathbf{I} + O(\Delta t/\tau_m)$, and so, retaining terms linear in Δt,

$$\mathbf{q}(t + \Delta t) \approx \mathbf{C}\mathbf{q}(t) + \frac{\Delta t}{\tau_c}[\mathbf{q}_c - \mathbf{q}(t)],$$

as introduced in section 4c.

APPENDIX C

Solvability of Steady-State 1D Chemical Transport Model with the Transilient Matrix

Let us assume that the columns of $\mathbf{C}$ are all linearly independent, though strictly this need not be the case. If nothing else, linear independence can be assured by taking a sufficiently small time step, since in the limit as the time step vanishes, the diagonal elements must become much larger than the off-diagonal elements. This small time step may be suboptimal for reasons given in the discussion of convective structure memory in appendix A. Nonetheless, in practice it seems likely (and existing calculations of transilient matrices support this notion) that transilient matrices for real flows will generally have linearly independent columns for any reasonable choice of time step.

Given linear independence of the columns of $\mathbf{C}$, that matrix will have N independent eigenvectors $\mathbf{q}_k$, which satisfy

$$\mathbf{C}\mathbf{q}_k = \lambda_k\mathbf{q}_k,$$

where the λ_k's are the eigenvalues as in section 4b, which may or may not be distinct. We can then expand any profile $\mathbf{q}$ in the eigenvectors, as in (13), and express all time dependence in terms of the coefficients a_k:

$$a_k(t + \Delta t) = \lambda_k a_k(t).$$

Since the eigenvalues satisfy (8), we can establish the solvability of (10) for any valid transilient matrix, hence any transport model that is local in time and is representable as an operator on the mean tracer profile. The Fredholm alternative (e.g., Hochstadt 1973) states that an equation of the form (11) has a unique solution for arbitrary $f(y)$ if and only if the associated homogeneous equation

$$\int_0^{N\Delta y} C(y|y')\overline{q}(y')\,dy' - \lambda\overline{q}(y) = 0 \tag{C1}$$

has no solution other than $\overline{q}(y) = 0$. Again ignoring the distinction between discrete and continuous versions of this theorem, we can see that the homogeneous form of (10), which is

$$\mathbf{C}\mathbf{q} - \left(1 + \frac{\Delta t}{\tau_c}\right)\mathbf{q} = 0, \tag{C2}$$

can have no nontrivial solution. This is because all eigenvalues of the matrix $\mathbf{C}$ are positive and less than or equal to unity, while the quantity $1 + \Delta t/\tau_c$ is always greater than unity. Hence, no eigenvector of $\mathbf{C}$ is a solution of (C2), and so by linearity no solution to (C2) can be constructed.

Since (10) has a unique solution for all $f(y)$, the solution to the time-dependent equation [(9)] will smoothly approach it, as long as $\Delta t/\tau_c$ is sufficiently small to avoid numerical instability. This is guaranteed by the property [(8)] that is satisfied by the eigenvalues of all the nonuniform eigenvectors.

REFERENCES

Aref, H., 1984: Stirring by chaotic advection. *J. Fluid Mech.,* **143,** 1–21.

Batchelor, G. K., 1949: Diffusion in a field of homogeneous turbulence. I. Eulerian analysis. *Aust. J. Sci. Res.,* **2,** 437–450.

Bercowicz, R., and L. P. Prahm, 1979: Generalization of K theory for turbulent diffusion. Part I: Spectral turbulent diffusivity concept. *J. Appl. Meteor.,* **18,** 266–272.

Budyko, M. I., 1969: The effect of solar radiation variations on the climate of the earth. *Tellus,* **21,** 611–619.

Corrsin, S., 1974: Limitations of gradient transport models in random walks and in turbulence. *Advances in Geophysics,* Vol. 18a, Academic Press, 25–60.

Ebert, E. E., U. Schumann, and R. B. Stull, 1989: Nonlocal turbulent mixing in the convective boundary layer evaluated from large-eddy simulation. *J. Atmos. Sci.,* **46,** 2178–2207.

Fiedler, B. H., 1984: An integral closure model for the vertical turbulent flux of a scalar in a mixed layer. *J. Atmos. Sci.,* **41,** 674–680.

Flierl, G. R., and W. K. Dewar, 1985: Motion and dispersion of dumped wastes by large-amplitude eddies. *Offprints from Wastes in the Ocean,* Kester et al., Eds., Vol. 5, *Deep-Sea Waste Disposal,* John Wiley and Sons, 147–169.

Garratt, J. R., 1992: *The Atmospheric Boundary Layer.* Cambridge University Press, 316 pp.

Haidvogel, D. B., and I. M. Held, 1980: Homogeneous quasi-geostrophic turbulence driven by a uniform temperature gradient. *J. Atmos. Sci.,* **37,** 2644–2660.

Hama, F. R., 1962: Streaklines in a perturbed shear flow. *Phys. Fluids,* **5,** 644–650.

Hochstadt, H., 1973: *Integral Equations.* John Wiley and Sons, 282 pp.

Lamb, R. G., and D. R. Durran, 1978: Eddy diffusivities derived from a numerical model of the convective planetary boundary layer. *Il Nuovo Cimento,* **1,** 1–17.

——, W. H. Chen, and J. H. Seinfeld, 1975: Numerico-empirical analyses of atmospheric diffusion theories. *J. Atmos. Sci.,* **32,** 1794–1807.

