62

# Quantitative Diagnostics of Stratospheric Mixing

by

Adam Harrison Sobel

B.A., Wesleyan University (1989)

Submitted to the Department of Earth, Atmospheric and Planetary Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1998



Author .....
Department of Earth, Atmospheric and Planetary Sciences
September 3, 1997

Certified by .....
R. Alan Plumb
Professor of Meteorology
Thesis Supervisor

Accepted by .....
Thomas H. Jordan
Chairman, Department of Earth, Atmospheric and Planetary Sciences

# Quantitative Diagnostics of Stratospheric Mixing

by
Adam Harrison Sobel

Submitted to the Department of Earth, Atmospheric and Planetary Sciences
on September 3, 1997, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

## Abstract

This thesis addresses the planetary-scale mixing of tracers along isentropic surfaces in the extratropical winter stratosphere. The primary goal is a more fully quantitative understanding of the mixing than is available at present.

The general problem of representing eddy mixing in a one-dimensional mean representation of a two-dimensional flow is discussed. The limitations of the eddy diffusion model are reviewed, and alternatives explored.

The stratosphere may, for some purposes, be viewed as consisting of relatively well-mixed regions separated by moving, internal transport barriers. Methods for diagnosing transport across moving surfaces, such as tracer isosurfaces, from given flow and tracer fields are reviewed.

The central results of the thesis involve diagnostic studies of output from a shallow water model of the stratosphere. It is first proved that in an inviscid shallow water atmosphere subject to mass sources and sinks, if the mass enclosed by a potential vorticity (PV) contour is steady in time, then the integral of the mass source over the area enclosed by the contour must be zero. Next, two different approaches are used to diagnose the time-averaged transport across PV contours in the model simulations. The first is the modified Lagrangian mean (MLM) approach, which relates the transport across PV contours to PV sources and sinks. The second is called "local gradient reversal" (LGR), and is similar to contour advection with surgery.

The model includes a sixth-order hyperdiffusion on the vorticity field. Except in a thin outer "entrainment zone", the hyperdiffusion term has only a very weak effect on the MLM mass budget of the polar vortex edge. In the entrainment zone, the hyperdiffusion term has a significant effect. The LGR results capture this behavior, providing good quantitative estimates of the hyperdiffusion term, which is equivalent to the degree of radiative disequilibrium at a PV contour. This agreement shows that the main role of the hyperdiffusion is to remove filaments. It is argued that these results do not depend on the details of the small-scale dissipation.

Using a more direct type of trajectory-based calculation, the "transilient matrix" for the shallow water model flow is constructed. The matrix is used as the basis for a one-dimensional chemical transport model of the two-dimensional shallow water flow. A highly idealized representation of (true) latitude-dependent chemistry is included. The one-dimensional model represents the two-dimensional model reasonably well, but is surprisingly insensitive to some details of the transilient matrix. The transilient matrix calculations also show, as expected, that the model polar vortex is extremely isolated from its exterior.

The various different diagnostics, taken together, allow a comprehensive description of the Lagrangian circulation in the model's winter extratropics to be composed, including the relationships between parcel trajectories and PV contours in different flow regions.

Thesis Supervisor: R. Alan Plumb
Title: Professor of Meteorology

## Acknowledgments

I have surely forgotten to mention everyone who deserves to be mentioned, and I apologize.

Thanks are due first to my advisor, Professor R. Alan Plumb. Alan's reputation as a scientific researcher needs no comment, but his merits as a teacher and mentor are known directly only to a privileged few. Not least among these is his ability to maintain a particularly collegial and stimulating working environment, in which it has been a pleasure to work during these past four years.

Alan has allowed me great independence in even the most basic aspects of my research. He has managed to provide me with support, encouragement, and critical advice, while still letting me think for myself. It is not my job to judge the value of the results, but the experience has been tremendously rewarding, and I appreciate having been given the chance.

Reading of previous students' theses reveals that it is a standard formality to thank the other members of one's committee besides one's advisor. In this case, however, Professors Kerry Emanuel, Glenn Flierl, and Richard Lindzen, deserve it. During my four years here at MIT, each of them has been generous with his time, knowledge, and scientific wisdom, in many individual discussions with me, not to mention classroom teaching. It has been a great privilege to learn from Professors Emanuel, Flierl, and Lindzen, and I offer them my sincerest thanks. I also thank Professor Mario Molina for serving on my committee at the outset, when I submitted my thesis proposal (which, for better or worse, had little to do with the end result).

During my first year and a half here, while he was working with Alan as a postdoc, Darryn Waugh served (perhaps partly due to coercion on my part) as an informal second advisor to me, enabling me to "get up to speed" in the research I was then beginning, much more quickly than I would have otherwise. He provided computer codes, help with innumerable technical tasks, and wise scientific advice. This sort of help, from someone advanced enough in research to know the answers to a beginner's questions, but not so far along as to have forgotten what it was like to be a beginner himself, cannot be easily overrated. After Darryn's departure it became more difficult to get him to debug my FORTRAN, but he has continued to provide helpful advice and encouragement. Lorenzo Polvani has also helped me a great deal. He has done this first of all by providing the shallow water model used here as well as hours of instruction and assistance in its use. Besides this, in several detailed discussions of my research, he helped me to clarify where it was going and why. He and Darryn have both greatly enlivened the intellectual atmosphere here in room 54-1725, and I thank them both.

The time I have spent in this computer room would not have been the same, either, without any of its other inhabitants. The people here, working with Alan in various capacities, have been the source of endless discussions, debates, and jokes, scientific and otherwise. These have been needed distractions, enabling me to spend all day here without losing my marbles, even when the work was not going well. Janusz Eluszkiewicz, Gavin Esler, Will Heres, Chia-Hui (Juno) Hsu, Tieh-Yong Koh, Moto Nakamura, Jessica Neu, Chantal Rivest, and Xinyu Zheng have all contributed to making this an interesting and friendly place to spend my days. Will deserves special thanks for keeping the computer system running with professionalism and good humor.

I have also benefited greatly from the friendship, intellectual stimulation, and assistance of many people here who work in other rooms besides 54-1725. Danny Kirk-Davidoff, Nili Harnik, Gerard Roe, and Amy Solomon have been good friends and stimulating colleagues

during my entire four years here. It was a pleasure to share an office with Sudharshan Sathiyamoorthy during the first two years, and I have missed him since. Michael Chechelnitsky has always been willing to put his own work down at a moment's notice, either to lend his considerable talents to helping me solve my mathematical problems, or to begin interminable debates on unanswerable philosophical questions. I have also had helpful discussions at various stages with Myles Allen, Brian Arbic, and Johan Nilsson.

Jane McNabb has, tirelessly and skillfully, helped me through many minor bureaucratic difficulties which could have otherwise been major ones. Tracy Stanelun and Joan Windover have also been nothing but helpful and kind.

In addition, I would like to acknowledge a number of people outside MIT for help of various sorts. Ross Salawitch taught me a great deal about stratospheric chemistry while we were collaborating on a project during my second year at MIT, valuable background for any student of stratospheric transport. Richard Goody has taught me a lot about atmospheric radiation and related subjects, and shared his valuable, well-earned insights into the whole enterprise of studying the atmosphere. Bjorn Stevens read and commented perceptively on several drafts of my first research paper (Sobel et al. 1997), and introduced me to many interesting ideas from the field of turbulent entrainment in boundary layers. I have appreciated greatly many stimulating discussions with Paul Kushner, on stratospheric dynamics and transport and a wide range of other subjects, as well as the friendship and hospitality of him and his wife, Catherine, during several visits to GFDL. Peter Haynes and Martin Juckes pointed out existing results which were closely related to one derived here (section 6.2) and helped me to understand the relationships between the various results (explained in appendix B). Stephen Wiggins explained some aspects of his lobe dynamics method to me; if I were to do this thesis again starting now, I would take more time to understand, and possibly use this method. I also thank Roger Atkinson, David Dritschel, and Isaac Held for helpful discussions.

My parents, Cynthia and Gerald Sobel, are responsible in uncountable ways, besides the obvious ones, for my having reached this point. I am truly blessed to have had their love and support during my first thirty years of life, and to have had my sister, Melissa, who has kept life interesting for all of that time save the first two years.

My beloved wife, Marit Larson, bears direct responsibility for this thesis. She suggested to me in the first place that I ought to study meteorology, an idea which probably would not otherwise have entered my mind. Most importantly of all, and despite the fact that I have failed to solve any global environmental problems, she has continued to make my life a happy one these past four years.

> If from the ginkgo tree a single little yellow leaf falls and rests on the lawn, the sensation felt in looking at it is that of a single yellow leaf. If two leaves descend from the tree, the eye follows the twirling of the two leaves as they move closer, then separate in the air, like two butterflies chasing each other, then glide finally to the grass, one here, one there. And so with three, with four, even with five; as the number of leaves spinning in the air increases further, the sensations corresponding to each of them are summed up, creating a general sensation like that of a silent rain, and — if the slightest breath of wind slows their descent — that of wings suspended in the air, and then that of a scattering of little luminous spots, when you lower your gaze to the lawn. Now, without losing anything of these pleasant general sensations, I would like to maintain distinct, not confusing it with the others, the individual image of each leaf from the moment it enters the visual field, and flollow it in its aerial dance until it comes to rest on the blades of grass.

Italo Calvino, *If on a Winter's Night a Traveller*, translated from the Italian by William Weaver.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Preface

The subject of this thesis is the global-scale, quasi-isentropic mixing which occurs in the extratropical winter stratosphere. The research reported here deals in large part with two questions.

The first is how some meaningful properties of the mixing, that is, properties which can be inserted into appropriate budget equations for quantities being mixed, can best be diagnosed from observed or numerically simulated data. The problem here lies in choosing the right properties as much as in performing the diagnosis. The second question concerns the way in which the mixing interacts with radiative and chemical processes to maintain the structure of the polar vortex in both dynamical and chemical tracers.

The two questions are related. For example, if the polar vortex is defined as a pool of air with high potential vorticity (PV), then the mixing of that quantity opposes the tendency of radiation to concentrate it over the poles. A definition of mixing which allows the balance between these processes to be meaningfully quantified is clearly a useful one, since it relates the mixing in a direct way to another process which is relatively well understood, and more straightforwardly observable.

In the process of organizing the material, I have made some attempt to separate these two questions for the sake of clarity. However, this has proved difficult, due to the connection mentioned immediately above. As a result, the material has (as is usual) been organized largely along the lines of the distinct types of calculations which have been carried out, and the various scientific issues are raised as results appear which are relevant to them.

A third, related issue is discussed here as well. This is how the mixing can best be parameterized in reduced-dimensional (e.g. zonally averaged) chemical transport models. This issue is not tackled in its entirety, and in fact some important aspects of it are not discussed at all. Rather, a number of subsidiary issues are discussed in considerable detail. particularly in chapters 3 and 7.

## 1.2 Synopses of chapters

Chapter 2 contains some basic review material on stratospheric dynamics and transport. There is nothing new here, and a number of good reviews covering most of the important issues can be found in the literature, so I have kept this brief.

Chapter 3 is a general theoretical discussion of eddy transport in two dimensions (2D).

The emphasis is on nonlocal mixing by large eddies, and as such the discussion is relevant to the stratospheric surf zone. The thinking recorded here was driven by the question of whether, or rather subject to what limitations, stratospheric mixing can be represented by an eddy diffusivity, as is done in all zonally averaged chemical transport models, and if not what alternatives are possible. This question is not truly answered. In fact it is probably not answerable in general; how good or bad a diffusive parameterization is depends on for what it is being used and on how the diffusivity, which in general may be a function of both space and time, is obtained. Because the line of thought pursued here is partially abandoned in the remainder of the thesis, it may seem something of a non sequitur. If so, this chapter may simply be viewed as a stand-alone review and discussion of some ideas about mixing. It contains a detailed introduction to the *transilient matrix*, a device which is central to chapter 7.

Chapter 4 reviews various techniques which have been used to estimate the rate at which material crosses a tracer isosurface (in 3D) or contour (in 2D) from given flow and tracer fields. Atmospheric and oceanic "transport barriers" can usually be conveniently defined as coinciding with such an isosurface, or a finite region bounded by two isosurfaces, though other definitions are defensible (and one such is briefly discussed). The discussion here is thus not only relevant to the permeability of the polar vortex edge, but also to other transport problems, most prominently the permeability of the extratropical tropopause, and some explicit discussion of the latter problem is included.

Chapter 5 provides necessary background for chapters 6 and 7, by introducing the shallow-water model and the details of the particular simulations to be analyzed.

Chapter 6 contains what in my view are the most significant results of the thesis. The simple but important result is derived that an inviscid shallow-water atmosphere in an MLM steady state (which does not imply steady flow fields) must be in net (meaning, averaged over the area enclosed by PV contours) radiative equilibrium, or rather the shallow-water analog of it. This result is new, being a generalization of earlier results which required steady flow fields.

The terms in the MLM equation are computed from the output of a shallow water model of the stratosphere. The main vortex turns out to be effectively nearly inviscid, and thus near radiative equilibrium. The surf zone is far from radiative equilibrium, implying that it is effectively a relatively viscous region. This clarifies the distinction between the vortex and the surf zone in a quantitative way.

Following this, a scheme using trajectory methods for diagnosing the transport across tracer contours due to filamentation events is applied to the shallow water data. The method, labeled "local gradient reversal", is original but borrows its essential concept from the technique of contour surgery. The transport computed by this method is compared to the transport caused by the hyperdiffusion term, whose presence in the model allows departures from MLM radiative equilibrium in steady state. The two are shown to agree well in some important respects, despite that no information about the nature of the hyperdiffusion (other than that the qualitative knowledge that it is a weak, small-scale dissipation) was included in the trajectory-based calculations. This offers support for the claim that, at least in one particular dynamical regime, this and similar trajectory techniques can be used to quantify transport — more specifically, the entrainment of vortex edge air into the surf zone — without any direct consideration of small-scale mixing or dissipation.

Chapter 7 is constructed around a set of "brute force" transport calculations using the shallow water model output. The method here is the one referred to as "contour crossing" by Sobel, Plumb, and Waugh (1997, hereafter SPW). Those authors showed this

method to be highly sensitive to noise, but the shallow water data is free of noise, so this is not a problem. These calculations are used to construct the "transilient matrix" (Stull 1984, Stull 1993) for the model flow, which is in turn used to construct a one-dimensional (1D) parameterization of the full two-dimensional (2D) model, including a term meant to represent latitude-dependent chemistry in a very simplified way. The 1D model is then compared to the 2D model (with the same chemistry term). The agreement is good, but complications to be explained in the chapter render this a less than completely satisfying result. The complications themselves, however, are of some interest. Also, the results of the transilient matrix calculations provide a complement for those of chapter 6. Since the contour crossing calculations are fully Lagrangian, they have direct implications for chemical tracer fluxes across PV contours, whereas the technique used in chapter 6 only determines a net mass flux, which could have cancelling components of opposite sign and thus hide tracer fluxes of larger magnitude than the mass flux. The results of chapter 7 show that this is not the case for the shallow water flow studied here. This information allows a comprehensive description to be constructed of the Lagrangian circulation in the model's winter extratropics.

Chapter 8 summarizes the most important conclusions of the thesis, and concludes with a few additional remarks.

## 1.3 A note about terminology

Problems of the sort studied here tend to breed semantic difficulties. Terms such as *stirring*, *mixing*, and *transport*, not to mention *wave breaking*, may have different meanings to different people, or sometimes no clear meaning at all. Two extreme and opposite approaches to this difficulty are possible. One is to spend a lot of effort in carefully defining each term so that there is no risk of confusion. The second is to use the different terms casually without defining them. The danger in the second approach is obvious. The first approach may seem the better one, but it too can mask, if not sloppy thinking, then at least poor scientific judgment. In my view, what is important in science is that one be able to compute or measure quantities which have some use. Excessive attention paid to terminology can sometimes be a distraction, hiding the fact that nothing useful has been produced. By a useful quantity I mean, for example, that knowledge of it enables some other distinct quantity to be predicted or constrained. The chain of quantities being predicted or constrained should eventually lead back to something of intrinsic interest to someone other than the particular scientists doing the predicting.

I will try to steer a practical course which avoids the pitfalls of both approaches. In some parts of the discussion I will use the above mentioned terms, and other related ones, loosely and without definition. These will be parts where in my judgment the precise meaning of the term in question is either not essential or apparent enough from context, and so being too careful about the definition would only divert attention from the argument. On the other hand, a central goal of this research is to isolate quantities related to transport, mixing, stirring, wave breaking, etc. (pick your favorite) which can be put directly to some practical purpose, i.e. as terms inserted into appropriate budget equations which may also include other processes as discussed above. I will try to make precisely clear what these quantities are, how they can be computed, and for what they can be used. To the extent that I achieve this goal, arguments about what word best describes the process being quantified will be irrelevant.

A special case is the word "diffusion". I will use this word in only one specific sense, to denote a linear flux-gradient relationship:

$$\mathbf{F} = -K\nabla q$$

where $q$ is some property, $\mathbf{F}$ is the flux of that property, and $K$ is the diffusivity, which in general may be a function of space and time. However, there are still two partially distinct sorts of diffusion which will arise. One is "small-scale" diffusion, which acts on the full, un-averaged, instantaneous tracer field, and is assumed to be weak. The other is "large-scale" diffusion, which acts on a tracer field which has been averaged over one spatial dimension and possibly time, and presumably represents the average effect of large-scale eddies. The most relevant example of this is the latitudinal diffusion used in zonally averaged stratospheric chemical transport models, which is meant to represent the effects of planetary waves. Which meaning is intended should be apparent from context, as long as the distinction is kept in mind.

# Chapter 2

# The Stratosphere: Generalities

Some salient aspects of extratropical stratospheric dynamics and transport are reviewed here. Andrews et al. (1987, p.360) explain in detail why "... the gross characteristics of transport in the meridional plane can be modeled in terms of a combination of advection by a mean meridional mass circulation and quasi-isentropic mixing by large-scale eddies." This statement succintly delineates the essential conceptual backdrop of the present work, the subject of which is the quasi-isentropic mixing component. Some substance will be added below to the above-quoted statement, but not nearly enough to make the rest of the thesis fully comprehensible for someone not already reasonably familiar with the subject. Any such reader is referred to the text by Andrews et al. (1987), reviews such as McIntyre (1991), Schoeberl and Hartmann (1991), and Holton et al. (1995), and the references contained therein. Conversely, any reader who is already familiar with the stratospheric dynamics and transport literature of the past few decades, and is in a rush, can safely skip this chapter. It is included solely for completeness.

## 2.1 Basic facts

The stratosphere is of course strongly stratified. This means that vertical motions of any sort are constrained to be weak, and irreversible ones require a relatively large amount of diabatic heating. The extratropical circulation is westerly in winter and easterly in the summer, with temperature fields approximately in thermal wind balance. Further details of the climatological flow fields can be found in many sources, such as Andrews et al. (1987). Particularly notable for our purpose is the strong westerly "polar night jet" which roughly circles the pole, centered on average at about 60-70 degrees latitude (depending on hemisphere, season, etc.) and whose intensity increases with height up to the stratopause. The mass of cold air enclosed by and including the jet is called the "polar vortex".

The winter flow, including the polar vortex, shows large deviations from zonal symmetry, particularly in the northern hemisphere, while the summer flow does not. Charney and Drazin (1961) used basic theoretical arguments to show that planetary-scale Rossby waves, generated in the troposphere, can propagate upwards only when the winds are westerly, providing a clear explanation for this difference between summer and winter. Their arguments also showed that upward propagation is only possible for waves of the largest horizontal scales, explaining the observed fact that the deviations from zonal symmetry in winter tend to be planetary in scale, lacking the synoptic-scale variability associated with tropospheric weather.

The crude quantity used most often by dynamicists to represent the relative importance of radiation in the stratosphere is the "radiative time scale". This can be formally defined by considering the decay of a small-amplitude thermal disturbance superimposed on an otherwise steady basic state (e.g. Goody and Yung 1989, ch.10). In general, the time scale so defined depends strongly on the spatial scale of the thermal disturbance, being shorter for smaller-scale disturbances. The vertical scale is the one which matters, since it is generally constrained by the effects of stratification and rotation to be smaller by roughly Prandtl's ratio (coriolis parameter $f$ divided by buoyancy frequency $N$) than the horizontal scale. Because the vertical and horizontal scales of atmospheric disturbances are so strongly coupled, though, one can speak meaningfully of the radiative time scale of a disturbance with a given horizontal length scale.

For motions with horizontal length scales of order 1000 km or greater, throughout much of the stratosphere, the radiative time scale is long compared to the dynamical time scale, which might be defined as the length scale of the motion divided by a typical horizontal wind speed. For example, in the lower stratosphere the radiative time scale is typically estimated to be comparable to a month, whereas the time scale for a parcel of air moving at 30 $ms^{-1}$ to travel 10,000 km is about 3 days.

These constraints, namely stratification, rotation, and the relative slowness of radiative heating and cooling, imply that over their intrinsic dynamical time scales, extratropical stratospheric motions of large horizontal scale can be considered to be layerwise approximately two-dimensional on isentropic surfaces. This removes one of the two processes mentioned in the earlier quotation from Andrews et al. (1987), namely the meridional overturning, from consideration. Extreme advantage will be taken of this simplification in the remainder of this thesis, in that I will assume that a single-layer model can represent enough important features of the stratospheric flow to be a useful stand-in for that flow, for the purpose of studying large-scale isentropic mixing.

## 2.2 The meridional circulation

Over time scales longer than a month, the effects on tracers of the stratosphere's meridional overturning cannot be neglected. The general sense of this secondary circulation was first inferred by Brewer (1949) and Dobson (1956), and has since been referred to as the "Brewer-Dobson circulation". In an annual average, the Brewer-Dobson circulation rises in the tropics, splits and moves towards both poles, and sinks at middle and high latitudes. During solsticial seasons, the circulation occurs primarily between the tropics and the winter hemisphere. The meridional circulation in the summer hemisphere is relatively weak and poorly known.

Since the Brewer-Dobson circulation is thermally direct (at least in the winter hemisphere), it is tempting to explain it as thermally driven, analogously to the tropospheric Hadley cells. Because the stratospheric overturning extends to high latitudes, however, such an explanation is untenable. The reason for this can perhaps be most quickly understood by an argument such as the following.

Imagine the atmosphere to be zonally symmetric, so that angular momentum is conserved following the motion in the absence of sources and sinks. Outside the equatorial regions, the earth's rotation and spherical geometry induce a strong meridional gradient of angular momentum in the atmosphere. The Brewer-Dobson circulation requires a nonconservative mechanism for parcels to reduce their angular momentum so that they can move

poleward. The dissipation of planetary-scale Rossby waves can, if we suspend consistency long enough to allow their effects into our zonally symmetric atmosphere, provide such a mechanism, since these waves carry westward momentum up from the earth's surface. The circulation also requires diabatic (which in the stratosphere means, for the most part, radiative) heating and cooling so that the parcels can move irreversibly in the vertical. Through continuity of mass the poleward and vertical circulations are coupled, and so in a diagnostic sense one can think of the circulation as resulting from a balance between these two types of nonconservative process. The question thus naturally arises which process is in the more fundamental sense the cause of the circulation.

It is more or less universally agreed, as per the discussion above, that radiation in the stratosphere can act as a relaxation-like damping on the temperature field. Any externally imposed change to an air parcel's temperature (away from some steady state) will be eventually damped out through an exchange of photons between the parcel and the rest of the atmosphere, the earth's surface (or clouds which may obscure it), and space.

There is no analogous mechanism for damping out an externally imposed perturbation to a parcel's angular momentum, other than the exceedingly weak one provided by molecular friction. Since the source of upward-propagating planetary waves is outside the stratosphere, and the momentum source due to these waves has only one sign, that source is most sensibly viewed as externally imposed from the stratosphere's point of view. Hence the angular momentum budget is more strongly constrained than the heat budget is. This is the reason for the by now standard dynamicists' assertion that *the Brewer-Dobson circulation is driven by waves*. The associated radiative signature is taken to be a response rather than a cause. While there are still some who are uncomfortable with this idea, as far as the extratropical branch of the circulation goes no coherent challenges to it have been made. On the other hand, recent work indicates that rather different dynamics may have an important role in the tropics (R. A. Plumb, personal communication).

## 2.3 Quantitative techniques; coordinate systems

It is desirable to make the qualitative discussion above more quantitative. In particular, the interaction between the planetary waves and the posited axisymmetric flow was described in a very fuzzy way. When the flow is not very close to axisymmetric, there is no single most obvious way to separate the "mean flow" from the "waves" (also sometimes called "eddies", or "disturbances"), in order to understand the interaction between these two aspects of the flow. A number of approaches to this problem of "wave-mean flow interaction" have been developed, a subset of which will be briefly reviewed here.

### 2.3.1 Zonal mean, transformed Eulerian mean

The most obvious and straightforward approach is to simply zonally average the equations of motion. This is the version of the standard Reynolds average which is appropriate for spherical geometry. The waves are then defined as local deviations in any quantity from that quantity's zonal average. This leads to the presence of "eddy correlation" or "eddy flux" terms in the equations for the zonal mean quantities.

For reasons to be discussed shortly, it has proven advantageous in stratospheric dynamics to work with versions of the zonal mean equations which have been altered in a particular way, originally proposed by Andrews and McIntyre (1976). These altered equations are known as the *transformed Eulerian mean* (TEM) equations. The TEM equations contain

only zonally averaged quantities. The difference between the TEM and standard zonal mean equations is that the mean vertical and meridional velocities are redefined to include certain eddy correlation terms (see Andrews et al. 1987, eq. 3.5.1, p. 128). The mean velocity vector in the meridional plane, thus redefined, is known as the *residual mean velocity.*

The advantage of this redefinition is twofold. First, it isolates all interactions between zonal mean quantities and deviations from the zonal mean ("waves") into a single quantity, appearing in the zonal momentum equation, which can be written as the divergence of a flux, known as the *Eliassen-Palm flux*, or EP flux. This flux has different explicit forms depending on the specific geometry (spherical, local cartesian) as well as the particular form of the equations used (primitive, quasi-geostrophic etc.) This is a substantial simplification over the zonal mean case in which multiple, apparently unrelated, eddy correlation terms appear. Additionally, the EP flux divergence can be shown to vanish if the waves are small-amplitude, steady, and conservative and the underlying basic flow is purely zonal. The eddy correlation terms in the unmodified zonal mean equations do not vanish in this case.

Second, under assumptions which are restrictive but not so restrictive as to be irrelevant, the residual mean velocity is equal to the *effective transport velocity* defined by Plumb and Mahlman (1987), which at least under idealized conditions is the relevant velocity for studying tracer transport in the meridional plane. Recent attempts to diagnose the detailed structure of the residual mean circulation of the middle atmosphere include those of Rosenlof and Holton (1993), Rosenlof (1995) and Eluszkiewicz et al. (1996).

Under quasi-geostrophic assumptions on a $\beta$-plane, it can be shown straightforwardly that the EP flux divided by the mean density is equal to the northward eddy flux of quasi-geostrophic PV (Andrews et al. 1987, eq. 3.5.9, p. 130). This means that at least to a useful degree of approximation, we can rephrase our earlier assertion to say that *horizontal PV mixing drives the meridional circulation* of the stratosphere. Understanding how such mixing occurs is therefore essential to a complete understanding of the stratospheric circulation. It is important to remember that the PV flux must be given in the TEM equations; any closure assumption for relating that flux to mean quantities (such as the mean northward PV gradient) is external to TEM theory itself.

### 2.3.2 The generalized Lagrangian mean

Another framework which has been developed is known as the *Generalized Lagrangian Mean* (GLM), developed by Andrews and McIntyre (1978). In this case the "mean" is found by averaging over a set of fluid parcels, which may be in motion, rather than over a fixed spatial region such as a latitude circle. This device allows results concerning the interaction of mean and disturbance quantities which are both simple and exact, avoiding the restrictive assumptions typically associated with the TEM system. This is not a minor point, as some of those assumptions — perhaps most obviously that of small wave amplitude — are routinely and strongly violated in the stratosphere. The GLM equations are very difficult to apply in practice, however, and in addition can effectively break down for sufficiently large disturbance amplitude. As a result, despite their clear value from a theoretical perspective, they have seen very little use since their development.

### 2.3.3 Modified Lagrangian mean

The modified Lagrangian mean (MLM) coordinate was originally suggested by McIntyre (1980), developed from a practical or "synoptic" perspective by Butchart and Remsberg

(1986), and has recently been developed from a more theoretical perspective by Nakamura (1995, 1996). The surfaces of this coordinate are formed by isopleths of quasi-conserved tracers. MLM diagnostics are used to a considerable extent in this thesis. Detailed review of them is postponed to chapter 4.

## 2.4 Tracer transport

The Brewer-Dobson circulation acts to cause meridional (north-south) gradients on a given isentropic surface of tracers which have either sources or sinks at levels higher than the surface in question (obviously, the sources and sinks directly cause gradients at the levels where they act). Examples are chemical tracers such as $N_2O$ or $CH_4$. Horizontal mixing acts to reduce or eliminate such gradients. The competition between these two processes determines the slopes of tracer isopleths. Species which have no significant sources or sinks in a given region can be expected (and are observed) to have identical isopleth slopes in that region, as explained by Plumb and Ko (1992). The results from that study, as well as from the follow-up by Plumb (1996), allow powerful inferences about transport to be drawn from collections of simultaneous pointwise observations of multiple chemical tracers, even though these may have sampled only a very small fraction of the stratosphere. Arbitrarily chosen recent examples, relating to the currently important question of tropical-midlatitude exchange in the stratosphere, are the studies of Volk et al. (1996) and Minschwaner et al. (1996). Such studies demonstrate the advantages of working in "tracer space".

### 2.4.1 Planetary wave breaking

Nonetheless, since we do not live in tracer space, it is ultimately necessary to understand how tracers, including PV, are mixed in physical space along isentropes. In the extratropics in winter, *planetary wave breaking* (McIntyre and Palmer 1983, 1984, 1985)is believed to play an essential part. There is a large observational. modeling, and theoretical literature on the dynamics of breaking planetary waves and the transport they induce. There has been some argument over whether the term is an appropriate one (Rood 1985; McIntyre and Palmer 1985). The view taken here is that it is useful as a heuristic concept for qualitative discussion, though more precise terminology is needed for quantitative purposes.

Essentially, the idea is that in the absence of waves, all tracer isopleths would lie parallel to latitude circles. However, planetary waves distort the isopleths into other shapes. When the waves reach large enough amplitudes, they break. This means, in practice, that the tracer contours fold over on themselves. The folded regions tend to be stretched into "filaments" by the shear and strain in the flow, growing ever thinner, until finally some small-scale diffusion-like process (at the moment unspecified) finally causes them to lose their identities by mixing with their immediate environments. Various aspects of this process have been studied directly in observations (e.g. McIntyre and Palmer 1983, 1984; Leovy et al. 1985; Dunkerton and Delisi 1985; Baldwin and Holton 1988), in single-layer models (e.g. Juckes and McIntyre 1987; Juckes 1989; Salby et al. 1990; Polvani and Plumb 1992; Bowman 1993; Norton 1994; Waugh et al. 1994b; Polvani et al. 1995), and in three-dimensional models (O'Neill and Pope 1988; Saravanan and Dritschel 1994; Polvani and Saravanan 1997; Waugh and Dritschel 1997). Recently, many investigators have used observational analyses of winds and tracers to produce high-resolution reconstructions of episodes of planetary wave breaking. Most of these have used Lagrangian trajectory methods (e.g. Pierce and Fairlie 1993; Waugh et al. 1994a; Plumb et al. 1994; Dahlberg and Bowman

1994; Chen 1994; Manney et al. 1994; Norton and Chipperfield 1995; Eluszkiewicz et al. 1995; Schoeberl and Newman 1995). Eulerian or semi-Lagrangian advection schemes, in particular that of Prather (1986), have also been used (Orsolini et al. 1995, 1997; Wauben et al. 1997).

Planetary wave breaking is observed to occur in some regions but not others. This leads to the well-mixed "surf zone" (McIntyre and Palmer 1984), or region of weak large-scale tracer gradients, being bounded on the poleward side by the polar vortex edge. Because the mixing at the surf zone comes (to at least a first approximation) to an abrupt end at the latter boundary, the interior of the polar vortex is isolated from its exterior, a fact which has important and well-known implications for ozone chemistry (e.g. Schoeberl and Hartmann 1991). All definitions of the surf zone and the vortex edge have a certain amount of arbitrariness, but a plausible and typical definition of the boundary is that of Nash et al. (1996). Those authors define the polar vortex "edge region" as the region of large PV gradients, and its outer and inner boundaries as the places where, in an average around the PV contours, the PV gradient changes most sharply[1]. Issues related to this are also discussed by SPW. On the equatorward side, the surf zone is bounded by the somewhat less well understood "subtropical edge", or "subtropical barrier". Evidence for the existence of the latter is briefly reviewed in Polvani et al. (1995, hereafter PWP).

Studies such as those mentioned above have led to considerable insight into the way in which planetary wave breaking mixes chemicals and PV. However, the understanding gained has been primarily qualitative. For example, it is probably not possible at present to derive from observations some quantitative measure of PV mixing which could be combined with knowledge of nonconservative PV sources (i.e. from radiative calculations) to yield an accurate prediction of the width of the surf zone or the size of the polar vortex.

Likewise, while the debate about the degree of chemical isolation of the polar vortices (Arctic and Antarctic) seems to have died down recently[2], presumably due to a majority agreement that both are much closer to being "containment vessels" than "flowing processors", a close reading of the few papers which have attempted to address the question quantitatively through kinematical calculations could lead one to think that the question was still open to debate[3]. Evidence for the isolation of the vortex comes also from sources other than kinematical calculations, such as *in situ* measurements of chemical species (e.g. Hartmann et al. 1989), or gross dynamical constraints (McIntyre 1995). However, given the existence of reasonably reliable stratospheric wind analyses in the extratropics, one might hope that use of these to study the mixing directly in a *forward* sense could form a large part of the basis for a quantitative understanding of stratospheric mixing. Inferring details of the mixing from chemical tracer observations must always involve an *inverse* problem in some sense, in that the mixing one is studying is that which has already occurred rather than that which is happening presently, and that mixing more or less by definition involves a loss of information.

With a bit of tuning, general circulation models (GCM's) produce simulations of the winter stratosphere which are reasonably good, but simulation is not the same as understanding. In fact, describing quantitatively and concisely the mixing in a GCM is not much

[1] that is, where the second derivative of PV with respect to *equivalent latitude* reaches local maxima in absolute value — equivalent latitude will be defined later.

[2] this debate was intense for a while. See, for example, Hartmann et al. 1989, Tuck 1989, Tuck et al. 1993, and McIntyre 1995.

[3] See SPW for further discussion of the issues here and review of the relevant studies.

easier than doing so from observations. One avoids problems associated with the measurement and analysis processes, but some of the most limiting difficulties are due to a lack of appropriate theoretical and technical tools for describing the mixing in a robust and quantitative way. We do not understand the mixing process well enough to know how to cook it down to a small set of useful numbers. The difficulties which prevent the simplest and most familiar approaches from being adequate to the task are discussed in the next two chapters.

## 2.5 Small-scale mixing

The discussion so far has assumed that the isentropic mixing of tracers is due primarily to breaking planetary waves. One might reasonably ask what role, if any, is played by small-scale processes such as 3D turbulence.

### 2.5.1 Filament breakdown

In the discussion of wave breaking above, it was assumed that when the filaments produced by that process reach sufficiently small scales, they are dissipated by some process. Recent work has shed some light on how this might occur. Haynes and Ward (1993) argued that radiation would act quickly to damp out the temperature structure associated with small-scale PV anomalies, leading to the dissipation of those anomalies. Notice that this mechanism does not similarly eliminate chemical tracer anomalies. Theoretical investigations by Prather and Jaffe (1990) and Haynes and Anglade (1997) show that in the presence of vertical shear, PV and tracer filaments can be expected to be rapidly strained to very small vertical scales, enabling even a very weak diffusion (whether eddy or molecular) to disperse them after a modest amount of time. The latter of those two studies indicated that the vertical straining is, however, slaved to the horizontal straining, so that despite that the diffusion ultimately occurs in the vertical, two-dimensional (i.e. single-layer isentropic) tracer fields can be used to diagnose the process. Waugh et al. (1997) have studied the decay of vortex filaments in data from the NASA ER2 aircraft, using the model of Prather and Jaffe (1990) to derive an effective diffusion coefficient. A similar study of the same data has also been carried out by Balluch and Haynes (1997).

These studies suggest that vortex filaments can reach a minimum horizontal scale of about 10-100 km before diffusion becomes important. Whether the diffusion is purely molecular, or whether some dynamical mixing process plays a role, is still not entirely clear. Since the filaments are locally isolated strips of PV, one might naively expect them to be barotropically or baroclinically unstable, which would presumably accelerate their breakdown. However, analytical and numerical studies (Dritschel 1989b, Waugh and Dritschel 1991, Dritschel and Polvani 1992) suggest that at least as long as they remain in the extratropics, the filaments will be stabilized by the strong shear induced by the nearby polar vortex.

### 2.5.2 Gravity wave breaking, etc.

It is sensible to wonder whether there is some other source of turbulence in the stratosphere which might lead to significant mixing, apart from that which dissipates filaments. A number of studies relevant to this question have been carried out by applying ideas from turbulence and gravity wave theories to the analysis of horizontal wave number spectra of

temperature, winds, and tracers derived from ER2 observations (Bacmeister et al. 1990, 1992, 1996; Bacmeister 1993; Strahan and Mahlman 1994). Taken collectively, these studies suggest that any small-scale turbulence which may be present in the stratosphere is not strong enough, most of the time, to cause tracer fluxes comparable to those due to large-scale processes. The exceptions to this are spatially and temporally intermittent episodes of the breaking of topographically forced gravity waves. The relative rareness of these events means that their collective effect can by no means be modeled by a diffusivity that is either temporally or spatially uniform. It seems possible that the only satisfactory way to account for them is by representing them explicitly, for example using the parameterization proposed by Bacmeister et al. (1994).

### 2.5.3 The bottom line

In the remainder of this thesis, I will take the following to be true, based on the current state of knowledge as reviewed above.

1. Mixing of tracers in the stratosphere by small-scale processes is weak, in general.
2. What small-scale mixing there is is distributed unevenly in space and time.
3. The mechanisms causing the small-scale mixing are not entirely understood.

The net implication of these points is that a detailed, accurate and quantitative accounting of tracer transport due to small-scale mixing is either impossible or at least impractical, except on a more or less case study basis, such as when high-resolution *in situ* data are available. This implies that any technique for diagnosing large-scale mixing which depends on accurate knowledge of small-scale mixing has a poor observational basis.

The role of small-scale mixing, its interaction with large-scale dynamics, and the extent to which the details really need to be known will be considered further in chapters 4 and 6.

# Chapter 3

# Eddy Transport

## 3.1 Introduction

This chapter contains a general theoretical discussion of transport by eddies. The emphasis will be on transport in bounded domains, and on situations in which the largest eddies are comparable in size to the domain itself.

Section 3.2 defines the problem to be addressed in the remainder of the chapter. This is not merely a formality. One can find much previous work which deals in some way with the issues of interest here, and I will of course make use of existing results. But to my knowledge the central questions have not been addressed in an organized way from the perspective taken here, which is a purely *kinematic* one. Taking the kinematic view allows a clear analysis of some important issues, though it entirely excludes others from consideration. Section 3.3 presents G. I. Taylor's (1921) derivation of a diffusion equation for tracer transport by isotropic, homogeneous turbulence. This illustrates nicely the definition of the problem given in section 3.2. Taylor solves the problem in a particular asymptotic limit. Section 3.4 discusses integral or matrix representations of eddy transport (which are continuous and discrete mathematical expressions of the same idea). These ideas have seen the most use in the field of boundary layer meteorology, in particular by Roland Stull and colleagues. Some theoretical aspects of the matrix representation are explored here, and I argue that it is the most general possible model of eddy transport which one can formulate which retains certain desirable properties. The asymptotic solution obtained by Taylor is contained within this representation as a special case. Section 3.5 discusses a very simple representation of transport by large eddies, which may be derived as an approximation to or truncation of the transilient matrix model. A separate, heuristic derivation of the same simple representation is also presented. This representation may be thought of as being valid in the opposite limit to diffusion. It is shown that its effects cannot be mimicked, except in the crudest sense, by a diffusion model, even if the diffusivity is allowed to be an arbitrary, tunable function of space. Section 3.6 summarizes, based on the material in the preceding sections, the argument that diffusion is an inappropriate model for mixing by large eddies, and that alternatives are preferable. Section 3.7 discusses the relevance of some of these ideas to quasi-horizontal mixing in the stratosphere specifically, and uses a heuristic argument to illustrate in another way the relationship between the diffusion and large-eddy limits.

## 3.2 Definition of the problem

### 3.2.1 Advection

Consider an incompressible fluid flow $\mathbf{u}(\mathbf{x},t)$. For maximum simplicity the flow will be taken to be two-dimensional so that $\mathbf{x} = (x,y)$ and $\mathbf{u} = (u,v)$. The fluid carries a tracer $q(\mathbf{x},t)$. That is, the tracer obeys the Eulerian equation

$$\frac{\partial q}{\partial t} + \mathbf{u} \cdot \nabla q = S \tag{3.1}$$

where $S$ represents any sources or sinks of the tracer. Strictly speaking one should include a true (i.e. molecular) diffusion term on the right hand side of (3.1), but I will assume that the Reynolds number is very large. Then diffusion may be ignored, and tracer transport considered entirely determined by $\mathbf{u}$. A physically equivalent, but mathematically rather distinct way of representing the tracer evolution in this case is the Lagrangian formulation

$$\frac{d\tilde{\mathbf{x}}_i}{dt} = \mathbf{u}(\mathbf{x}_i,t) \tag{3.2}$$

where $\tilde{\mathbf{x}}_i$ is the location of a particular fluid parcel that has been "tagged" at some initial time with the label $i$ and with a particular value of the tracer $q_i$. That value evolves according to

$$\frac{dq_i}{dt} = S(\tilde{\mathbf{x}}_\mathbf{i},t)$$

With $S(\tilde{\mathbf{x}}_\mathbf{i},t)$ still the source term but now evaluated at the position of parcel $i$. Each fluid parcel is in principle considered infinitesimal in size, and in order for (3.2) to contain as much information as (3.1) the former must be solved for an infinite number of parcels. In numerical computations both representations must of course become finite-dimensional and approximate, and each has its advantages.

Throughout this discussion, (3.1) and (3.2) will be considered from a purely *kinematic* point of view. That is, $\mathbf{u}(\mathbf{x},t)$ will be taken to be given[1]. Strictly speaking, this means that $q$ will be considered a passive tracer. However, this approach may be extended to study the advection of active tracers as well, at least in an initial value sense, if the initial conditions and source terms are constrained to be only those which are consistent with the flow $\mathbf{u}$. That is, the relation between $\mathbf{u}$ and $q$ is taken to be already built in, and the advection is studied in a diagnostic sense. For a truly passive tracer, such as a chemical species whose reactions do not release or consume energy in sufficient quantities to noticeably affect the

[1]Due to the wide range of scales typically present in atmospheric motions, it may seem an excessive idealization to assume that the flow is known entirely down to the molecular scale. The above statement may be qualified to take this into account, with little damage to the ensuing arguments. It is assumed that we know the flow field with sufficient resolution so that (1) may be solved with an artificial diffusion or diffusion-like term to mop up tracer features below some certain scale, without introducing significant inacuraccy into the representation of larger scale tracer features. This is essentially the thinking behind the "Large Eddy Simulation" (LES) technique of modeling turbulent flows. There are some sticky issues concerning subgrid-scale (SGS) closure, for example see Kraichnan (1976) and Lilly (1989). Nonetheless I will assume that the transport by large eddies is insensitive to the details of the small eddies' behavior. This assumption is not necessarily always valid, especially near boundaries. However, since all questions about SGS closures must in any case become inconsequential at sufficiently high resolution, the reader who is concerned about them may simply imagine that the flow is known with resolution high enough to allay his or her concerns.

flow, any set of initial conditions, boundary conditions, and source terms is consistent with the problem since none of these affects **u**. This may seem a trivial observation, but it is important to what follows.

If we are given **u**, $S$, boundary conditions, and (for the initial value problem) initial conditions, we can in principle solve either (3.1) or (3.2) numerically without any fundamental difficulty. Now, though, let us take some form of spatial average of $q(x, y, t)$ in order to obtain a field, $\overline{q}$, which depends on only one spatial coordinate. The simplest way to do this is to perform a standard Reynolds average over one of the "natural" spatial coordinates, say $x$, to obtain a field $\overline{q}(y, t)$. However, circumstances can arise in which it is advantageous to average over some other coordinate, such as one defined by curves in $x$-$y$ space, which may even move in time. The distinction is not crucial to anything in this chapter, so $\overline{q}$ may be taken as a simple average over $x$ here. For the average over $x$, (1) becomes

$$\frac{\partial \overline{q}}{\partial t} + \overline{v}\frac{\partial \overline{q}}{\partial y} + \frac{\partial}{\partial y}\overline{v'q'} = \overline{S} \tag{3.3}$$

where primes represent deviations from the average. The third term on the left hand side of (3.3) will be called the *eddy flux divergence*, *kinematic flux divergence*, or when no confusion is possible, simply the *flux divergence.*

### 3.2.2 Two steps to parameterization

The problem to be discussed here is how to represent the eddy flux divergence by some form of operator on the mean tracer field $\overline{q}(y)$. It is crucial to make clear exactly what this does and does not entail.