Larson, V. E., 1999: The relationship between the transilient matrix and the Green's function for the advection-diffusion equation. *J. Atmos. Sci.,* **56,** 2447–2453.

McIntyre, M. E., 1991: Atmospheric dynamics: Some-fundamentals with observational implications. *Proceedings of the International School of Physics 'Enrico Fermi', CXV Course,* J. C. Gille and G. Visconti Eds., North-Holland, 313–386.

——, and T. N. Palmer, 1983: Breaking planetary waves in the stratosphere. *Nature,* **305,** 593–600.

——, and ——, 1984: Surf zone in the stratosphere. *J. Atmos. Terr. Phys.,* **46,** 825–849.

——, and ——, 1985: A note on the general concept of wave breaking for Rossby and gravity waves. *Pure Appl. Geophys.,* **123,** 964–975.

Nakamura, N., 1995: Modified Lagrangian-mean diagnostics of the stratospheric polar vortices. Part I: Formulation and analysis of GFDL SKYHI GCM. *J. Atmos. Sci.,* **52,** 2096–2108.

Newman, P. A., 1988: Mixing rates calculated from potential vorticity. *J. Geophys. Res.,* **93,** 5221–5240.

——, M. R. Schoeberl, and R. A. Plumb, 1986: Horizontal mixing coefficients for two-dimensional chemical models calculated from National Meteorological Center data. *J. Geophys. Res.,* **91,** 7919–7924.

Ngan, K., and T. G. Shepherd, 1999a: A closer look at chaotic advection in the stratosphere. Part I. Geometric structure. *J. Atmos. Sci.,* in press.

——, and ——, 1999b: A closer look at chaotic advection in the stratosphere. Part II. Statistical diagnostics. *J. Atmos. Sci.,* in press.

Ottino, J. M., 1989: *The Kinematics of Mixing: Stretching, Chaos and Transport.* Cambridge University Press, 346 pp.

——, 1990: Mixing, chaotic advection, and turbulence. *Annu. Rev. Fluid Mech.,* **22,** 207–253.

Pavan, V., and I. M. Held, 1996: The diffusive approximation for eddy fluxes in baroclinically unstable jets. *J. Atmos. Sci.,* **53,** 1262–1272.

Pierce, R. B., and T. D. A. Fairlie, 1993: Chaotic advection in the stratosphere: Implications for the dispersal of chemically perturbed air from the polar vortex. *J. Geophys. Res.,* **98,** 18 589–18 595.

Pierrehumbert, R. T., 1991a: Chaotic mixing of tracer and vorticity by modulated traveling Rossby waves. *Geophys. Astrophys. Fluid Dyn.,* **58,** 285–319.

——, 1991b: Large-scale horizontal mixing in planetary atmospheres. *Phys. Fluids,* **3A,** 1250–1260.

Plumb, R. A., 1979: Eddy fluxes of conserved quantities by small-amplitude waves. *J. Atmos. Sci.,* **36,** 1699–1704.

——, and J. D. Mahlman, 1987: The zonally averaged transport characteristics of the GFDL general circulation/transport model. *J. Atmos. Sci.,* **44,** 298–327.

Prather, M. J., 1994: Lifetimes and eigenstates in atmospheric chemistry. *Geophys. Res. Lett.,* **21,** 801–804.

——, 1996: Time scales in atmospheric chemistry: Theory, GWPs for CH_4 and CO, and runaway growth. *Geophys. Res. Lett.,* **23,** 2597–2600.

Raymond, D. J., 1994: Convective processes and tropical atmospheric circulations. *Quart. J. Roy. Meteor. Soc.,* **120,** 1431–1455.

Roberts, P. H., 1961: Analytical theory of turbulent diffusion. *J. Fluid Mech.,* **11,** 257–283.

Sobel, A. H., and R. A. Plumb, 1999: Quantitative diagnostics of mixing in a shallow water model of the stratosphere. *J. Atmos. Sci.,* **56,** 2811–2829.

Stull, R. B., 1984: Transilient turbulence theory. Part I: The concept of eddy-mixing across finite distances. *J. Atmos. Sci.,* **41,** 3351–3367.

——, 1988: *Introduction to Boundary Layer Meteorology.* Kluwer Academic, 666 pp.

——, 1993: Review of non-local mixing in turbulent atmospheres: Transilient turbulence theory. *Bound.-Layer Meteor.,* **62,** 21–96.

Taylor, G. I., 1921: Diffusion by continuous movements. *Proc. Roy. Math. Soc.,* **2,** 196–212.

Taylor, H. M., and S. Karlin, 1994: *An Introduction to Stochastic Modeling.* Academic Press, 566 pp.

Thuburn, J., and M. E. McIntyre, 1997: Numerical advection schemes, cross-isentropic random walks, and correlations between chemical species. *J. Geophys. Res.,* **102,** 6775–6797.

Wiggins, S., 1992: *Chaotic Transport in Dynamical Systems.* Springer-Verlag, 301 pp.

Yang, H., K. K. Tung, and E. Olaguer, 1990: Nongeostrophic theory of zonally averaged circulation. Part II: Eliassen–Palm flux divergences and isentropic mixing coefficients. *J. Atmos. Sci.,* **47,** 215–241.