In models of many sorts it is necessary to *parameterize* one or more physical processes. The need arizes because the process affects the rest of the model dynamics in a major way, yet the process is too complex to be represented explicitly in a realistic manner. What one needs then is a "sub-model" whose inputs are the variables resolved explicitly by the larger model, and whose outputs represent the effects of the process being parameterized. Arakawa (1994), in a general discussion of issues involved in parameterization, points out that most parameterizations are *diagnostic*[2]. A diagnostic parameterization depends only on the instantaneous values of the resolved variables in the main model, and only represents those aspects of the parameterized process which affect the resolved model variables. In other words, the parameterized process is not explicitly predicted, by a separate prognostic equation, nor is any information about its past behavior used. To limit the scope of discussion, I will hereafter consider only diagnostic parameterizations.

The process of parameterizing the flux divergence term in (3.3) can be broken into two distinct steps. The distinction being made here is not always (or even commonly) made in describing parameterizations of this term, and the two separate steps may not be immediately discernable in all derivations or justifications of such parameterizations. Nonetheless, at least most of the time, one can identify them after the fact. The two steps are:

1. Develop a *transport model.* This is distinct from the larger model within which the parameterization will function. It is an equation, set of equations, or other set of rules

[2] Arakawa strictly refers to parameterizations of cumulus convection, but the statement is true of parameterizations of other physical processes as well.

for determining the desired output, that is, the eddy flux divergence, from the *mean* tracer field. The transport model will contain some number of *parameters* to which values must be assigned.

2. Hypothesize a relationship between the explicitly resolved model fields (other than the tracer field) and the needed parameters. This is *closure*.

My use of terminology is not quite standard. The term "closure" is sometimes used to encompass both steps.

Note that step (2) makes sense only within the context of a particular choice of model as per step (1). For example, consider an atmospheric boundary layer problem with both shear and buoyancy effects. Discussions of this can be found in textbooks such as those by Stull (1988) and Garratt (1992). A simple and common approach to this problem is the use of a diffusion model to parameterize the eddy fluxes (step 1). This is often called "K-theory", or, "first-order closure", though according to the definitions of terms given above this is not the closure itself but the necessary precursor to closure. Perhaps one might say that this choice defines a certain class of closures. Even after choosing the diffusion model, without some way to specify the diffusivity as a function of the mean flow, the problem is not in fact closed. In any case, diffusion is a valid model if the turbulent eddies have spatial scales small compared to the scale on which the mean shear varies (as will be discussed in more detail shortly). This is a kinematic condition. Then, knowing something about the dynamics behind the turbulence, a simple and common approach is to postulate a relationship between the diffusivity and some form of Richardson number (step 2). This is the closure assumption, and necessarily involves some knowledge of the dynamics.

In designing our operator which will give us $\frac{\partial}{\partial y}\overline{v'q'}$ in terms of $\overline{q}$ we are only tackling step 1. This kinematic step can be isolated from any consideration of dynamics because of the assumption that the flow is fixed and known. Thus having formulated the transport model, one is free to use this knowledge of $\mathbf{u}$ — in its full, un-averaged form — to obtain the values of the parameters. If we know $\mathbf{u}$, why not just compute $\overline{v'q'}$ directly? The answer is that if the model is to be a consistent description of the flow's transport properties, a *single* operator is required, dependent on some statistics of $\mathbf{u}$ only, which will tell us the eddy flux divergence of *any* tracer (or more realistically, any of some not too narrow class of tracers) once that tracer's mean profile, $\overline{q}$, is given. The correlation $\overline{v'q'}$ clearly cannot be computed directly without knowledge of $q'$.

Once the desired operator, or transport model, has been obtained, the parameters can be computed for particular observed or simulated flows. The model can then be used without closure as a diagnostic tool for studying various aspects of the behavior of those flows, for example the way in which transport by those flows interacts with chemistry to determine the distributions of chemical species. The model can also be used as a basis for a parameterization, in which case the closure problem must be faced. The degree to which any closure scheme can succeed is clearly constrained, however, by the appropriateness of the model on which it is based. Because of this it seems wise to understand the inherent limitations of various transport models before attempting closure.

Again, the kinematic approach has the advantage that it allows us to approach this question independently of the flow dynamics. We can learn to understand how a flow transports material simply by looking at the flow itself, without worrying about whether the eddies are produced by shear, buoyancy, rotational effects, etc. We can consider the differences and similarities between flows whose dynamics are entirely different. If we are fortunate we may gain intuition, and perhaps even specific tools, of broad applicability.

### 3.2.3 Condensed descriptions

To compute the eddy flux directly, one needs $\mathbf{u}$ and $q$ at sufficient resolution so that the transporting eddies are well represented. The number of necessary data is then on the order of the number of grid points in the full 2D grid. The number of parameters required for any useful transport model must be much smaller than this; a useful transport model is a condensed description. Because of this, each model is guaranteed to fail in some situation, simply because some degrees of freedom which are essential to the true system's behavior in that situation have been truncated in the model. In general, one expects that the more complex the transport model, and the larger the number of parameters, the wider the range of situations in which it is at least theoretically capable of being accurate. The best transport model for a given problem is the one that strikes the optimal balance between the conflicting demands of simplicity and accuracy. It is reasonable to suppose that no model will be optimal for all problems, so that it is worthwhile to consider different transport regimes. By a regime I mean a broad class of flows $\mathbf{u}$ which share some loosely defined characteristics.

There can be situations in which no transport model can do a good job. For example, consider a tracer whose variability in the $x$ direction is as large as or larger than that in the $y$ direction. In this case $\overline{q}$ simply cannot tell us enough of what we need to know; the true answer depends strongly on the information that has been left out. In some cases this may be ameliorated by choosing a different coordinate over which to average, in others perhaps not. In any case, in the remainder of this chapter it will be assumed, unless indicated otherwise, that $\mathbf{u}$ and $S$ are such as to cause the tracer to be, on average, substantially more homogeneous in one direction than the other.

## 3.3 Taylor's derivation of the eddy diffusion equation

In a classic paper, G. I. Taylor (1921) showed that transport by isotropic, homogeneous turbulence could be modeled by a diffusion equation, which can be considered a prototypical transport model (it is certainly the most popular). The only parameter is the diffusivity, which Taylor showed to be a function of certain statistical properties of the flow. Taylor's theory is particularly valuable because these statistical properties, which are all that the theory requires, are rigorously definable and experimentally measurable. This should be compared with the theory of Prandtl (1925)[3], which relies on a fictitious "mixing length". However, Taylor's derivation (and all which have come after) has the important limitation that the diffusion equation obtained is valid only for time and space scales large compared to the largest eddies. This point is so central here that it is appropriate to go through the most important steps in Taylor's paper in order to justify it.

### 3.3.1 Derivation and key results

Taylor starts with a derivation of "normal" diffusion by particles undergoing a random walk on a one-dimensional, discrete lattice. Each time step $i$, a particle moves a distance $x_i$ which is equal to either $+d$ or $-d$, with a 50 percent probability of either one. Let $X_n$ be the root mean square distance, in the sense of an ensemble average over many realizations, moved

[3] as reviewed, e.g., by Sutton (1949).

by a particle after $n$ time steps. Then

$$X_n^2 = [(x_1 + x_2 + ...x_n)^2]$$

where a square bracket represents an expectation value, or average over the ensemble of all possible realizations. Since the probability of the particle's moving forward or backward at each step is entirely independent of what happened at the previous step, all cross terms $2[x_i x_j]$ vanish for $i \neq j$, so we have

$$X_n^2 = nd^2$$

or

$$X_n = dn^{1/2}$$

Since the total elapsed time is simply $n\Delta t$ where $\Delta t$ is the time step, the familiar result is that the RMS displacement increases as the square root of the time. This implies that the expectation value of the particle density, call it $\rho$, obeys a diffusion law with a constant diffusivity

$$\frac{\partial \rho}{\partial t} = K \frac{\partial^2 \rho}{\partial x^2}$$

where $x$ is the fixed Eulerian coordinate. The diffusivity, $K$, in this case is equal to $d^2 \Delta t^{-1}$, or $v^2 \Delta t$ where $v \equiv d\Delta t^{-1}$.

Taylor next modifies the previous case so that at each time step the direction of the particle's motion is correlated with its motion at the last step, by a correlation coefficient $c$. Partial correlations across more than one time step are zero, so that $[x_i x_{i+2}] = c^2$, etc. This leads directly to

$$X_n^2 = nd^2 + 2d^2((n-1)c + (n-2)c^2 + ... + c^{n-1})$$

Summing the series, and making the substitutions $t = n\Delta t$ and $d = v\Delta t$ leads to

$$X_n^2 = v^2\{(\frac{1+c}{1-c}\Delta t)t - \frac{2c(1-c^n)\Delta t^2}{(1-c)^2}\}$$

Since Taylor's interest is, as per the paper's title, ultimately in diffusion by continuous movements, the natural next step is to take the limit as $\Delta t \to 0$. All quantities can be kept finite in this limit only if $1 - c$ decreases linearly with $\Delta t$, that is

$$\lim_{\Delta t \to 0} \frac{\Delta t}{1-c} = \tau$$

where $\tau$ is a constant, so named (here, not by Taylor) because it has the dimension of time. The result, dropping the subscripts $n$, is

$$[X^2]^{1/2} = v\{2\tau t - 2\tau^2(1 - e^{t/\tau})\} \tag{3.4}$$

When $t$ is small, expanding the exponential to second order gives

$$[X^2]^{1/2} = vt$$

that is, the RMS particle displacement grows in a manner that is mathematically more like

*advection* than diffusion. For large $t$, however, the asymptotic result is

$$[X^2]^{1/2} = v(2\tau t)^{1/2} \tag{3.5}$$

or a diffusion with diffusivity $2v^2\tau$.

Despite the simplicity of Taylor's above calculation, it contains an essential insight into the nature of turbulent transport. Taylor notes that (where I have changed only the labeling of mathematical symbols to be consistent with my usage here)

> The constant $\tau$ evidently measures the rate at which the correlation coefficient between the direction of an infinitesimal path in the migration and that of an infinitesimal path at a time t, say, later, falls off with increasing values of t.

$\tau$ is a sort of eddy turnover time, and it determines how long one has to wait before the asymptotic result (3.5) becomes valid. It is important to understand the physical meaning of the limit process as $t \to \infty$. Taylor cannot change the eddy turnover time for a given flow, but he can and does choose to study the flow's transport properties only over time scales long compared to the eddy turnover time.

Taylor next goes on to consider diffusion by true continuous movements. He first obtains some general results for statistical properties of continuously varying quantities. The results of this will be taken for granted here; for the most part they consist of the definition of the correlation coefficient and the derivation of its properties.

Following this Taylor considers isotropic, homogeneous turbulence. $u$ is the velocity component in the $x$ direction, and the statistical quantities $[u^2]$ and $R_\xi$ are known. $R_\xi$ is "the correlation between $u$ for a particle at any instant, and the value of u for the same particle after an interval of time $\xi$",

$$R_\xi = \frac{[u(t)u(t+\xi)]}{[u^2]}$$

where $u(t)$ is evaluated following a fluid parcel, so that $R_\xi$ is a Lagrangian quantity. By the definition of $R_\xi$, the result (proved by Taylor) that it is an even function of $\xi$, and a few more simple steps, Taylor obtains the result

$$[X^2] = 2[u^2] \int_0^t \int_0^{t'} R_\xi d\xi dt' \tag{3.6}$$

for the displacement $X$ of a particle in the turbulent flow after a time $t$.

The asymptotic behavior of (3.6) is the same as that of (3.4); the only difference is that the correlations between successive infinitesimal displacements are now allowed to have any form, except that they are assumed to vanish for displacements separated by sufficiently large time increments. Because of this, Taylor can define a limit $I$

$$\lim_{t\to\infty} \int_0^t R_\xi d\xi = I$$

For times short enough that $R_\xi$ does not drop significantly from unity, (3.6) becomes

$$[X^2]^{1/2} = [u^2]^{1/2} t \tag{3.7}$$

while for times much larger than that required for $R_\xi$ to drop to zero,

$$[X^2]^{1/2} = (2[u^2]It)^{1/2} \tag{3.8}$$

which is the limit of true "eddy diffusion".

Later investigators called $I$ the "Lagrangian integral time scale". It has an Eulerian counterpart which is defined similarly, except that the correlation coefficient used is that between the velocity at two different times but the same point in physical space, rather than following a particle. These time scales are probably the most widely used quantitative definitions of an eddy turnover time. In a later paper, Taylor (1935) defined the length

$$l = [u^2]^{1/2} I$$

which is now called the "Lagrangian integral scale", also having an Eulerian counterpart in which $I$ is replaced by the Eulerian integral time scale.

### 3.3.2 Scale separation in space

With these definitions, a limitation of the diffusion model can be seen which may not be immediately apparent from Taylor's analysis. To obtain the result (3.8), Taylor took the limit of large time, but did not make any explicit restrictions in the space domain. There is such a restriction nonetheless. (3.8) is valid only for $t \gg I$, but from that expression, and the definition of $l$, we can see quite clearly that in this case $[X^2]^{1/2} \gg l$. If we are in a *bounded* domain, whose size is comparable to $l$ so that *the largest eddies fill the domain*, then (3.8) can *never* be valid, no matter how long we wait. A derivation of the eddy diffusion model which is Eulerian, and makes the need for a scale separation in space more directly apparent, can be found in Frisch (1989). A derivation somewhat similar to Frisch's will also be presented later in this chapter.

This restriction on the spatial scale of the eddies is of great importance for geophysical fluid dynamics (GFD), since that subject often must concern itself with eddies of various sorts which do span their domains. Good examples are a convective tower which extends from the surface to the tropopause in the tropics, or more relevant here, a planetary wave which stirs the entire stratospheric surf zone in a single coherent circulation. These may be considered fundamentally *inhomogeneous* flows, as opposed to the homogeneous turbulence for which Taylor's analysis (and most of what is commonly known as turbulence theory) applies. A good discussion of the distinction, from a GFD point of view, can be found in Haidvogel and Held (1980). In the present discussion the adjective "inhomogeneous" is taken to be roughly synonymous with the phrase "containing domain-size eddies". Weakly inhomogeneous flows, in which the eddies are small but their statistical properties vary on spatial scales large compared to the eddies, are considered locally homogeneous and essentially similar to truly homogeneous flows.

Transport by inhomogeneous flows must generally be studied on a problem-by-problem basis, since the inhomogeneities are controlled by forcing and boundary conditions; no universality assumption can be applied.[4] However, this is not to say that there are not

[4] In fact, it may be argued that even the most isotropic and homogeneous turbulence in the real world is neither as isotropic nor as homogeneous as classical theory would have it, and that evidence for the universality of scalars in small-scale, high Reynolds number turbulence is not unshakeable; see Sreenivasan (1991). Nonetheless the distinction between this classical turbulence and the strongly inhomogeneous variety

some general concepts which may be at least loosely applicable to a number of different inhomogeneous problems. In the next section I will discuss the idea that transport due to inhomogeneous flows with domain-size eddies can be most generally described by an integral, rather than a differential equation.

## 3.4 Integral or matrix representations

### 3.4.1 Introduction and anecdotal history

If the diffusion model is bound to fail when the eddies are domain-size, what other model can we use? Batchelor and Townsend (1956) very briefly discussed the case in which the domain is still unbounded spatially, but one is interested in times of the order of the integral time scale; the difficulty is essentially the same one. They noted that

> When the probability distribution of $X$ does *not* have the normal form (as is frequently the case at values of $t - t_0$ that are not large) there is no reason to expect $Q$ and $P$ to satisfy *any* differential equation. Turbulent diffusion is not a local effect, since the rate of increase of dispersion of a particle about its initial position depends on the statistical persistence of the particle velocity over a finite time, and a description of the diffusion by some kind of integral equation is more to be expected.

The italics are original, and $X$ is a particle's displacement, $P(\mathbf{x}, t)$ is the probability of finding an initially marked parcel of fluid at location $\mathbf{x}$ at time $t$, and $Q(\mathbf{x}, t|\mathbf{x}', t_0)$ is the probability density function of the displacement $\mathbf{x}-\mathbf{x}'$ at time $t$ of the parcel which was at $\mathbf{x}'$ at $t_0$. The reader is referred to the original paper for more explanation. The main point here is that, while Batchelor and Townsend did not pursue the issue further, they recognized that in the regime of interest here an integral equation would be the most appropriate description of turbulent transport.

Roberts (1961) studied the diffusion of passive scalars in homogeneous turbulence, using the direct interaction approximation of Kraichnan (1959). Roberts obtained integral equations for the evolution of the probability distributions of both a single particle's displacement (Taylor's $X$) and the relative displacement of two particles, in the most general case in which the time is not restricted to being asymptotically small or large. However, in the interest of analytic progress, Roberts' analysis of the single-particle case (which is the one of most interest here) was restricted to limits of small and large time. in which a differential equation could be recovered.

Charney and Eliassen (1964) in their paper on hurricanes, noted that

> Since the time-scale of cumulus development is small in comparison with the time-scale of development of the depression, and the cumulus cells extend throughout the whole troposphere, the vertical transfer of latent heat by cumulus convection appears as an integral effect, more analogous to radiative than to convective or diffusive transfer.

One must assume that the word "convective" is meant in the sense synonymous with "advective" rather than in the sense associated with buoyancy-induced motions. The diffusion

is clearly worth making.

model generally fails in a particularly obvious and dramatic way to capture transport by atmospheric moist convection (where the word is meant in the latter sense), so that this is a rare transport problem for which the diffusion model has been little used.

### 3.4.2 Matrix representation: definition

The most concerted effort to develop a theoretical framework along the line suggested by Batchelor and Townsend's remark has been made in more recent years, predominantly by Roland Stull and his co-workers. Two independent but nearly simultaneous studies, those of Fiedler (1984) and Stull (1984) suggested that the effect of turbulent mixing in the vertical on a mean tracer profile in the planetary boundary layer should be modeled by an integral operator

$$\frac{\partial q(y)}{\partial t} = \int_0^{y_t} C(y|y')q(y')dy' \tag{3.9}$$

where $y_t$ is the top of the domain and $C$ is some kernel, or when discretized in time and space, by a matrix operator

$$\mathbf{q}(t + \Delta t) = \mathbf{C}\mathbf{q}(t) \tag{3.10}$$

Here $\mathbf{q}$ is an $n$-dimensional vector whose elements are the values of the tracer $q$ at the nodes of a one-dimensional lattice representing the (zonal. horizontal, etc.) mean profile,

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

where $y_b$ is the bottom of the domain so that $y_b + n\Delta y = y_t$ and $\mathbf{C}$ is an $n \times n$ matrix. It should be noted that the roles of the kernel $C(z|z')$ in (3.9) and the matrix $\mathbf{C}$ are not directly analogous due to the implicit form of (3.10), though they can easily be made so by rewriting (3.10) explicitly in terms of the tendency and redefining the matrix accordingly. In the implicit form, the matrix $\mathbf{C}$ was named the *transilient matrix* by Stull (1984). Fiedler (1984) derived the integral representation starting from the "spectral diffusivity" parameterization of Bercowicz and Prahm (1979). while Stull hypothesized it directly on intuitive physical grounds. In fact equation (3.10) had been used at least once prior to 1984 to parameterize convective mixing (Fung et al. 1983) but without much attendant discussion.

Equations (3.9) and (3.10) describe a process akin to the shuffling of a deck of cards. $\mathbf{C}$ 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 $\Delta t$. $\mathbf{C}$ is a function of the time step $\Delta t$; the nature and implications of this dependence will be discussed in detail below. It can also, of course, be a function of time itself.

### 3.4.3 Matrix representation: mathematical properties

The mathematical properties of (3.10) and related equations will be discussed in some detail, but it is appropriate to mention at this early stage the most important and basic ones. Definitions of formal terms concerning stochastic processes are taken from the text by Taylor and Karlin (1994). First, when the matrix $\mathbf{C}$ is known (as it is, in the kinematic approach) (3.10) 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, (3.10) describes a *Markov process*. Third, since no negative tracer mixing ratios can develop from an initial

distribution which is all positive, all elements must be nonnegative. Fourth, by conservation of fluid mass and of the tracer, the rows and columns of **C** must all sum to unity:

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

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

### 3.4.4 Down-gradient flux is more general than diffusion

A fifth property, which is not obvious mathematically but may corroborate physical intuition, involves the flux due to mixing by a transilient matrix. If 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)

$$F_k = \frac{\Delta y}{\Delta t} \sum_{j=1}^{k} \sum_{j'=k+1}^{N} C_{jj'}(\overline{q}_j - \overline{q}_{j'}) \tag{3.11}$$

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 *if the gradient of the tracer is everywhere of one sign, the flux is everywhere down-gradient* no matter what the details of the transilient matrix. Nonetheless 3.10 is not, in general, equivalent to a diffusion equation. This suggests that one ought to avoid the assumption that down-gradient transport and diffusion mean the same thing. It should also be noted that if the tracer gradient changes sign, then (3.11) can (depending on the details of the tracer profile and the transilient matrix) yield up-gradient fluxes, whereas a diffusion equation, of course, cannot.

### 3.4.5 Physically motivated assumptions, and consequences

In all that follows I will make two further assumptions about all physically realistic transilient matrices. The first is that they are *irreducible*, meaning that there are no absolutely impermeable transport barriers. In other words, given an infinitely long time, each region of the fluid "knows about" (formally, *communicates* with) each other region. The second is that there is no unmixing, 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 from this and the double stochasticity of **C** that the *limiting distribution* of the process defined by **C** is the uniform distribution, that is, a vector with all elements equal. This formalizes the intuitive notion that mixing, acting in the absence of other process, must lead eventually to a state in which all material properties are homogenized.

### 3.4.6 Usefulness

A fairly large body of work has been built up since 1984 on "transilient turbulence theory", most of it by Stull and various collaborators. A thorough review, along with some discussion of other nonlocal turbulent transport models and a good list of references, can be found in Stull (1993). Within the boundary layer community, the theory has not been widely adopted

for the purpose towards which it seems to have been primarily intended, that is, as the basis of a parameterization for boundary layer turbulence in mesoscale or global atmospheric models. To my knowledge a transilient model has been used to represent boundary layer turbulence in only one study (Raymond and Stull 1990) in which it was found to be very computationally expensive, and at best marginally superior to much cheaper models. [5]

However, with regard to the problem of interest here, the transilient matrix is a more appropriate tool. This is the case for two related reasons:

1. The matrix representation (3.10) can be derived directly as an approximation to the raw Reynolds-averaged advection equation (3.3), and is exact in certain special cases.

2. Unlike the eddy diffusion model, (3.10) is a very *general* model of eddy mixing. In particular, no assumption about the size of the eddies is built in. The diffusion model, which makes such an assumption, is a special case of (3.10).

Because of point 2, (3.10) may be a more complex framework than is ultimately needed to parameterize some particular type of eddy mixing once that type is reasonably well understood. However, if one wishes to start with a clean slate, and approach a particular flow without assuming that one already understands how it transports tracers, then the most general framework is best. More restrictive simplifying assumptions can always be made later, when one has some legitimate basis for them.

### 3.4.7 Derivation of the transilient matrix

Here I derive (3.10). To my knowledge this has not been done previously from first principles, though it is not difficult. Fiedler's (1984) derivation uses the spectral diffusivity parameterization, which is essentially *ad hoc*. Stull's justification of it is intuitive rather than being based on any direct derivation.

Consider the Eulerian advection equation ( 3.1), in a bounded domain with no flux of tracer through the boundaries. Consider a discretization of the tracer field in coordinate space, so that $q_{ij}$ is the value of the tracer at $x = i\Delta x$, $y = j\Delta y$ with $\Delta x$, $\Delta y$ the grid spacings in x and y. For convenience take $\Delta x = \Delta y$. Also, let the grid spacing be sufficiently small that each grid box may be considered homogeneous; there is no sub-grid scale variability in $q$.

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

$$q_{ij} = A_{iji'j'} q_{i'j'}$$

where primes represent values at the previous time step, or in more compact notation as

$$\mathbf{q}(t + \Delta t) = \mathbf{A}(t)\mathbf{q}(t) \tag{3.12}$$

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

[5] It seems likely that this lack of great success is due to the intrinsic difficulty of the closure problem. The transilient model might, with optimal values of the matrix elements, be more accurate than other models with optimal values of their parameters. However, when the matrix elements must be obtained through closure assumptions, rather than from direct computation from detailed, eddy-resolving knowledge of the flow within the boundary layer, the results are unlikely to be optimal. In this case it is unclear that the additional complexity of the transilient model should lead to additional accuracy.

the current time step, but are independent of $\mathbf{q}$, by linearity. Note that explicit diffusion, since it is linear, can be incorporated into $\mathbf{A}(t)$ as well. In fact, linear sources or sinks of $q$ could be included also, but for some purposes it desirable to exclude them. This is because if one is ultimately interested in understanding the interaction of sources and sinks with transport, one would like the transilient matrix to represent the transport only, rather than a mixture of the two processes. Prather (1994, 1996) describes modes which are eigenvectors of a matrix which includes both transport and linearized chemistry. The transilient matrix, which includes only transport, is a special case of Prather's matrix.

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

$$\mathbf{q}(t + m\Delta t) = \mathbf{A}(t + [m-1]\Delta t)\mathbf{A}(t + [m-2]\Delta t) \cdots \mathbf{A}(t)\mathbf{q}(t) \tag{3.13}$$

We can write (3.13) more compactly as

$$\mathbf{q}(t + m\Delta t) = \mathbf{B}(t, m\Delta t)\mathbf{q}(t) \tag{3.14}$$

if we define

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

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, i.e. that the tracer $q$ is quasi-conserved over times $m\Delta t$.

Now define a standard "i-wise" (e.g. over $x$) average

$$\overline{q}_j = \frac{1}{N} \sum_{i=1}^{N} q_{ij}$$

where $N$ grid spacings span the domain in the $x$ direction, and perturbation

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

Then, letting primes denote indices at time $t$ and unprimes denote indices at $t + m\Delta t$, take the $i$-wise average of equation (3.14). In element form, writing out all the sums, this is

$$\overline{q}_j = \frac{1}{N} \sum_{i} \sum_{i'} \sum_{j'} B_{iji'j'}(\overline{q}_{j'} + \Delta q_{i'j'}) \tag{3.15}$$

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

$$\overline{q}_j = \frac{1}{N} \sum_{i=1}^{N} \sum_{i'=1}^{N} \sum_{j'=1}^{M} B_{iji'j'}\overline{q}_{j'} + \frac{1}{N} \sum_{i=1}^{N} \sum_{i'=1}^{N} \sum_{j'=1}^{M} B_{iji'j'}\Delta q_{i'j'} \tag{3.16}$$

where $M$ grid points span the domain in $y$. Now if we define

$$C_{jj'} \equiv \frac{1}{N} \sum_{i=1}^{N} \sum_{i'=1}^{N} B_{iji'j'} \tag{3.17}$$

then (3.16) becomes

$$\overline{q}_j = \sum_{j'=1}^{M} C_{jj'}\overline{q}_{j'} + \frac{1}{N}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j'=1}^{M} B_{iji'j'}\Delta q_{i'j'} \tag{3.18}$$

which is simply (3.10) with an additional term on the RHS, which will be called the error term in the discussion below.

It is easy to show that if the full 2D advection-diffusion operator conserves mass, the transilient matrix defined by (3.17) does also. Additionally, we can guarantee that the transilient matrix will have no negative elements, since the matrix **A** cannot have any; if it did it could produce negative values where there were none before, which advection cannot do[6].

It is good that the transilient matrix thus derived has the necessary mathematical properties, but we would also like it to yield an identity when inserted into (3.10). Clearly this occurs only when the error term is negligible. Note, however, that if this 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 $i'$. 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 which define the error term. The larger the *variances* about the means at each level are, the less likely are either of these assumptions to be true, in general, and the less accurate the transilient matrix formulation may be expected to be.

The above discussion shows that (3.10) is indeed the most general model possible of eddy transport which uses only the mean profile of the tracer, and which is local in time (the Markov property). If there is a known correlation between $i$-wise tracer variation and eddy dynamics — for example, consider a smokestack emitting some pollutant in a plume which tends to rise because it is hotter than its environment — then one might design a better parameterization by using this information. However, such considerations are excluded by the kinematic approach taken here.

### 3.4.8 Convective structure memory and optimal time step

The error term clearly vanishes in the special case when all $\Delta q_{i'j'}$'s vanish, so that the tracer is exactly $i$-wise homogeneous at the initial time. This situation was enforced by Ebert et al. (1989) in their direct, 3D computations of transilient matrices from large eddy simulations of a planetary boundary layer. They did not use (3.17) for their computations. Rather, they ran their simulation until the turbulence statistics were stationary, paused the simulation, infused each model level with a separate and layerwise homogeneous tracer, and measured how much of each level's tracer had migrated to each other level after various amounts of time had passed. The error term will also vanish if those parcels whose tracer values are not identical to the $i$-wise means at their respective levels $j$ do not move in the $y$ direction, or if there is exact cancellation between the transports of parcels with positive

[6]However, it should be pointed out that while the true physical advection process is both linear and positive definite, the only numerical advection schemes which have both these properties are first-order upstream schemes, which are very diffusive. For a relevant discussion, see Thuburn and McIntyre (1997).

and negative $\Delta q_{i'j'}$'s, but these are clearly very special cases.

In some situations, it may be possible to minimize the errors due to convective structure memory by an appropriate choice of coordinate. The "equivalent latitude" coordinate used in chapter 7 is an example of this approach.

In the study of Ebert et al., the error term did come into play when those authors addressed the question of whether, as one would hope,

$$\mathbf{C}(n\Delta t) = [\mathbf{C}(\Delta t)]^n \tag{3.19}$$

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. They found that the above relation was never exactly satisfied, though the deviations from it were small when the time step was near the "convective time scale", $t^*$, essentially the turnover time of the largest eddies. Ebert et al. only initialized their tracer once, so in comparing, for example, their $\mathbf{C}(2\Delta t)$ with their $[\mathbf{C}(\Delta t)]^2$, it can be seen that the transport during the second time step will be influenced by the departures from layerwise homogeneity which have developed during the first. They referred to this effect as "convective structure memory", and discussed it qualitatively.

It makes some intuitive sense that what they called $t^*$ - essentially the time for a parcel to traverse the domain once caught up in a large-eddy circulation - should be the optimum time step no matter what the dynamical origin of the eddies. For much shorter time steps, the large eddies do not have time to "turn over" once, so their net effect on the tracer distribution cannot be fully evaluated. More formally, while an eddy is in "mid-turnover", large deviations of local tracer values from the layerwise means can be expected to cause the error term in (3.19) to be systematically large. On the other hand, if the time step is longer than $t^*$ by a sufficiently large factor, in the absence of sources or sinks the tracer will become homogeneous in a single time step, and information about the way homogeneity is reached will be inaccessible.

The claim that the self-consistency of the transport model (3.10) requires $\Delta t \sim t^*$ is a crucial one. This is because, absent constraints imposed by a particular application, the time step is a free parameter, and in general the properties of the transilient matrix depend on it. Therefore one needs some criterion for determing the right time step, at least to order of magnitude (small variations in time step are immaterial if within a range where 3.19 holds). It would be desirable to have a more general justification of this claim on theoretical grounds. Lacking this the best justification we can give is twofold. First, as mentioned above the results of Ebert et al. provide direct support of the claim for the particular flow that they simulated.

Second, we can consider theoretically the particular class of flows, admittedly rather special, for which the transilient matrix formulation is exact. These are flows in which eddies appear only intermittently, separated by time intervals in which the flow is quiescent (at least in the $y$ direction, using the convention of section 2) and during which the tracer distribution somehow becomes completely homogenized in the $x$ direction (by the same convention). The idea of such a flow is formalized in section 3.5.2.

Consider an infinite, plane parallel shear flow between two horizontal plates. Periodically, a heating strong enough to drive a convective circulation is applied to the lower plate for long enough that the circulation turns over approximately once. Then it is turned off again, giving the shear time to approximately homogenize any tracers in $x$ (the horizontal dimension). In this case it should be clear that the optimal time step is the period at

which the forcing is turned on and off (which we can make comparable to the duration of an "on" event by making the shear strong enough that $x$-homogenization is rapid), and the beginning of each time step should be the time just before the forcing is turned on. Then the mixing statistics can be represented exactly since the tracers will be $x$-homogeneous at the beginning *and end* of each time step. This guarantees that the result of each matrix multiplication according to (3.10) will give the correct initial condition for the next. If the time step is shorter than an event then one can no longer guarantee this; more than one time step will be needed to span an event, and at the beginning of the second time step the $x$-wise variations will in general be non-negligible. Whether the time step can be significantly longer than an event depends on how strong the mixing by each event is. As discussed above, if the time step is chosen longer than the time scale for homogenization, important information about the way the mixing occurs is lost.

In general, if the intensity of the large eddies varies substantially over time scales much longer than $t^*$, but some time-averaged representation of the mixing is desired, the best procedure, as noted by Ebert et al. (1989), is to compute many separate matrices for periods of $t^*$, and then average the matrices. If a longer time step is desired, this matrix can be raised to a higher power. This procedure avoids the loss of information that results from computing a single matrix over a time $nt^*$, where $n \gg 1$, and taking the $n$th root.

The proposition that statistics of mixing, such as the transilient matrix, are best evaluated over time periods on the order of the large-eddy turnover time will be reinforced by results shown in chapters 6 and 7.

### 3.4.9 Interaction of large-eddy transport and chemical or radiative processes: the Fredholm equation

The transilient matrix equation (3.10) 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 PV, which is affected by radiative heating. The simplest form of such a term which still allows some interesting dynamics is that of a Newtonian relaxation. That is, we can study the system

$$\mathbf{q}(t+\Delta t) = \mathbf{C}(t,\Delta t)\mathbf{q}(t) + \frac{\Delta t}{\tau_e}(\mathbf{q_e} - \mathbf{q}(t)) \tag{3.20}$$

where $\mathbf{C}$ is the transilient matrix, $\mathbf{q_e}$ is a given profile representing the equilibrium tracer distribution in the absence of transport, and $\tau_e$ is a chemical or radiative time scale. The factor $\Delta t$ in the second term on the RHS comes in because the transilient matrix represents an amount of mixing occurring in a finite time $\Delta t$. The assumption has been made that $q$ can be taken close enough to constant over this time interval that the relaxation does not need to be explicitly evaluated as an integral over the time step; essentially this is a restriction that $\tau_e$ be large compared to $\Delta t$. The same restriction also guarantees the assumption, already made implicitly, that each parcel conserves its tracer value during the journey from layer $j'$ to layer $j$.

It is interesting to consider steady state solutions to (3.20). In this case we have

$$\mathbf{Cq} - (1 + \frac{\Delta t}{\tau_e})\mathbf{q} = -\frac{\Delta t}{\tau_e}\mathbf{q_e} \tag{3.21}$$

The continuous version of (3.21) may be written in the form

$$\int_0^{N\Delta y} C(y|y')q(y')dy' - \lambda q(y) = f(y) \tag{3.22}$$

which is known as the *Fredholm equation of the second kind.* This is a fairly well-studied equation, and so we benefit from past efforts of capable mathematicians. The reduction from continuous to discrete form raises some 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.

### 3.4.10 Properties of eigenspectrum

Let us assume that the columns of $\mathbf{C}$ are all linearly independent, though strictly this need not be the case. Linear independence can be assured by taking a sufficiently small time step, since as the time step vanishes $\mathbf{C}$ becomes the identity matrix. This small time step may be suboptimal for other reasons, which may be partially gleaned from the discussion above of convective structure memory. Nonetheless in practice it seems likely that linear independence will not require violation of the consistency constraint, and this will be assumed hereafter.

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 $\lambda_k$'s are the eigenvalues, which may or may not be distinct. We can then expand any profile $\mathbf{q}$ in the eigenvectors,

$$\mathbf{q} = \sum_k a_k \mathbf{q_k}$$

and express all time dependence in terms of the coefficients $a_k$:

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

The fact that all realistic transilient matrices have the uniform distribution as their unique limiting distribution, and have no negative elements, means that

$$0 \leq \lambda_k < 1 \tag{3.23}$$

for all eigenvectors except the one corresponding to the uniform distribution, for which $\lambda_k = 1$.

### 3.4.11 The Fredholm alternative

Given these properties of the eigenspectrum, we can establish the solvability of (3.21).

The *Fredholm alternative* (e.g. Hochstadt 1973) states that an equation of the form (3.22) has a unique solution for arbitrary $f(y)$ if and only if the associated homogeneous equation

$$\int_0^{N\Delta y} C(y|y')q(y')dy' - \lambda q(y) = 0 \tag{3.24}$$

has no solution other than $q(y) = 0$. Again ignoring the distinction between discrete and continuous versions of this theorem, we can see that the homogeneous form of (3.21), which

is

$$\mathbf{Cq} - (1 + \frac{\Delta t}{\tau_e})\mathbf{q} = 0 \tag{3.25}$$

can have no nontrivial solution. This is because all eigenvalues of the matrix C are positive and less than or equal to unity, while the quantity $1 + \frac{\Delta t}{\tau_e}$ is always greater than unity. Hence no eigenvector of **C** is a solution of (3.25), and so by linearity no solution to (3.25) can be constructed.

If (3.21) has a unique solution for all $f(y)$, the solution to the time-dependent equation (3.20) will smoothly approach it. This is guaranteed by the property (3.23) which is satisfied by all the nonuniform eigenvectors.

## 3.5 Relaxation to the average

### 3.5.1 Derivation from transilient matrix model

Rather than using the transilient matrix as in (10), we can use the explicit form, that is

$$\mathbf{\Delta q} = \mathbf{q}(t + \Delta t) - \mathbf{q}(t) = (\mathbf{C} - \mathbf{I})\mathbf{q} \tag{3.26}$$

with **I** the identity matrix. We also redefine **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 **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 $\epsilon_k = \lambda_k - 1$. $-\epsilon_k$ is the inverse of the decay time scale for each eigenvector, scaled by $\Delta t$.

In general, the spectrum of $\epsilon_k$'s may have any form. However, we can envision a couple of cases which allow a drastically simplified treatment.

One case is that in which all the $\epsilon_k$'s corresponding to nonconstant eigenvectors are similar in value, so that $\epsilon_k \approx \epsilon$ where $\epsilon$ is a constant. Then we have

$$\Delta a_k = \sum_k \epsilon_k a_k \approx \epsilon \sum_k a_k$$

which corresponds to

$$\Delta \overline{q}_j \approx \epsilon \overline{q}_{j'}$$

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

$$\frac{\partial \overline{q}}{\partial t} \approx \tau^{-1}(\hat{q} - \overline{q}) \tag{3.27}$$

where

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

with $L$ the size of the domain. That is, 3.27 represents relaxation of the mean tracer field

everywhere towards its global average, with a single time scale $\tau$, where

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

The limiting process should be viewed as representing the choice to view the system only on time scales long compared to the eddy turnover time $t^*$. Equation 3.27 will be referred to hereafter as the "relaxation to the average" (RA) model. The transilient matrix corresponding to this approximation has each element along the diagonal equal to $1 - (1 - \frac{1}{N})\Delta t/\tau$ while every off diagonal matrix element is equal to $\Delta t/N\tau$.

In general we do not expect all eigenvalues to be similar, but if there is a "spectral gap", that is, a situation in which a few eigenvalues are similar and close to zero while the rest are much closer to -1, then the approximation is still justified. We can simply represent the system by a truncated expansion in terms only of those eigenvectors whose eigenvalues are close to zero, and only study it on the "slow" time scale such that any projection of the initial conditions on the quickly decaying eigenvectors is assumed to have vanished. This approach can, of course, fail if there is some forcing term that projects much more strongly on the quickly decaying eigenvectors than on the slowly decaying ones. Even then, the truncated representation may still be useful. For example, if the slowly decaying modes have large physical scales compared to the others, we may simply say that we care only about the large scale.

In using the model (3.27) one is not, however, actually truncating the tracer distribution so that it only projects on a single eigenvector. Rather one assumes that all the (nonconstant) eigenvectors decay at the same rate. However the main point is that the decay of the longest-lived components of the tracer structure will be represented accurately.

It is important to realize that the homogenization time scale $\tau$ need not be the same as the eddy turnover time $t^*$. The latter is roughly the time for a parcel caught up in a large eddy to traverse the domain, while the former is the time scale over which the entire domain approaches homogeneity due to mixing. The latter can be shorter than the former if the eddies are small in the unresolved dimension, i.e. narrow, widely spaced plumes or filaments. Hence it is not necessarily inconsistent to use the RA model to represent the transport of species which have sources and sinks whose time scales are comparable to $\tau$. Finally, it is worth noting that in the limit as $\tau$ 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.

### 3.5.2 Heuristic derivation and further discussion

In this section the RA model will be explored in more detail. My interest in it derives from the fact that it captures an essential property of large-eddy mixing, its nonlocalness, while nonetheless being simple enough to allow analytical manipulation.

In the preceding section, the RA model was derived as an approximation to mixing by a general transilient matrix. Here, an alternate, heuristic derivation is presented, which is of further use in that it suggests a somewhat more general model. of which (3.25) is again a special case, yet which is still substantially simpler than the general transilient matrix. 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, with arbitrary structure in the $y$ direction but homogeneous in the $x$ direction. This amounts

essentially to an assumption that there is a highly anisotropic mixing process acting in the $x$ direction, and that we apply a separation of time scales and consider only the mixing in the $y$ direction. The latter occurs through episodic events which occur very quickly but mix only a small amount of material per event, so that the time-averaged effect is one of very slow mixing relative to that in the $x$ direction.

Every time interval $\Delta t$, an "eddy" which spans the domain in $y$, entirely homogenizes a region of width $W_1$ (which is much smaller than $W_0$) and length $L$. This situation is sketched schematically in figure 3-1. Hence,

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

Now, let us take the limit as $\Delta t$ vanishes, but requiring that $W_1$ vanish in a directly proportional manner, that is,

$$\lim_{\Delta t \to 0} \frac{W_1}{W_0 \Delta t} \equiv \tau_m^{-1} \tag{3.28}$$

This leads immediately to (3.27).

The meaning of the limiting process is, again, that we are focusing our interest on time scales very much longer than $\Delta t$ (which is the appropriate eddy turnover time in this idealized model). If we are to allow non-uniform tracer distributions to exist in our model, we must assume that even on this "slow" time scale the eddy mixing is not strong enough to completely homogenize the domain. This requirement forces us to let $W_1$ vanish with $\Delta t$. In essence this implies that transport across the domain is occurring via plumes or filaments which are very narrow compared to the domain width (the $x$ direction) but which traverse the domain length (the $y$ direction) very quickly. Consistently with this requirement, relations of the form of (3.27) are used in some deep convective parameterizations (Betts 1986, Betts and Miller 1986, Raymond 1994).[7] In general, the modifier "very quickly" must be interpreted by comparison to some other time scale in the problem, such as that of a nonconservative process which tends to drive the tracer distribution away from homogeneity.

### 3.5.3 More general mixing profiles

Now consider a situation like that considered above, except now the width of the region homogenized, $W_1$, is a function of $y$. This situation is depicted in figure 3-2. Steps directly analogous to those taken above lead to, in the limit of vanishing time step,

$$\frac{\partial \overline{q}}{\partial t} = \tau_m^{-1}(y)(\tilde{q} - \overline{q}) \tag{3.29}$$

with, at any given $y$,

$$\lim_{\Delta t \to 0} \frac{W_1(y)}{W_0 \Delta t} \equiv \tau_m^{-1}(y) \tag{3.30}$$

[7] Though in those cases the meaning is not only kinematic but dynamic as well, in that the circulation is imagined to cease when the dynamically important tracer (e.g. equivalent potential temperature) attains a uniform distribution, according to considerations of hydrodynamic stability. This is quite different, especially as regards what one assumes about what happens to *other* tracers, from a situation in which the circulation doing the mixing continues to act, but simply cannot homogenize the tracer any further because it is already homogeneous. The latter is implied by the kinematic approach.

and

$$\tilde{q} = \frac{\int_0^L \tau_m^{-1}(y) q(y,t) dy}{V} \tag{3.31}$$

with

$$V = \int_0^L \tau_m^{-1}(y) dy$$

We now have a model of intermediate simplicity, in that the function $\tau_m(y)$ is a one-dimensional object, as opposed to the two-dimensional transilient matrix or the zero-dimensional single time scale of (3.27).

In the model (3.29), of which (3.27) is a special case, we can find the kinematic flux of tracer at any point, $F(y)$, simply by using the expression

$$\frac{\partial \overline{q}(y)}{\partial t} = -\frac{\partial F(y)}{\partial y} \tag{3.32}$$

so that

$$F(y) = -\int_0^y \tau_m^{-1}(y)(\overline{q} - \tilde{q}) dy + F_B \tag{3.33}$$

Where $F_B$ is the contribution of fluxes from the boundaries. (3.33) reduces, in the case where $\tau_m$ is a constant so that (3.27) holds, to

$$F(y) = -\tau_m^{-1}[\int_0^y \overline{q}(y) dy - \hat{q}y] + F_B \tag{3.34}$$

When $\tau_m$ is everywhere very small compared to time scales for chemical or other nonconservative forcings on the tracer, we expect to have in the interior of the domain $\overline{q} \approx \tilde{q} \approx \hat{q}$, and in that case the flux will be dominated by boundary contributions if there are any, $F \approx F_B$ everywhere. A more interesting case is that in which $\tau_m$ varies so that it is small somewhere but not everywhere, for example small in the interior but becoming large at the boundaries.

### 3.5.4 Relaxational mixing in the presence of relaxational chemistry

Consider now the model (3.29) with a chemistry term (or other nonconservative term if $q$ is not a chemical species) added, that is

$$\frac{\partial q(y)}{\partial t} = \tau_m^{-1}(y)(\tilde{q} - q(y)) + \tau_e^{-1}(q_e(y) - q(y)) \tag{3.35}$$

where $q_e$ is some given chemical equilibrium profile, $\tau_e$ is a chemical time scale, and the overbar has been dropped so that $q(y)$ now represents the mean tracer profile. Assume no-flux boundary conditions. In principle $\tau_e$ might just as well be a function of $y$ as $\tau_m$, but taking the former as constant will be sufficient to demonstrate the main points at issue here. It is important to keep in mind that while the two terms on the RHS of (3.35) appear to have the same form there is a fundamental difference between them, in that the first conserves the global integral of the tracer while the second does not.

Of particular interest are steady state solutions of (3.35). These are

$$q(y) = [\tau_m^{-1}(y) + \tau_e^{-1}]^{-1}[\tau_m^{-1}(y)\tilde{q} + \tau_e^{-1} q_e(y)] \tag{3.36}$$

though this is not a complete solution, as $\tilde{q}$ is not yet determined. Integrating (3.35) in $y$ from 0 to $L$ gives $\hat{q} = \hat{q_e}$. Mixing cannot change the steady-state global tracer burden from its chemical equilibrium value no matter what the profiles of $q_e$ or $\tau_m$; however if the chemistry were not linear in $q$, or if $\tau_e$ were allowed to vary in space, this result would no longer hold in general. Even in the present case, knowing this fact does not solve the problem because it is $\tilde{q}$, not $\hat{q}$ that needs to be expressed in terms of the given quantities. To do this, set the LHS of (3.35) to zero, divide the entire equation by $\tau_m(y)$, and then integrate from 0 to $L$. The result is

$$\tilde{q}\int_0^L \tau_m^{-2}(y)dy - \int_0^L q\tau_m^{-2}dy + V\tau_e^{-1}(\tilde{q_e} - \tilde{q}) = 0 \tag{3.37}$$

In principle the problem is solved, since (3.36) and (3.37) are two equations in two unknowns $q$ and $\tilde{q}$. The solution for $q$ in (3.36) can be substituted into (3.37) to give a closed equation for $\tilde{q}$, the solution of which can then be substituted back into (3.36) to give a completely determined solution for $q$. Since (3.37) is an integral equation, this procedure might in general have to be carried out numerically, for example by some iterative procedure, but there should be no difficulty beyond this.

## 3.6 Can diffusivity successfully represent large-eddy mixing?

It was shown in section 3 that when the eddies doing the mixing are large, the assumptions leading to the diffusion model are violated and the model is theoretically indefensible. This is far from a new conclusion. It is evident in Taylor's original (1921) paper, and instances in which this limitation of the diffusion model are more specifically pointed out are many. The inapplicability of the eddy diffusion model to mixing by stratospheric planetary waves, in particular, has been noted more than once (e.g. McIntyre 1991; Pierrehumbert 1991a,b). Nonetheless, to my knowledge all zonally symmetric stratospheric models currently in use parameterize horizontal mixing as diffusion (e.g. see Prather and Remsberg 1993). In part, as suggested previously, this may be due to the lack of alternatives to the the diffusion model which are both simple and well-understood. Diffusion may be wrong, but at least it is simple, and the modeler has a good intuitive grasp of how it will affect the model's behavior!

There may also be something of the feeling that, no matter what the theoretical issues, as long as the fluxes are down-gradient one ought to be able to tune $K_{yy}$ so as to get things right, at least within the substantial error bars imposed "beforehand" by the chemistry and radiation codes, model parameters obtained from observations, etc.[8] It is beyond this author's power to entirely disprove this last notion, but the development of the preceding sections allows us to make the criticism of the eddy diffusion model at least a bit more pointed.

[8]Of course, in general one might prefer that $K_{yy}$ not be treated as a tuning parameter, but should either be determined by an explicit planetary wave breaking parameterization (e.g. Garcia 1991) or from observations (e.g. Newman et al. 1986; Newman 1988; Yang et al. 1990). The more strongly the conditions for the validity of the diffusion model are violated, the less likely are such techniques to yield optimal $K_{yy}$'s. "Optimal" is meant here in a sense to be discussed in section 7.2.4.

### 3.6.1 Illustration - $K$ derived from RA model

Let us return to the case in which $\tau_m$ is a constant. Then the steady state solution of (3.35) is

$$q = [\tau_m^{-1} + \tau_e^{-1}]^{-1}[\tau_m^{-1}\hat{q}_e + \tau_e^{-1}q_e] \tag{3.38}$$

Now let us focus on a still more special case, in which $\tau_m = \tau_e = \tau$. This is not an arbitrary choice, or one made purely for the sake of simplicity. Its significance will be discussed in section 3.6.3. Also, let $q_e(y) = \alpha y$ where $\alpha$ is a constant; this choice on the other hand is purely for simplicity's sake. With these choices the solution is:

$$q(y) = \frac{1}{2}(\alpha y + \frac{\alpha L}{2}) \tag{3.39}$$

Using the results of the previous section, and again assuming no flux through the boundaries, it is then easy to show that the flux at any point $y$ is given by

$$F = \frac{\alpha}{4\tau}(y^2 - yL) \tag{3.40}$$

Now, knowing exactly what the solution is for the tracer distribution and for the flux, we can answer the question: even though we know that the transport is not diffusive, can we model it diffusively if we allow ourselves complete freedom in fitting a diffusivity $K(y)$ to the results? For this tracer, considered alone, the answer is of course yes, since the fluxes are (indeed, as was shown, must be) down-gradient; one can simply write

$$F = -K\frac{\partial q}{\partial y}$$

and solve for $K$. In this case one obtains

$$K(y) = \frac{yL - y^2}{2\tau} \tag{3.41}$$

Having successfully completed this exercise, let's consider a different tracer, for which $q_e = \beta y^2$, but still take the mixing and chemical time scales equal. In this case the steady state solution is

$$q(y) = \frac{1}{2}(\beta y^2 + \frac{\beta L^2}{3}) \tag{3.42}$$

and, going through the same exercise as before to find the flux and fit a function $K(y)$ to the solution, one obtains

$$K(y) = \frac{L^2 - y^2}{2\tau} \tag{3.43}$$

which does *not* have the same functional form as (3.41) . In fact, (3.43) does not vanish at $y = 0$; since $\frac{\partial q}{\partial y}$ does vanish there for the tracer for which (3.43) 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.

In other words, the artificially derived "diffusivity" is affected by the chemistry terms, which is inconsistent with the very formulation of the problem as stated at the outset. If the diffusion model is substituted for the model used here, clearly one ought to expect incorrect results no matter how $K(y)$ is tuned, as long as more than one species is transported, assuming that different species will in general have different sources and sinks. Even if only

one species is transported, there was no other way to obtain the diffusivity but by solving the entire problem first, in which case there is no need for the diffusivity! While in less idealized situations things will not be so clear-cut, one may expect something like this to be true whenever the eddies are large, since then theory can offer no help in deriving $K$ from observed or simulated data. If one then wants to derive an accurate flux-gradient relationship for a species, one is left using the brute force method used here, but with measured or numerically computed fluxes and gradients, and then again once these are known the derived diffusivity is no further help, since one cannot expect it to allow prediction of the flux of another species.

That said, what has been demonstrated by the above calculations is not that the diffusion model is "wrong" while the relaxation-to-the-average model is "right", for transport by any particular species of large eddies. What has been shown is that it is possible to formulate a simple, consistent model to parameterize transport by domain-size eddies, and that that a diffusion model cannot be made to mimic the former model's behavior in a consistent manner. The two models are inescapably different at a practical as well as a theoretical level, despite the fact that in many situations both will have down-gradient fluxes everywhere.

### 3.6.2 Diffusion vs. alternatives, in summary

It is worthwhile to summarize here some of the key points which have been made so far.

1. The diffusion model cannot be derived without assuming a *spatial* scale separation between the eddies and some other scale of interest. When the domain is bounded, the latter scale can be no larger than the size of the domain. If the eddies are the same size as the domain, the diffusion model is unjustifiable theoretically.

2. The most general model for eddy mixing which still has the desirable properties of being history-independent (Markov) and involving only the (zonally, horizontally, etc.) averaged tracer values is the transilient matrix model. This model requires no intrinsic assumptions about eddy size.

3. The transilient matrix formulation is only self-consistent if the time step is chosen comparable to the turnover time of the largest eddies. If only very small eddies are present, a tridiagonal (diffusion) matrix will result. On the other hand, if there are domain-size eddies, some parcels will traverse the domain in a single time step, so that the matrix will have nonzero elements far from the diagonal. In other words, a self-consistent parameterization of large-eddy mixing must be *nonlocal*.

4. The RA model can be derived as an approximation to the transilient model, by taking all of the transilient matrix' eigenvalues equal to the smallest ones. A heuristic derivation is also possible. The RA model is strongly nonlocal: it is perhaps appropriate to say that it represents the opposite limit to diffusion.

5. The RA model and the diffusion model are not interchangeable, but have practical as well as theoretical differences. Hence even if one is not impressed by purely theoretical arguments, one ought to consider the requirements of a given application carefully before blindly choosing the diffusion model. This is true even if fluxes of the properties of interest are everywhere downgradient.

### 3.6.3 Tracers with different time scales

A mathematical point has been swept under the rug in the preceding discussion. I have suggested that the RA model can be thought of as a truncation of the full transilient matrix representation, in which the decay time of the most slowly-decaying eigenvector is considered representative of the entire spectrum. This is, however, a plausible way to approximate a tridiagonal (diffusive) matrix as well, if one only cares about the evolution of the largest scale tracer variations! For this reason, the diffusion model provides a useful representation even of large-eddy mixing when only a rough mixing time scale is desired, as opposed to an explicit description of the distribution of eddy fluxes throughout the domain. The RA model is preferable to the diffusion model only when it is the closer approximation to the true transilient matrix (with a proper choice of time step). Even then, though, the extent to which the difference is significant in practice depends on the chemical time scale of the tracer.

Let us compare two models, an RA model with single time scale $\tau$ and a diffusion model with constant diffusivity $K$. The domain is again a "box" of width $W$ and length $L$. Let $K$ and $\tau$ be such as to give the same time scale for large-scale homogenization, that is, the most slowly decaying eigenvector has the same eigenvalue in the transilient matrices representing the two models (though in the RA model all eigenvectors have the same eigenvalues so "most slowly decaying" is not particularly meaningful). Now consider the behavior of some tracer which is being relaxed towards inhomogeneous chemical equilibrium profile with chemical time scale $\tau_e$. In the preceding section we considered only the case where this was the same as the mixing time scale. What if it is substantially different?

Without even solving the problem formally we can see that in steady state, if $\tau_e \ll \tau_m$ the tracer will simply keep its chemical equilibrium, while if $\tau_m \ll \tau_e$ it will become homogeneous, *in either transport model.* We might conclude from this that the RA and diffusion models are substantially different only when the chemical and mixing time scales are comparable. Within the narrow confines of the problem as formulated so far this is true. Even this is not a great flaw in the preceding argument. If one is interested in chemistry one may typically want to transport a number of species in a chemical transport model. These will likely have a range of time scales, and some of those are likely to be comparable to the mixing time scale. In a middle atmosphere model which resolves the vertical dimension, tracers which are long-lived in some region will be shorter-lived in others, due to the increase in solar UV flux with altitude. Many species may thus be expected to have a region somewhere in which the mixing and chemical time scales are comparable.

### 3.6.4 Practical stratospheric considerations

As far as the parameterization of stratospheric isentropic mixing is concerned, what should be compared to the mixing time scale is not just the true chemical time scale, but the time scale of the fastest process which causes horizontal (i.e. isentropic) inhomogeneity of a species. This includes vertical (cross-isentropic) advection, which in regions of strong vertical motion and strong vertical gradients (e.g. the high latitudes in winter, for species like $N_2O$ and $CH_4$) is much faster than chemistry.

Perhaps most importantly, the existence of transport barriers must lead to severe practical problems for anyone wishing to diagnose a diffusivity profile from observations. At the polar vortex edge, tracer fluxes become small while gradients become large, relative to their surf zone values. Modeling the transport by a diffusion model thus requires a severe drop

in the diffusivity at the vortex edge. The flux then is the product of two quantities which vary rapidly in space, leading to a strong likelihood of large errors due to any uncertainty in either quantity.

## 3.7 Relaxation to the average within subdomains separated by leaky transport barriers

### 3.7.1 The stratosphere is more than a surf zone

The above discussion has, hopefully, served to make clear some general notions about the nature of mixing by large eddies. The RA model has been proposed as an alternative to diffusion, if simplicity is desired. If a higher level of computational complexity is acceptable, the transilient matrix allows greater accuracy to be obtained. The RA model, as presented above, is adequate for illustrating the theoretical ideas which have been the primary focus of this chapter, but as even a highly idealized model of stratospheric mixing it is in one particular respect too oversimplified. This is, as might be guessed from the preceding discussion, its lack of transport barriers.

I have been referring to eddies which "span the domain". In the stratosphere, the true domain is the entire planet, and while there are eddies which are of planetary scale, there are no eddies which truly span the globe from pole to pole. Inasmuch as the preceding discussion is taken as referring to the stratosphere, we should associate the "domain" with the wintertime midlatitude surf zone, which is indeed spanned by the circulation(s) associated with the major anticyclone(s). To the lowest approximation, the surf zone may be considered a separate "domain" since the mixing within it is more rapid than the mixing across either its poleward or equatorward edge. However, in the end one does not want to consider the surf zone in isolation from the rest of the stratosphere, as its edges are certainly somewhat permeable. A complete theory must include the exchange of material across the transport barriers. Techniques for diagnosing this in real flows will be discussed in later chapters. Once this exchange has been quantified we would like to develop a version of the RA model which consistently incorporates multiple mixing regions separated by transport barriers.

### 3.7.2 Two or more boxes

The RA model can be easily extended in this manner. Simply consider two "boxes", each individually like the one in section 3.5.1. Each box has its own length, $L_1$, $L_2$, and its own "eddy width", $W_1$, $W_2$. The domain width, $W_0$, is of course the same for each box. The new feature is that we take the boxes to overlap slightly, by an amount $dL$. This is depicted in figure 3-3. The resulting derivation is the same as before, that is, writing $q_i$ for the mean tracer value in box $i$ (and again omitting the overbars) we have

$$\frac{\partial q_i}{\partial t} = \tau_i^{-1}(\hat{q}_i - q) \tag{3.44}$$

where $\tau_i$ is the mixing time scale obtained from the limiting process for box $i$. The only difference now is that the average $\overline{q}_1$ includes some of the same material as $\overline{q}_2$, since the boxes overlap. In fact (3.41) only applies outside the overlap region. Within the overlap region, we have, writing $q_o$ for the tracer there,

$$\frac{\partial q_o}{\partial t} = \tau_1^{-1}(\hat{q}_1 - q) + \tau_2^{-1}(\hat{q}_2 - q) \tag{3.45}$$

This approach can be extended to any number of boxes. To make a highly simplified model of the global stratosphere, for example, one might imagine having one box for the polar vortex, one box for the surf zone, a tropical box, and a single box for the rest of the summer hemisphere. Each box could have its own internal mixing time scale, and the mixing between two neighboring boxes would be set by the extent of their overlap with one another. The resulting model would be quite similar to the "pipe model" of Wofsy et al. (1997), though not quite as simple as the latter in that the concentration of each species is allowed to vary spatially within each box.

### 3.7.3 Infinitely many boxes

While not relevant to the stratosphere, to further illustrate the central point of this chapter it is instructive to allow the number of boxes to tend to infinity. Let's start again from the discrete mixing events, occurring every time step $\Delta t$. Let each box be the same, that is, all $W_i$'s are equal to $W_1$, and all boxes overlap by the same amount $\Delta L$. However, assume that the "eddies" represented by the homogenized regions of width $W_1$ do not overlap in $x$ for any two adjacent boxes, so that no two eddies have a claim on the same material in any overlap region. Within each box, we have as before

$$q_i(t + \Delta t) = (1 - \frac{W_1}{W_0})q_i(t) + \frac{W_1}{W_0}\hat{q}_i \tag{3.46}$$

where now

$$\hat{q}_i = (1 - 2\epsilon)\tilde{q}_i + \epsilon(q_{i+1}^o + q_{i-1}^o)$$

where $\epsilon = \Delta L/L$, $\tilde{q}_i$ is the average over box $i$ excluding the overlap regions, and $(q_{i+1}^o$ and $q_{i-1}^o$ are the average values within the two overlap regions. Consider the case $\epsilon \ll 1$. In this case we imagine that the variations of $q$ within a box will be small compared to the variations between boxes, so that away from overlap regions $q_i$ can be approximated reasonably well by $\hat{q}_i$ for purposes of understanding box-to-box transport. Within overlap regions, a reasonable assumption is that $(q_{i+1}^o$ and $q_{i-1}^o$ can be approximately represented by

$$q_{i+1}^o = \frac{1}{2}(\overline{q}_i + \overline{q}_{i+1})$$

which is in fact the steady state solution of (41) for equal time scales in each box. Substituting this into (42) and simplifying leads to

$$\Delta\hat{q}_i = \hat{q}_i(t + \Delta t) - \hat{q}_i(t) = \frac{W_1\epsilon}{2W_0}(\hat{q}_{i+1} + \hat{q}_{i-1} - 2\hat{q}_i) \tag{3.47}$$

Now divide through by $\Delta t$, multiply the right hand side by $L^2/L^2$. Then as before take the limit as $\Delta t$ vanishes, except now the quantity which must be held fixed to insure finite and definite results is the quantity $\epsilon L^2 W_1/W_0\Delta t$. Beyond this, however, I assume that $L^2$ vanishes proportionately to $\Delta t$, while the other quantities remain finite. If we simply rename $L$ as $\Delta y$, and define

$$\lim_{\Delta t \to 0} \frac{(\Delta y)^2 W_1 \epsilon}{2W_0\Delta t} \equiv K$$

then 3.47 becomes simply

$$\frac{\partial \hat{q}}{\partial t} = K \frac{\partial^2 \hat{q}}{\partial y^2} \tag{3.48}$$

a diffusion equation. This derivation is in fact very similar to that of Frisch (1991). However, having started by considering the mixing within a single box, telescoping out to view a limitless field of tiny connected boxes allows us to see with particular clarity the relationship between diffusion and large-eddy mixing, since both have been derived as limits of essentially the same model.

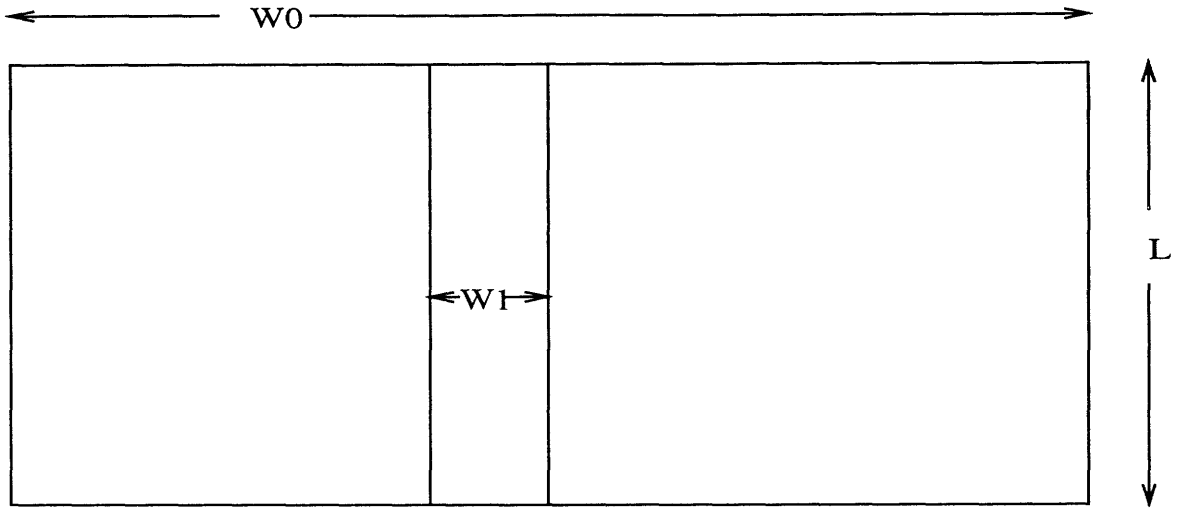


Figure 3-1: Schematic for heuristic derivation of RA model.

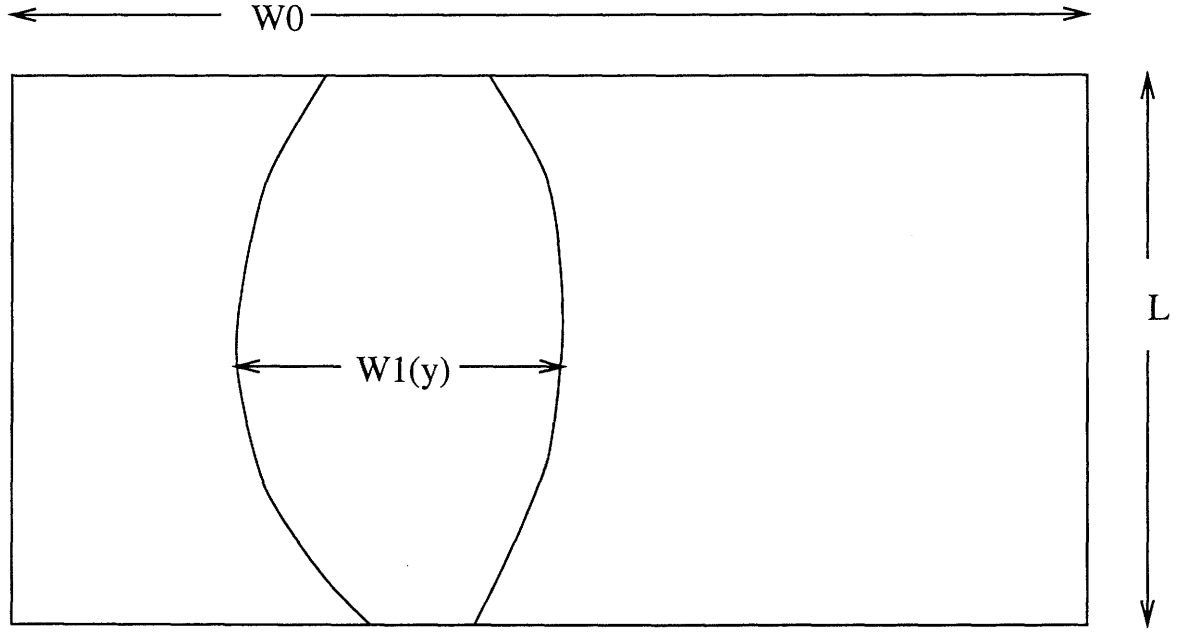


Figure 3-2: Schematic for heuristic derivation of RA model with a homogenized region of nonuniform width.

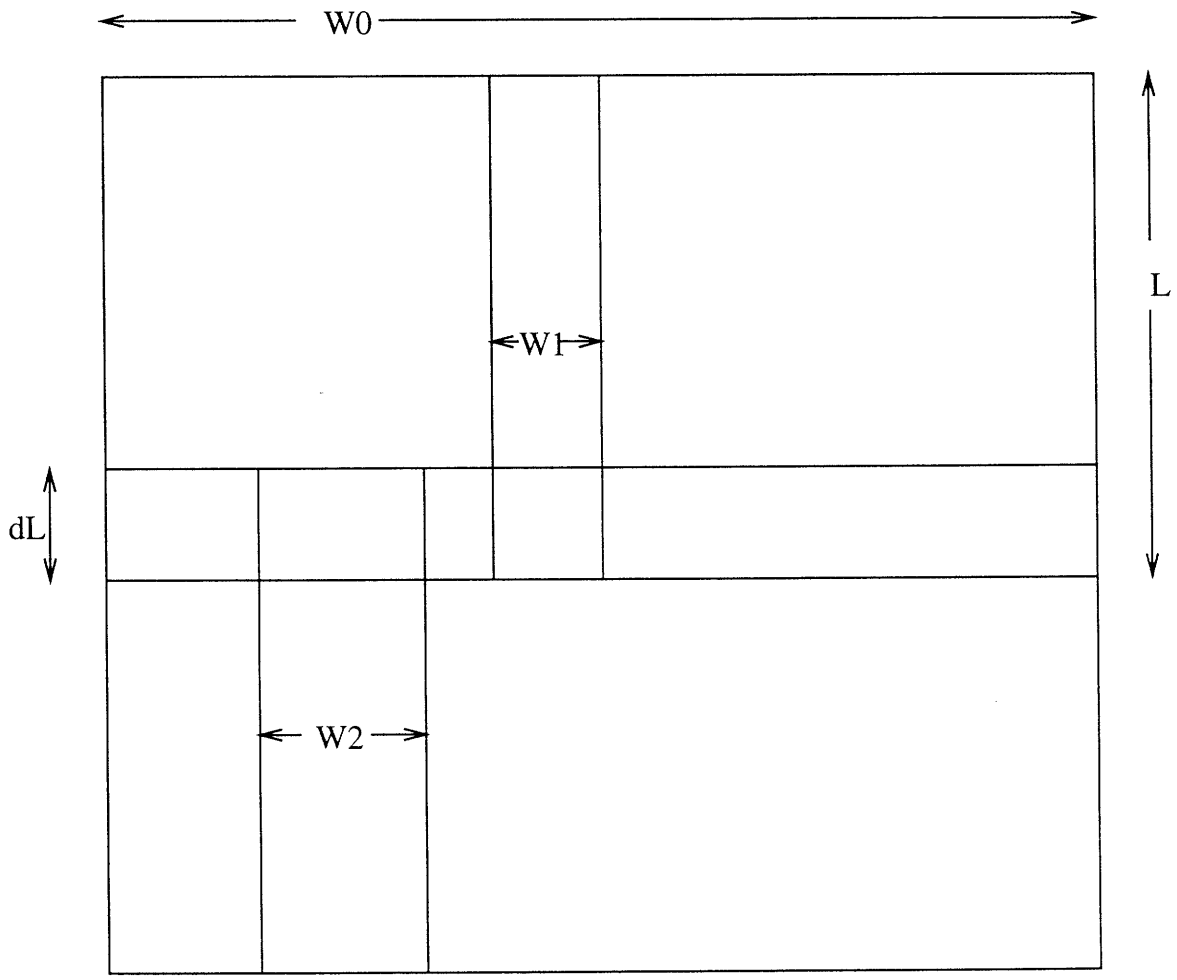


Figure 3-3: Schematic for heuristic derivation of RA model with multiple boxes.

# Chapter 4

# Transport across internal surfaces

## 4.1 Turbulence vs. waves

The preceding chapter discussed the relevance of different transport models to flow regimes with eddies of various sizes. I made the assumption early in the discussion that the coordinate was a standard physical coordinate, meaning that the coordinate surfaces had some reasonably simple geometry and, more importantly, were fixed in space. The discussion drew heavily on ideas from the study of turbulence. When the flow is fully turbulent, using a fixed coordinate framework is not only the simplest choice, but is physically reasonable.

The stratospheric flow of interest here, however, is not fully turbulent, at least not by all definitions. Most of the time in midwinter, the large-scale flow has a strong aspect of waviness to it due to the influence of planetary waves propagating up from the troposphere. When the waves break, the resulting process has the character of *chaotic advection* (Aref 1984; see also, e.g., Ottino 1990; Pierrehumbert 1991a,b; Wiggins 1992; Pierrehumbert and Yang 1993; Pierce and Fairlie 1993; Ngan and Shepherd 1997) rather than turbulence *per se*. The distinction is not rigorous, but loosely speaking, it means that complex tracer structures, corresponding to individual parcel trajectories which are *chaotic* in the sense of having sensitive dependence on initial conditions. are produced by flow fields which are relatively smooth and regular. This is as opposed to turbulent flows, in which both flow and tracer fields are chaotic.

Over much of time and space, however, the planetary waves are not large enough to break, so that fluid parcels undergo reversible undulations in latitude as they circle the pole. Since reversible undulations (by loose definition) induce no mixing, coherent large-scale gradients in PV and tracers can be sustained. However. the waves often have large amplitudes. This means that the result of Plumb (1979), which provides support for the diffusion model for flows characterized by small-amplitude waves in the presence of damping, is generally not valid since the result was obtained by assuming infinitesimal wave amplitude.

In fact, the large wave amplitudes, together with the large space and time scales of the waves, imply that not only is the eddy diffusivity approximation invalid, but the zonal mean itself is of limited utility. Planetary wave "cycles" of wave growth and decay can have time scales of weeks. In a typical scenario where wave number one is dominant, the vortex might start roughly centered on the pole, then wander far off it (in spectral terms, a large increase in the wave number one component) and then wander back over the pole a week or several later, with other fluctuations of shorter space and time scales superimposed on this motion. In a zonal mean framework, the wandering off the pole represents a large transport

of PV from the polar region to midlatitudes, while the wandering back represents a large transport back the other way. Yet clearly there is something misleading about this view of what has happened. Particularly in studies relevant to atmospheric chemistry, what is of interest is the degree to which different air masses, defined by chemical composition (which conveniently is often reasonably well correlated with PV), are mixed together. That degree could be zero in the case just described, yet if we are not careful to integrate the zonal mean transport over just the right period of time, we will find the latter to be large.

The same problem arises if one instead chooses to compute the transport with respect to the time-mean flow, since the fluctuations about this state are also finite-amplitude and may have long time scales. Said another way, the diagnosed time-mean flow will in general be sensitive to the duration and offset of the averaging period, as long as that period is shorter than a season, and so the transport with respect to this mean will probably be likewise sensitive. If the averaging period is comparable to a season in duration, the time mean will be poorly defined due to the intrinsic nonstationarity of the system at that time scale. Averaging over a year or more may be illuminating for some purposes, but is useless for understanding transport-chemistry interactions on shorter time scales, including all processes involving the polar vortex.

### 4.1.1 Reversible vs. irreversible

What is needed is a practical way to separate the "reversible" from the "irreversible" transport. What I mean by "irreversible" transport is: that which corresponds in some reasonably direct way to the intimate (that is, molecular) mixing of fluid with different properties. This is not the only valid definition. In the formal sense defined by classical statistical mechanics, the enstrophy cascade in perfectly inviscid 2D turbulence is irreversible, even though there is no molecular mixing (Carnevale 1982; Pierrehumbert 1991a; see also SPW). There are subtleties involved in understanding the relationship between these different notions of irreversibility. Chapter 6 will show nonetheless that at least in some situations, different definitions (or rather, their embodiments in different diagnostic techniques) give approximately the same results. This indicates that the underlying phenomenon is more robust than is any one way of defining it.

For the moment, let us keep the definition involving intimate mixing of different fluid parcels. Then, one particular alternative to zonal and time averaging which offers improved prospects for diagnosing the irreversible transport alone, without having to average out large reversible transports, is the use of a coordinate constructed from the isopleths of a quasi-conserved tracer.

It was realized about a decade ago, as attention paid to the polar stratosphere increased, that since the polar vortex is clearly defined in quasi-conserved tracers, it made sense to define a coordinate by contours of such a tracer. The "equivalent latitude", $\phi_e$, of a tracer contour whose value is $q$ is defined as the latitude which encloses the same area as the contour,

$$\phi_e(q) = sin^{-1}(1 - \frac{A(q)}{2\pi A_e^2} \tag{4.1}$$

where $A(q)$ is the area and $A_e$ the radius of the earth. In this coordinate, the polar vortex has a very simple geometry. Its sharp edge is not smeared out by averaging, and reversible fluxes, such as described above due to the wobbling vortex, do not occur. This idea was first applied to actual data by Butchart and Remsberg (1986). There is a direct parallel to the common definition of the extratropical tropopause as a PV surface, and consequently

much of the discussion in this chapter applies to computations of extratropical stratosphere-troposphere exchange as well as to those of transport in and out of the polar vortex.

In most of the remainder of this chapter, I consider the problem of computing the transport of material across a tracer isosurface, given data sets representing the flow and tracer fields with some finite resolution in time and space, and possibly some noise as well. Although all the actual calculations performed in this thesis are two-dimensional, the problem is in some important respects the same in two or three dimensions. For the next two sections the three-dimensional problem will be considered, as is particularly relevant to the tropopause; all arguments carry over to the two-dimensional case with only obvious modifications. Some of the methods involve, as one step, the computation of large ensembles of Lagrangian trajectories, using one of several algorithms: reverse domain filling (RDF), forward domain filling, or contour advection (CA). Detailed descriptions of these methods, and some related discussion, are presented in appendix A.

## 4.2 Tracer coordinates: practical implementation

Let the isosurface be denoted by $\mathcal{C}$, and the unit normal to it by $\mathbf{n}$. If we are not to exclude arbitrarily any real physical situations, we must allow $\mathcal{C}$ to be arbitrarily convoluted, and to cross any coordinate axis more than once. The formulation of Wei (1987) assumes that the tropopause is a quasi-horizontal surface whose position in the vertical coordinate is a unique function of the horizontal coordinates

$$\mathcal{C} = Z(x, y) \tag{4.2}$$

but this assumption is violated in tropopause folds, which are generally assumed to be important contributors to stratosphere-troposphere exchange. To deal with this, Wei's formulation invokes the arbitrary and unjustifiable assumption that tropopause folds are cut off where they cross a vertical plane. In my view this is undesirable, and so assumptions like (4.2) should be avoided. Lamarque and Hess (1994) modified Wei's formulation to avoid this assumption, simply by defining a relation like (4.2) locally near the tracer surface, rather than globally, so that globally the surface can be multiple-valued in the vertical.

### 4.2.1 Direct approaches

#### Instantaneous flux calculations

It should be obvious that the mass flux across the surface is, at any given point,

$$F = \rho(\mathbf{u} - \mathbf{u}_c) \cdot \hat{\mathbf{n}} \tag{4.3}$$

where $\mathbf{u}$ is the fluid velocity and $\mathbf{u}_c$ the velocity of the surface itself. $\hat{\mathbf{n}}$ is the unit normal vector to the surface. Note that $\mathbf{u}_c$ has a component along the local tangent plane to the surface that is in principle indeterminate, but that component does not enter into the cross-surface mass flux and so is irrelevant. The above holds, of course, for any surface, not just a tracer isosurface.

If we desire the net transport across $\mathcal{C}$, we need only integrate (4.3) over that surface. If the fluxes of tracers (other than the one used to define $\mathcal{C}$) are of interest, integrations must be separately performed over areas of positive and negative $F$. The *net* mass flux relates in a meaningful way to tracer fluxes across $\mathcal{C}$ only when $F$ is known to have one sign over all

of $\mathcal{C}$.

Having (4.3), it appears as though the problem is almost trivially solved, except for a straightforward numerical calculation. In principle, no fancy tricks should be needed to compute the transport across the polar vortex edge, the midlatitude tropopause, or any other surface in which we may be interested. If we have data which allow us to determine the position of the surface at any time, then by suitable time differencing we can estimate $\mathbf{u}_c$, and then if we have data for $\mathbf{u}$ as well we can compute the RHS of (4.3) directly. In the atmosphere, we do have global, three-dimensional observational analyses for wind and PV, so we should be all set. The "model" of Wei (1987) for the computation of transport across the tropopause is, in fact, nothing but (4.3) written in a form that is independent of the choice of vertical coordinate, with the additional questionable assumption (4.2) added on, presumably for mathematical convenience.

The problem is simply the accuracy of the available data. As is well known, the process of computing a PV field from wind and temperature fields involves differentiation in space, and hence amplifies grid-scale noise. Three-dimensional gridded analyses of winds and temperatures are themselves subject to significant uncertainty, so PV is more so. This must lead to a component of $\mathbf{u}_c$ which is purely noise-induced. We may think of this simply as the "jitter" velocity of the analyzed PV isosurface. This will not likely be matched by identical jitter in $\mathbf{u}$, so that a spurious noise-induced component of transport will contaminate calculations using (4.3). Integrated over the whole surface and over some interval of time, we may reasonably expect this component to average to zero, but if we are interested not only in the net transport, but also in property fluxes, then the opposing mass transports in both directions across the surface are of interest. It is these which may be strongly contaminated by noise. Clearly the effect of noise will on average tend to increase the property fluxes across any surface rather than decrease them.

Despite both the importance of this point and its simplicity, it does not appear to be universally understood. It may be worthwhile to restate it in slightly more formal language.

Suppose we represent an observational analysis of a tracer field, $q_a$, by

$$q_a(\mathbf{x},t) = q(\mathbf{x},t) + \epsilon(\mathbf{x},t)$$

with $q(\mathbf{x},t)$ the true field and $\epsilon(\mathbf{x},t)$ an error due to instrument noise, incorrect assumptions made in the analysis, lack of resolution in the observational network, and the like. Assume that the analysis data are available at time intervals $\Delta t$. For sufficiently small $\Delta t$ we may approximate the normal component of the analyzed contour velocity, $u_{ca}$, as

$$\mathbf{u}_{ca}(\mathbf{x},t)\cdot\mathbf{n}(\mathbf{x},t) \approx \frac{\partial q_a(\mathbf{x},t)}{\partial t}(|\nabla q_a(\mathbf{x},t)|)^{-1}$$

and estimate the time derivative by a finite difference over $\Delta t$.

Expanding, we have

$$\mathbf{u}_{ca}(\mathbf{x},t)\cdot\mathbf{n}(\mathbf{x},t) \approx [\frac{\partial q(\mathbf{x},t)}{\partial t} + \frac{\partial \epsilon(\mathbf{x},t)}{\partial t}](|\nabla q_a(\mathbf{x},t)|)^{-1}$$

The key point to realize is that for the error term to be a serious problem, it is not at all necessary that

$$|\frac{\partial e(\mathbf{x},t)}{\partial t}| \gg |\frac{\partial q(\mathbf{x},t)}{\partial t}|$$

If the tracer isosurface of interest truly lies within a robust transport barrier, this implies almost by definition that $\mathbf{u} \approx \mathbf{u}_c$ over most of the surface most of the time, so that to the extent that the surface moves it undergoes primarily reversible undulations. The proper question, then, is whether

$$|\frac{\partial e(\mathbf{x},t)}{\partial t}| \gg |\frac{\partial q(\mathbf{x},t)}{\partial t} - \mathbf{u}(\mathbf{x},t) \cdot \nabla q_a| \tag{4.4}$$

that is, whether the spurious nonconservative term implied by the noise is much larger than the true nonconservative terms. If (4.4) holds, then the contribution of the error term will dominate the calculation of the cross-surface flux $F$ and the direct approach will suggest that the isosurface is far more permeable than it really is. In this case the calculation is essentially useless, unless the error contribution is small enough that the upper bound or "noise floor" that it sets is, despite its being much larger than the true transport, nonetheless sufficiently low as to provide a useful constraint on the problem of interest.

Data sets which we would consider quite "good" for other purposes are nonetheless likely to fail the test posed by (4.4), since it involves comparing the errors in the tracer field tendency not to the actual tendency but only to its nonconservative part, which is usually small compared to the advective part, this being more or less the practical definition of a "tracer".

To my knowledge (4.3) has been directly applied to real data only in studies of midlatitude stratosphere-troposphere exchange, in particular those of Hoerling et al. (1993) and Siegmund et al. (1996). It is difficult to assess with certainty the effect of noise on the results of these studies, since the authors themselves did not seem to consider the question directly. However we can estimate the noise level which would be required to produce the transports they deduced. It is worth mentioning that both studies used the formulation of Wei (1987), including its suspect treatment of tropopause folds.

Consider the results of Siegmund et al. (1996). Using ECMWF first-guess fields, those authors obtained the rather surprising result that the net extratropical stratosphere-to-troposphere mass flux is a small residual of opposing upward and downward components, each of which is typically 40 times larger than the net [1]. This suggests a tropopause which is much less of a barrier to tracer transport than is commonly assumed. A typical value they obtained for the instantaneous mass flux is of order $\sim 10^{-2} kgm^{-2}s^{-1}$. Letting the density of air be $\sim 0.3 kgm^{-3}$, a typical midlatitude tropopause value, one finds that this mass flux corresponds to a tropopause velocity, relative to the fluid velocity, of about $3 cms^{-1}$. Siegmund et al. used three-hourly data; a sustained velocity of $3 cms^{-1}$ over this period leads to a displacement of about $300m$. The authors quote the vertical resolution of the data as about 25 hPa in the tropopause region. This corresponds to more than $300m$, so that the very large cross-tropopause transports computed by Siegmund et al. (1996) could be *entirely* due to noise-induced fluctuations in tropopause height of less than one grid spacing. The fact that their flux computations were dominated by very small-scale features (see their figure 5, and note that in this plot the results have already been subject to heavy spatial smoothing) might be seen as further reinforcing this interpretation. Any uncertainty

[1] The net downward mass flux through the extratropical tropopause is relatively well constrained, certainly to order of magnitude and probably to a factor of two or better, by dynamical arguments and multiple diagnostic approaches applied to observations. See Holton et al. (1995) and references therein for further discussion. The separate upward and downward components of the mass flux, which must be known in order to estimate tracer fluxes, are much more poorly known at this time.

in the vertical velocity (and there surely is some) would only make the situation worse. The authors performed a sensitivity test in which they reduced the temporal resolution to 6 and then to 12 hours. This did reduce the computed transport significantly, especially for 12 hour resolution. The difficulty is that one might expect a genuine contribution from motions with time scales less than 12 hours, were these motions well-resolved, so that this sensitivity test does not directly prove anything about the effect of noise. The authors simply concluded that 3 or 6 hours was the preferable choice of time step, but this assumes a negligible contribution from noise and so seems unfounded as well.

### Contour crossing

The above discussion is meant as an example to illustrate the difficulty of the direct approach when applied to real data. Usually we are interested particularly in surfaces across which transport is inhibited, that is, transport barriers. Hence we are interested in transport where it is a minimum, and any spurious increase in computed transport due to noise effects is particularly unwelcome. A modification of the direct approach involves the use of Lagrangian trajectories to estimate a time integral of (1) directly, without explicitly computing $\mathbf{u} - \mathbf{u}_c$ at each instant in time first. This method, which has been called the "contour crossing" method by SPW, has been used in several studies of the polar vortex, including Dahlberg and Bowman (1994), Manney et al. (1994), Eluszkiewicz et al. (1995), and Knudsen et al. (1997).

Contour crossing works as follows. One initializes a large number of hypothetical point parcels throughout the domain of interest, i.e. on both sides of the transport barrier, and filling the space as densely as is computationally practical. Each parcel is marked by the value of the tracer field, e.g. PV, at its initial location. The parcels are then numerically advected forward in time for some specified period. At the end, one interpolates the tracer field valid at that time to each parcel's final position to determine a final tracer value. If the parcel's tracer value has changed, it has "moved" in tracer space, and if the change is such as to move it from one side of the isosurface to the other then it is considered to have crossed that surface. The parcel is considered representative of a finite volume of air, which is determined by the resolution of the grid on which the parcels are initialized. The parcels can be made to lie on a regular grid either at the beginning of the calculation or at the end. In the latter case the advection calculation must be done in reverse, in which case the method is called "reverse domain filling". See appendix A for further discussion of domain filling methods.

The disadvantage of contour crossing compared to the direct method is that the uneven distribution of particles which must obtain either at the beginning or the end of the calculation makes the identification of each particle with a fixed volume of air at least mildly suspect. Of course each particle does represent a finite volume, but that volume may become very elongated and distorted by the flow, and the extent to which the point particle's position continues to reliably represent the entire volume is in general unclear. If the flow is highly divergent (i.e. in a two-dimensional calculation) the problem becomes worse since substantial "holes" in the mass of particles can develop, leading to undersampling of those regions. In predominantly rotational flow regions this problem is ameliorated, and domain filling methods appear to represent the evolution of tracer fields in such regions quite well, as long as the calculations are not run too long[2].

[2]Ten to fifteen days appears to be a reasonable duration in the lower stratosphere.

However, the advantage of contour crossing relative to the direct approach is that the Lagrangian advection builds in an integrating effect which may be expected to reduce the effect of noise somewhat, more so for calculations of longer duration. Consider the tropopause example discussed above. The "jitter" would cause particles immediately on one side or the other of the tropopause to appear to cross back and forth repeatedly. However, at the end of the calculation, any particle can have crossed at most once in total. An implicit integration has been done. If there really is very rapid mixing across the interface then this is misleading. Notice that the maximum transport possible according to contour crossing (rather, the maximum which yields a zero net mass flux) is simply the exchange of all the fluid on one side with an equal mass of fluid on the other. Yet for sufficiently strong mixing and a sufficiently long calculation the entire mass of the region could turn over more than once. In this case there is either no transport barrier, or the duration of the calculation is far too long. Quantitatively, we might say that in using the contour crossing method, one effectively assumes that the residence time of a parcel on either side of the transport barrier is very long compared to the duration of the advection calculation. Hopefully, the assumption is made with good reason, or it can be checked after the fact. If it is correct, contour crossing will be superior to the direct approach in many situations, for sufficiently high particle density.

Nonetheless, noise-induced jitter can still pollute contour crossing calculations to a significant extent. SPW showed by direct example, as well as an idealized analytical thought experiment, that such calculations can lead to major overestimates of transport in and out of the polar vortex, for plausible calculation parameters and noise level in the data. A large contribution of noise to these calculations is also consistent with the results of the three other studies mentioned in the first paragraph of this section. Dahlberg and Bowman (1994) and Manney et al. (1994) both obtained substantially larger transport, on average, in and out of the vortex than did Eluszkiewicz et al. (1995). The latter used numerical model output while the former two used observational analyses. Numerical models, while having their own problems, are free of observational noise, and in general are constrained to exhibit greater consistency between the wind fields and quasi-conserved tracers than are observational analyses. Thus all else being equal we would expect Eluszkiewicz et al. (1995) to have the smallest transport of the three, if the dominant contributor to the estimates of the other two were noise. All else is not necessarily equal, though (the model used by Eluszkiewicz et al. does not necessarily mimic the real stratosphere perfectly), so the conclusion cannot be this clear-cut, but the comparison is nonetheless suggestive. We may likewise note that Lamarque and Hess (1994), who did a calculation comparable to that of Siegmund et al. (1996) but using a mesoscale model simulation rather than observational analyses, likewise obtained much smaller transports (per unit of tropopause area) than did Siegmund et al. (1996). This is despite the fact that Lamarque and Hess' calculation was not global, but centered on a region of space (and time) in which a tropopause fold was occurring, so that the transport might have been expected to be larger averaged over that region than over the entire extratropics as was the calculation of Siegmund et al. (1996).

We may reasonably conclude from all this that for very "clean" data sets, i.e. high-resolution numerical model output, either instantaneous flux calculations or contour crossing may give reasonably accurate results. For realistic observationally derived fields, the viability of both methods is at best suspect.

### 4.2.2 Less direct, advection-based approaches

Two other approaches have been developed for estimating transport across tracer surfaces, using Lagrangian advection methods in conjunction with more subtle techniques for diagnosing the irreversible transport from the results of the Lagrangian calculations. One of these methods will be the subject of the next chapter. Here, a brief introduction to the other, which should serve to illustrate the idea behind both, will be presented. At the moment, these methods have been used only in two dimensions, and so in this section we will refer to tracer contours rather than isosurfaces.

As a first step to either of these methods, one performs a Lagrangian advection calculation, using either CA or RDF with given (either observed or simulated) tracer and wind fields as inputs. In the stratospheric case, the calculations are typically done on isentropic surfaces in order to achieve the two-dimensionality. This is justified as long as the calculation's duration is short compared to the radiative time scale; the latter is of the order of a month or more throughout much of the stratosphere, being shorter at higher altitudes. One then has a representation of a high-resolution tracer field which has evolved from a lower-resolution tracer field. Next, an algorithm is invoked which distinguishes somehow between "large-scale" and "small-scale" features — typically, vortex filaments — in *physical* (as opposed to spectral) space. The filaments are then either literally or figuratively "chopped off". The idea is that the material in the filaments will be mixed with its immediate environment relatively rapidly.

If CA is used, the technique of contour surgery (Dritschel 1989a; see also Waugh and Plumb 1994) may be used in the manner developed by Waugh (1992) and subsequently used by D. Waugh and coworkers (Waugh et al. 1994a; Plumb et al. 1994; PWP; Waugh 1996) to remove the filaments from the parent contours. This involves applying the surgery algorithm repeatedly, with successively larger values of the scale cutoff parameter, until the desired value is reached, yielding a contour which differs from the original only in that the filaments have been removed. This has been dubbed the "coarse-graining" method. The area enclosed by the filaments is then summed and considered proportional to the amount of mass transported across the contour. Hereafter, the use of CA together with coarse-graining for the quantification of transport will be referred to as "contour advection with coarse-graining" (CACG). If desired, the area calculation can be replaced by a more accurate mass calculation by accounting for the variation of the density along the isentropic surface, though in many situations this makes only a modest difference (Nakamura 1995). A conceptually similar technique to CACG based on RDF, known as "local gradient reversal" (LGR), is presented in detail in chapter 6.

The advantage of these methods is that the tracer fields, which are typically noisy, are used only in the initial conditions, and not at the final time as in contour crossing. The tracer field at the final time is determined purely by the advection algorithm acting on the initial tracer field. Some, of the initial noise in the tracer field tends in practice to be smoothed out by the advection [3]. Also, since the tracer field is sampled only at a single time, jitter in that field cannot enter the calculation in any explicit way. The idea is that the transport, and the presence of barriers to transport, are determined most directly by the flow, and in many situations the flow is better known than the PV or other tracers.

[3] A more quantitative and detailed examination of this effect would be worthwhile. At this time no more than this subjective and qualitative statement, based on substantial experience, can be offered. An important caveat is that noise can also lead to the generation of spurious filaments. The competition between these two effects is discussed to some extent by SPW.

Additionally, the advection calculation is essentially an integrator of the flow properties, so that noise in the flow fields is averaged out somewhat and the results are most sensitive to large-scale flow features, which are relatively robust (Waugh and Plumb 1994).

The disadvantage of these methods is that any technique for distinguishing small-scale from large-scale flow features must have a certain *ad hoc* aspect, and typically one or more free (or quasi-free) parameters, in particular some sort of scale cutoff, must be invoked. Sensitivity tests to variations in these parameters are then essential.

## 4.3 Theory of tracer coordinates: modified Lagrangian mean diagnostics

Some theoretical aspects of the use of tracer coordinates have recently been worked out by N. Nakamura (1995, 1996). These studies outline an approach to computing transport across a tracer isosurface which is different from those discussed in the preceding part of this chapter. This section briefly reviews these studies and discusses some issues which they raise.

It is clear that in tracer coordinates, strictly speaking, there can be no transport without some sort of nonconservative process. By a nonconservative process I mean anything which causes the tracer to be not conserved following the fluid motion. Thus diffusion is included, though it is conservative globally, as opposed to a true source or sink. If there are no nonconservative processes, so that

$$\frac{\partial q}{\partial t} + \mathbf{u} \cdot \nabla q = 0$$

with $q$ the tracer, then tracer contours are material lines, and no fluid parcel can cross them.

The modified Lagrangian mean (MLM) framework developed by Nakamura (1995, 1996) makes the relationship between nonconservative processes and transport in a tracer coordinate explicit. The framework requires two independent quasi-conserved tracers which can be used as "vertical" and "meridional" coordinates. Potential temperature is a good choice for the vertical coordinate; PV can be used as a meridional one, though a chemical such as $N_2O$ would work as well.

### 4.3.1 3D MLM

In 3D, the quantity of primary interest is the mass enclosed by a tracer contour, defined as

$$m(q,\theta,t) = \int\int_A \sigma dA = \int_q^{q_{max}} dq' \oint_{\delta A'} \frac{\sigma dl}{|\nabla_\theta q'|} \tag{4.5}$$

where $A$ is the area enclosed by a contour of constant $q$, $\delta A$ represents the contour itself, and $\nabla_\theta$ is the horizontal gradient operator on the isentropic surface. The above expression assumes that $q$ increases towards the pole, as PV in the northern hemisphere does. $q_{max}$ is the "equivalent pole" or global maximum value of $q$. If it decreases, the sign of the inequality in the integral over $q$ must be reversed. $\sigma$ is the isentropic density,

$$\sigma = -\frac{1}{g}\frac{\partial p}{\partial \theta}$$

Nakamura defines the mean of any quantity in this framework as

$$\mathcal{M}(\,) = \int\int_A (\,)\sigma dA = \int_q^{qmax} dq' \oint_{\delta A'} \frac{(\,)\sigma dl}{|\nabla_\theta q'|} \tag{4.6}$$

so that $m = \mathcal{M}(1)$. After being subjected to this averaging operation, the mass continuity equation becomes

$$\frac{\partial m}{\partial t}|_{q,\theta} = -\frac{\partial \mathcal{M}(\dot{q})}{\partial q}|_{\theta,t} - \frac{\partial \mathcal{M}(\dot{\theta})}{\partial \theta}|_{q,t} \tag{4.7}$$

where $\dot{q}$ and $\dot{\theta}$ are the nonconservative "sources" (as defined above) of $q$ and $\theta$ respectively. Though in the above expression it is not immediately apparent, it is important to keep in mind the factor $|\nabla q'|$ in the rightmost definition of the operator $\mathcal{M}$ given in (4.7). This means that if gradients in the vicinity of a tracer contour are large, the nonconservative terms must be relatively large in order to change the mass enclosed by the contour greatly. Conversely, where gradients are weak, weak nonconservative processes will have a large effect on the MLM mass budget.

### 4.3.2 Shallow water MLM

For later reference, I present here the shallow water version of the MLM equation; the necessary modifications are straightforward. The shallow water equations are presented in chapter 5.

In the shallow water system, $m$ is defined by

$$m(q,t) = \int\int_{A(q,t)} h dA \tag{4.8}$$

where $h$ is the fluid layer thickness. Of course, (4.8) actually defines a volume rather than a mass: a constant density has been scaled out since it does not enter the problem at hand in any way. The shallow-water MLM mass continuity equation is then

$$\frac{\partial m}{\partial t}|_{q,t} = -\frac{\partial}{\partial q}\int\int_{A(q,t)} h\dot{q} dA + \int\int_{A(q,t)} Q dA \tag{4.9}$$

where $Q$ is an arbitrary mass source. There is, of course, no more vertical derivative $\frac{\partial}{\partial \theta}$, since there is no vertical coordinate. The first term is identical to the first term in 4.7, with the substitution of $h$ for the isentropic density $\sigma$.

### 4.3.3 Discussion

In the earlier sections of this chapter, the focus was on methods for directly diagnosing the rate of transport across tracer isosurfaces in data sets. To the extent that the nonconservative terms acting on the two tracers required by MLM theory are known, the theory provides an independent and rigorous way of computing the transport. There are several problems with this approach, however.

The first is that from stratospheric observations, certainly not all, and perhaps none, of the relevant nonconservative terms are known with the necessary accuracy for calculations of this sort to be useful. Processes one might expect to be important are radiative heating, and some sort of small-scale mixing, due perhaps to breaking gravity waves (though the

latter are observed to occur only intermittently at least in the lower stratosphere).

The second problem, not entirely distinct from the first, is that tracer contours can become extremely convoluted in the stratosphere. This is not only demonstrated directly by aircraft observations of chemical tracers (which provide a one-dimensional slice through the tracer field), but is to be expected based on the presumption that any effective "eddy diffusion" one might posit to model the effect of small scale mixing in the stratosphere is small. Some information about the sort of fine-scale structures one can expect can be gained from advection calculations and numerical models, but all stratospheric observations with global or near-global coverage, e.g. those from satellites, give only highly smoothed pictures of tracer fields. This means that the coordinate surfaces themselves are not well known when real data are interpreted in the MLM framework. This is a rather strange and unfamiliar difficulty.

Nakamura (1996) developed some further diagnostic techniques for use in the MLM framework. A brief discussion of these will serve to demonstrate both the theoretical elegance of the MLM framework and the difficulty of applying it directly to real data. The central quantity is the "equivalent length", which is formally not the same as the actual contour length, but in practice is generally quite close to the latter. The two are exactly equal when the tracer gradient's magnitude is constant around the contour. $L_e$ thus is strongly related to the amount of fine-scale structure in the tracer field, and is impossible to diagnose accurately from observations in regions of significant stretching and folding. Assuming that small-scale mixing in the atmosphere can be represented by a single constant diffusivity $D$, the transport across a contour is directly related to the product $DL_e^2$. Leaving aside the issue of whether invoking a uniform diffusivity is plausible, this result is not as simple as it looks. For a fixed flow field but lower $D$, tracer contours can become more convoluted, so that $L_e$ itself depends on $D$ in some way which is quantitatively unknown, though we may expect that $L_e$ will in general be a decreasing function of $D$, all else fixed.

Consider a state of statistical equilibrium in which there is a fixed, steady source of tracer in some region of space, a sink of equal strength somewhere else, and chaotic mixing due to a time-dependent, nondivergent flow conspires with some weak diffusion to maintain $L_e^2$ statistically steady in time for all tracer contours. Imagine that the flow field is prescribed, but the diffusion coefficient $D$ acting on the tracer is allowed to vary between realizations of this thought experiment. Take the tracer source to vary only on scales much larger than that over which diffusion acts.

As $D$ is allowed to approach zero, it is plausible to hypothesize that the net transport across the contours ought to become independent of $D$ as long as the latter is not identically zero. Assume is that in a statistical sense, the large-scale configuration of the tracer contours does not vary as a function of $D$. Because of this, the area-integrated mass tendency within a given contour due to the large-scale tracer source is likewise independent of $D$. Since the diffusive transport, related to $DL_e^2$, must balance the net tracer source, it follows that the former is independent of $D$ as well. This implies that in this limit[4], $L_e^2 \sim D^{-1}$. If this is true, then is it not rather awkward to try to estimate the product $DL_e^2$ by first estimating $D$ and $L_e^2$ separately? It would seem more sensible, and less susceptible to errors, to try to find some way to estimate the product directly.

It may be noted that the former argument is most clearly valid for regions where chaotic mixing is active, so that kinematic stirring controls the rate at which diffusion can destroy

[4]It is assumed here that $D$ does not vanish identically. The case of zero $D$ is theoretically interesting, but surely represents a singular limit. In a real sense it is thus irrelevant to any real physical fluid.

tracer variance. In true transport barriers, marked by bands of sharp tracer gradients which deform in a purely reversible wavelike manner, the transport may remain dependent on $D$ even in the small $D$ limit, as in the nonchaotic flows studied by Flierl and Dewar (1985). In this case kinematic approaches are useless, and one cannot quantify the transport without estimating $D$ and $L_e$. The latter is relatively easy in this case since the contours are smooth, but estimating $D$ remains at least difficult.

With these MLM diagnostics, then, one has a compact and elegant framework for computing transport in tracer space. The problem is that it requires knowledge of two quantities which are at present only very weakly constrained by observations. Nonetheless one may anticipate that these diagnostics may be useful for diagnosing numerical model output, since there one has not only access to all nonconservative terms, but also a complete representation of the tracer field, since there is no subgrid-scale variability.

Nakamura and Ma (1997) computed $L_e$ from $N_2O$ fields taken from satellite observations. They were, sensibly, unwilling to claim either that their calculations of $L_e$ were accurate given the low resolution data, or that they could justifiably assign a value for $D$. Because of these uncertainties they could not obtain quantitative estimates of transport, and settled for qualitative information such as the location of transport barriers.

## 4.4 Lobe dynamics

One other, rather different approach to computing the permeability of transport barriers from flow data has recently seen substantial development, and as such deserves mention here. This approach is based on ideas from dynamical systems theory. Some interesting work on mixing in idealized, engineering-like contexts from a perspective of this sort is reviewed by Ottino (1989, 1990). However, the only attempts to date of which I am aware which use dynamical systems ideas to obtain actual transport rates (as opposed to less directly useful quantities, such as Lyapunov exponents) for flows resembling real geophysical ones are all very recent, and owe more to the work of Stephen Wiggins (e.g. Wiggins 1992). The attempts to which I refer are those of Duan and Wiggins (1996), Malhotra and Wiggins (1997), Miller et al. (1997), Rogerson et al. (1997), and Ngan and Shepherd (1997).

In this approach, what is of interest is the transport across the *manifolds* of the flow's *hyperbolic fixed points* (these terms will not be defined here; the interested reader can consult the references), rather than across tracer isosurfaces. For simple periodic flows, the technique known as *lobe dynamics* allows particularly efficient computation of transport across the manifolds (e.g. Wiggins 1992). A satisfactory explanation of the mathematical basis and practical implementation of this approach is beyond the scope of this thesis [5]. However, a couple of general remarks are hopefully not out of order, though their meaning may not be entirely clear without recourse to the references mentioned above.

The first remark is that until very recently, the technique had been used only on purely periodic flows. As originally developed, the technique relied heavily on a "strobed" view of the flow, or *Poincare map*, which is an inherently periodic object. Real geophysical flows are never purely periodic, and so one might question how these techniques might be extended to geophysical problems. Miller et al. (1997) extend the technique to an aperiodic flow through a plausible, though technically *ad hoc*, numerical procedure. However, they do not address in great detail the question of what the technique's results really mean in this case,

[5] as well as being probably beyond the ability of the author.

given that the technique still relies heavily on essentially periodic concepts. Further, their figure 6 suggests that the aperiodic flow considered is still quite close to periodic, so that one can legitimately wonder how quickly the technique's applicability will degrade as the flow more closely approaches the "red" spectrum typical of real geophysical flows.

This issue has been addressed by Malhotra and Wiggins (1997), who also present methods for extending lobe dynamics to truly aperiodic (as opposed to quasiperiodic) flows. They provide a more detailed mathematical basis for the procedure, making use of results due to Beigie et al. (1994). Some problems, however, remain open at present.

The technique described by Malhotra and Wiggins (1997) requires (as must any technique utilizing manifolds) that the locations of the hyperbolic points in the flow be determined as a first step. In periodic flows these points are fixed in space, and their location can be straightforwardly determined. In aperiodic flows they can move, and locating them is more difficult. The authors discuss methods for locating the hyperbolic points only for flows which can be separated into a small-amplitude time-dependent part[6] superimposed on a much larger-amplitude steady part. Many realistic flows will not fit this description. Wiggins (personal communication) suggests that in strongly time-dependent, aperiodic flows, the hyperbolic points can still be located, though the methods for doing so are still being developed.

Another point is that even if one is completely satisfied by the appropriateness of computing transport across manifolds in aperiodic flows, these manifolds are still few in number (the prototypical 2D "cat's eye" flow is bounded by two such curves), and as such give only limited information on mixing in the space, though one can argue that the manifolds define the transport barriers, and so it is the most important information which is obtained. However, it is clear that at least in principle, a tracer coordinate offers more information, as it allows transport to be quantified throughout the space. Also, one expects that near the manifolds, contours of any quasi-conserved tracer will align along the manifolds, so that the two descriptions are similar, and by using a tracer coordinate one is not losing the ability to define the transport barriers correctly.

On the other hand, manifolds have the advantage over tracer contours that they are defined in terms of the flow only, and as such have greater generality than do contours of any particular tracer. While all tracers whose sources and sinks are sufficiently weak should eventually develop nearly coincident contours, the fact that the coordinate is, in general, influenced by the sources and sinks may be seen as undesirable from a certain theoretical point of view. There may also be instances in which tracer data of adequate quality are not available but flow data are; lobe dynamics is probably the only viable approach in this case.

In summary, dynamical systems approaches at present have both advantages and limitations. The cursory evaluations given here of these advantages and limitations should certainly not be taken as definitive. Lobe dynamics has been shown to be very useful in highly idealized problems. It has only very recently begun to be developed along the lines necessary for application to real geophysical problems. Because all techniques for diagnosing mixing have significant limitations, the best approach to any important problem is probably to use a number of different methods. Lobe dynamics clearly should be considered part of the suite of potentially useful tools.

[6] which, however, can be of arbitrary form.

# Chapter 5

# Methods

## 5.1 The model

The shallow water model is the one used by PWP. The model dynamical equations are

$$\frac{\partial \zeta}{\partial t} = -\nabla \cdot (\mathbf{u}\zeta_a) + \nu\nabla^6\zeta$$

$$\frac{\partial \delta}{\partial t} = -\frac{1}{2}\nabla^2(\mathbf{u}\cdot\mathbf{u}) + \mathbf{k}\cdot\nabla\times(\mathbf{u}\zeta_a) - g\nabla^2(h+h_b)$$

$$\frac{\partial h}{\partial t} = -\nabla\cdot(h\mathbf{u}) + Q$$

where $\zeta$ is the relative and $\zeta_a$ the absolute vorticity, $\delta$ the divergence, $\mathbf{u}$ the fluid velocity, $h_b$ the fixed bottom topography, and $h$ and $Q$ are again the layer thickness and mass source. As in PWP, $Q$ takes the form of a relaxation back to a prescribed, zonally symmetric "radiative equilibrium" state, $h_e$:

$$Q = \tau_h^{-1}(h_e - h)$$

The model equations are solved on the sphere by a spectral method; the code (without the mass source and topography) is described by Browning et al. (1989). The model has been modified by L. Polvani to carry a passive tracer, using the same advection algorithm as is used for the dynamical fields. For this work, two nonconservative terms have been added which act on this tracer. The resulting equation for the tracer, labeled $a$, is

$$\frac{\partial a}{\partial t} = -\mathbf{u}\cdot\nabla a - \tau_a^{-1}(a-a_e) + a(h-h_e)(h\tau_h)^{-1} + \nu\nabla^6 a$$

the term $\tau_a^{-1}(a-a_e)$ is a highly idealized "chemistry" term which relaxes the tracer towards an equilibrium value $a_e$. $a_e$ will be prescribed as a function of latitude here. The term $a(h-h_e)(h\tau_h)^{-1}$ exists to compensate for the effect of the height source term on the tracer budget. Multiplying the above equation by $h$, and using the dynamical equation for $h$ yields the result

$$\frac{\partial(ah)}{\partial t} = -\nabla\cdot(ah\mathbf{u}) - \tau_a^{-1}h(a-a_e) + h\nu\nabla^6 a$$

In other words, the only nonconservative processes acting on the total tracer amount are the hyperdiffusion and the chemistry term. The term $a(h-h_e)(h\tau_h)^{-1}$ prevents the mass source from adding or removing tracer. Other choices are possible. One might add a term in

the equation for the column amount $ah$ but not in the equation for the concentration $a$, so that fluid added or removed by the mass source would have the same tracer concentration as already present in the column. There is some arbitrariness in the choice, as there is no demonstrably correct way to represent the effect of vertical advection in a single-layer model. The main results of the present research turn out to be insensitive to this choice. The reason is that, as will be seen, in the polar vortex region, the mass source nearly vanishes in an average around PV contours, which is the form of averaging used.

It should be made clear that the purpose of this thesis is *not* to elaborate on the basic properties of the model simulations, nor to explore the parameter space of the model. Much of this has been done by PWP, including many experiments not explicitly mentioned in that paper (R. A. Plumb, personal communication). The approach here is to take simulations similar or identical to those performed by PWP as providing plausible representations of certain aspects of the stratospheric flow. The simulations are then viewed as a test bed for trying new techniques and examining the validity of theoretical ideas. Rather than running the model over a wide range of conditions and applying a few simple diagnostics to the results, we take a particular simulation (or two) and apply a variety of detailed, and in large part new diagnostics to them for the purpose of studying the transport in a more thorough way than is typical. The goal is not only to gain understanding of the dynamical and transport properties of a stratosphere-like flow, but also to take the problem of diagnosing these properties as seriously as it deserves to be taken (but often is not).

Even besides the lack of 3D effects, the details of the simulations are not as realistic as they could be if a more complex model were used and/or more tuning of the forcings, boundary conditions etc. were performed. For example, the polar vortex is a bit too large and too strong relative to the real one. However, I take the view that there is some value in using a simple model. The fact that the model requires little tuning to produce a reasonable polar vortex may be seen as a sign that the polar vortex is in its most basic aspects a fairly robust structure. It should be made clear that the simulations studied here represent a particular dynamical regime, in which the vortex remains at all times coherent and no major warmings occur. The real polar vortices are observed to be in this regime most of the time, so it is worthy of study. Nonetheless the absence of more major disturbances is a limitation, whose implications are discussed further in chapter 8.

## 5.2 Simulation particulars

Two model runs have been diagnosed in this study. In both runs, the radiative relaxation time is set to $\tau_e = 10$ days, and the "radiative equilibrium" profile $h_e(y)$ is the same as used by PWP (shown here in figure 5-1). The topographic forcing is also the same as in PWP,

$$h_b = H_B(1 - e^{-t/\tau})e^{-[(\phi-\phi_0)/\Delta\phi]^2}cos(\lambda)$$

with $H_B$ equal to 1500 meters, $\phi_0$ 45°N, $\Delta\phi$ 15°, and $\lambda$ is longitude. $\tau$ is a "ramp-up" time, taken to be 5 days. In the first run, which is identical to that shown in figure 8 of PWP except that it was run for 150 instead of 100 days, the hyperdiffusivity was set to $5 \times 10^{26} m^6 s^{-1}$. This gives a 5 hour damping time for total wave number 42, so that even though the model was run at T85, the effective resolution is really approximately T42. This run will be referred to as the "T42" or "low-resolution" run. The second run was identical except for having a hyperdiffusion coefficient of $1 \times 10^{25} m^6 s^{-1}$, a factor of 50 smaller than in the first experiment. This yields a true T85 resolution, with a damping

time of about 5 hours on total wave number 85. This one will be referred to as the "T85" or "high-resolution" run.

Model output was archived every 24 model hours, at a spatial resolution of $1° \times 1°$. Since T85 spectral resolution corresponds roughly to 1.4° grid resolution, this means that the fields were slightly oversampled to avoid aliasing.

All the diagnostics have been performed for a 50 day period, beginning on day 80 of the given model run, by which time the model was in a nearly statistically steady state. We chose a 50 day period because model wave breaking events tend to occur somewhat irregularly, with quiet and disturbed periods of roughly 10-20 days duration. 50 days was judged long enough to allow us to interpret averages over this period as reasonably representative of the model's climatology for this particular set of forcings.

Figure 5-2 shows the profiles of PV vs. equivalent latitude on the first and last days of our 50-day period, for both runs. Figure 5-3 shows the mass inside each contour (defined by equation 4.8) as a function of PV on these two days. In both runs the model is clearly near a steady state, as defined by these diagnostics, during this period. Also, the structure is very nearly the same between the two runs, supporting the assertion that the dynamics are at most weakly dependent on the hyperdiffusivity.

Figure 5-4 shows the time evolution, beginning on day 80 of the T42 run, of the mass enclosed by three PV contours: a surf zone contour ($0.8 \times 10^{-8} m^{-1} s^{-1}$), an "entrainment zone" contour ($1.1 \times 10^{-8} m^{-1} s^{-1}$), and a contour in the main vortex edge ($2.5 \times 10^{-8} m^{-1} s^{-1}$). The meaning of these labels for different regions of PV (or equivalent latitude) space will be made clearer in the next chapter. For the moment, notice that in the first two contours, there are large, slow fluctuations, while in the third there is much less variability, both absolutely and as a fraction of the average mass.

Because of the importance of the hyperdiffusion to the work presented in chapter 6, the T42 and T85 runs will both be analyzed in that chapter. In chapter 7, only the T42 run will be analyzed.

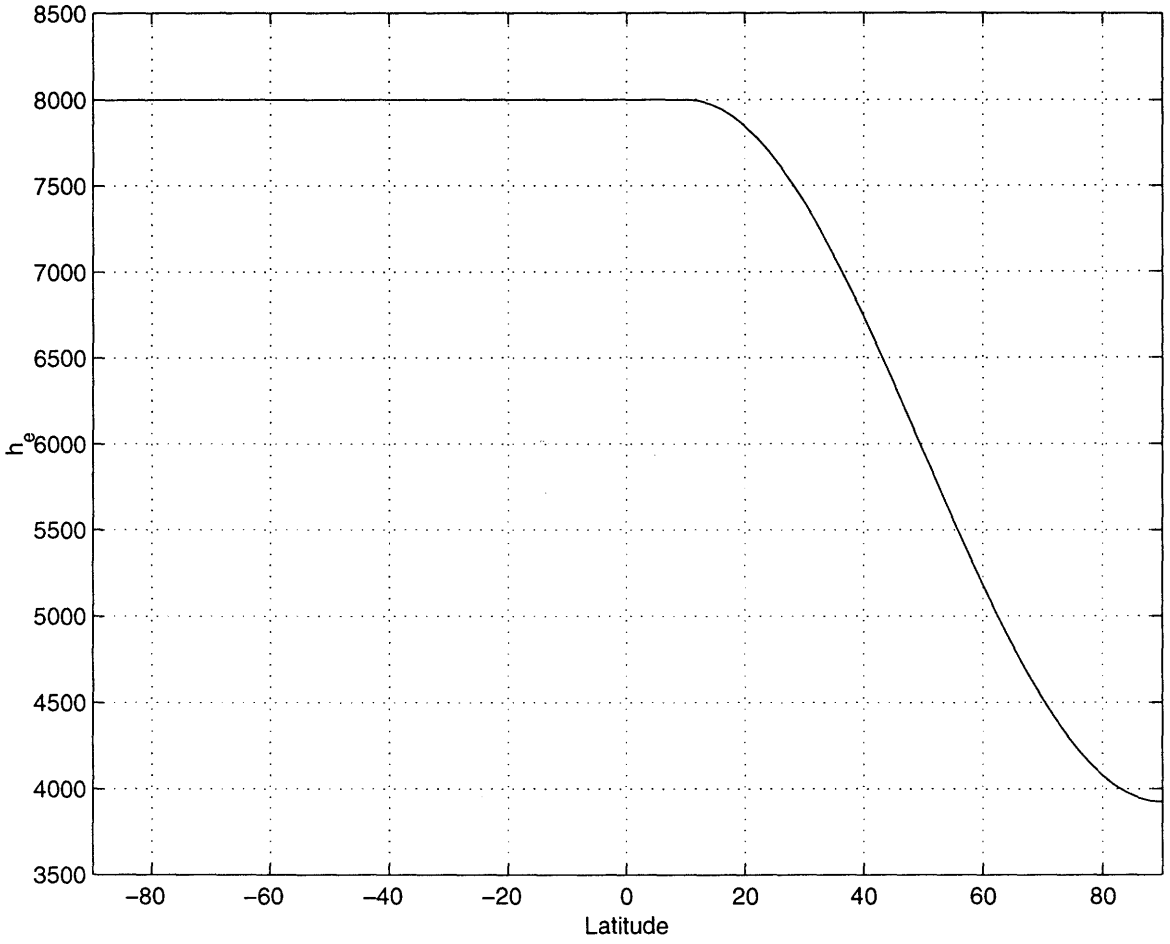


Figure 5-1: "Radiative equilibrium" layer thickness profile, $h_e$ (meters), used in the shallow water model, following PWP, as a function of latitude.

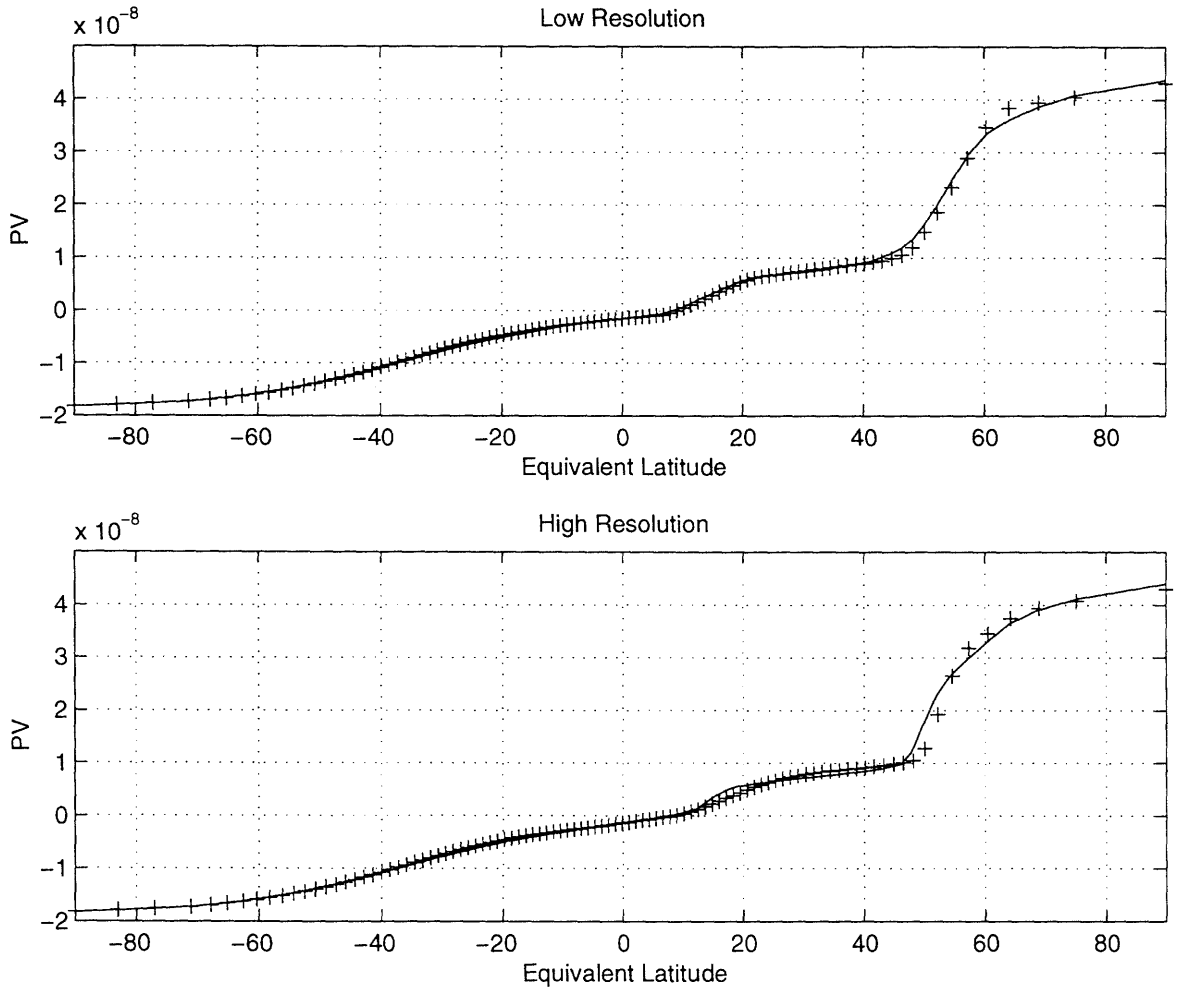


Figure 5-2: PV vs. Equivalent latitude, for both the low-resolution and high-resolution runs. Solid lines represent day 80, pluses day 130.

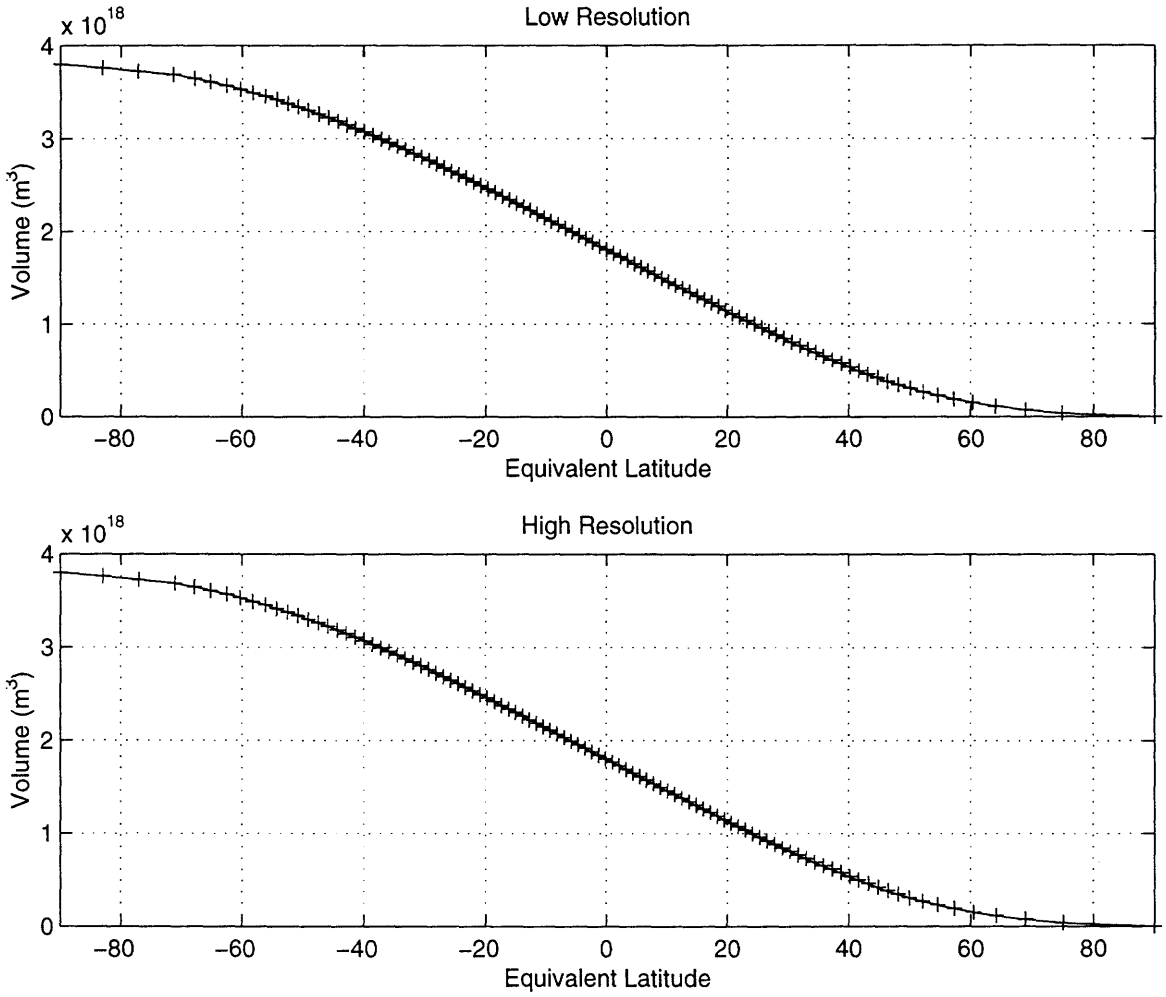


Figure 5-3: Mass enclosed by PV contours (on the northern side) vs. Equivalent latitude, for both the low-resolution and high-resolution runs. Solid lines represent day 80, pluses day 130.

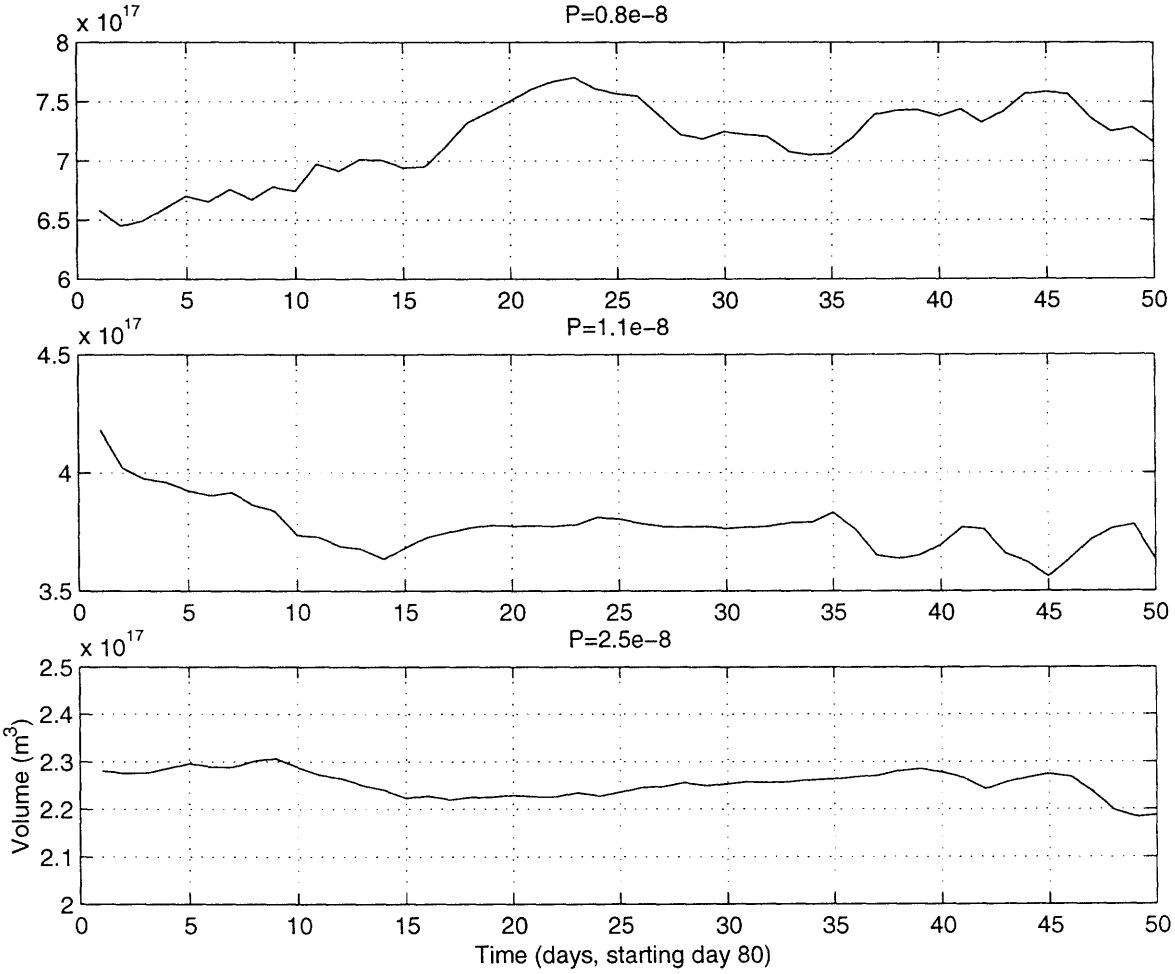


Figure 5-4: Mass enclosed by three different PV contours as a function of time, starting on day 80, for the low-resolution run.

# Chapter 6

# Shallow water results I: MLM and LGR Diagnostics

## 6.1 Introduction

This chapter contains the most significant results of the thesis. It focuses explicitly on the forced-dissipative, shallow-water polar vortex. Interest centers on the nature of the balances which maintain the vortex in a statistically steady state, the roles of the nonconservative processes and their relationships to the large-scale dynamics, and on the quantitative diagnosis of the exchange of mass between the vortex and its exterior. By the use of MLM diagnostics, the roles of the nonconservative processes can be understood in a detailed and quantitative way. This in turn forms the basis for a definition of transport which is precise and has a theoretical basis. The LGR method, to be described, is then shown to give results roughly consistent with this definition. LGR and CACG are also shown to give similar results to each other. The end result is an improved understanding of the shallow water polar vortex, and explicit justification for an approach to computing transport which up to now has been defended solely by plausibility arguments.

## 6.2 Inviscid MLM steady state in the shallow water system

A particularly simple result is derived here which applies to the shallow water MLM system when PV is used as the coordinate, and the system is inviscid but contains a mass source. This result is new, and is of intrinsic theoretical interest. It will also prove useful in interpreting the numerical results to follow.

By a well-known identity, the first term in equation 4.9 may be rewritten in terms of an area integral around the boundary of $A$, the result of which is

$$\frac{\partial m}{\partial t}|_{q,t} = \oint_{\delta A(q,t)} \frac{h\dot{q}}{|\nabla q|} dl + \int\int_{A(q,t)} Q dA \tag{6.1}$$

where $\delta A$ is the tracer contour bounding the region $A$, and $dl$ is an increment of scalar distance around $\delta A$. Note that for the above to remain finite we must assume that $|\nabla q|$ does not vanish identically anywhere, except at isolated maxima and minima. If we assume that the maxima and minima are quadratic (using a Taylor expansion and assuming that the leading order terms do not vanish), then the gradient will go to zero linearly at such

points. However, the contour length will shrink to zero linearly as well, so that the contour integral in (6.1) will remain finite.

Consider the MLM mass budget of a shallow-water atmosphere which is inviscid but subject to thermal forcing. Let the tracer which defines the MLM coordinate be the shallow water potential vorticity $P$,

$$P = \zeta_a / h$$

For general $Q$, but no forcing on $\zeta_a$, one obtains

$$\frac{dP}{dt} = -\frac{PQ}{h} \tag{6.2}$$

Substituting (6.2) into (6.1), the result is

$$\frac{\partial m}{\partial t}|_{P,t} = -P \oint_{\delta A(P,t)} \frac{Q}{|\nabla P|} dl + \int_P^{P_{max}} dP' \oint_{\delta A(P',t)} \frac{Q}{|\nabla P'|} dl \tag{6.3}$$

where the area integral in the second term has been rewritten in terms of contour integrals. $P_{max}$ is the "equivalent pole", that is, the global maximum value of $P$. Now defining

$$F(P) = \oint_{\delta A(P,t)} \frac{Q}{|\nabla P|} dl$$

(6.3) becomes

$$\frac{\partial m}{\partial t}|_{P,t} = -PF(P) + \int_P^{P_{max}} F(P') dP' \tag{6.4}$$

Now consider a steady state, so that the LHS is zero. Taking the derivative with respect to $P$, and noting that for positive argument, the integral on the RHS increases as the lower bound decreases, we have

$$F(P) + 2PF'(P) = 0 \tag{6.5}$$

We assume that the steady state holds over a region bounded by the equivalent pole (global maximum value of $P$) on one side, and some lower value of $P$ on the other, and that $P$ has only one sign in this region. By the argument given above concerning local maxima and minima, $F(P)$ remains finite at the equivalent pole. This, and equation 6.4 with the time derivative set to zero, imply that

$$P_{max} F(P_{max}) = 0$$

which implies that $F(P_{max})$ vanishes, since by assumption $P_{max}$ does not. This provides the necessary boundary condition on (6.5), whose general solution is

$$F(P) = CP^{\frac{-1}{2}}$$

with $C$ a constant of integration. The boundary condition requires $C = 0$, so that

$$F(P) = 0$$

is the solution throughout the region. This implies that

$$\int \int_A Q dA = 0$$

also, or in other words that there can be no net radiative disequilibrium in an MLM steady state, unless other nonconservative processes besides thermal forcing are added. This result is a generalization of those of Schneider (1987) and McIntyre and Norton (1990), who reached similar conclusions under assumptions of steady flow. I emphasize that the present result only assumes that $m(P,t)$ is constant in time; this is a much weaker assumption. The relationship between these various results is explained in more detail in appendix B.

Note that locally at each contour, the definition of $F(P)$ means that the integral around the contour of $Q/|\nabla P|$, rather than of $Q$ itself, must vanish. If the steady state is statistical rather than exact, the same conclusion applies to the time-averaged MLM thermal forcing (i.e. it must vanish). This is so since any sensible time-averaging operator is linear, and so the above analysis can simply be repeated with all terms time-averaged.

## 6.3 Numerical results — forced-dissipative vortex dynamics

### 6.3.1 Balances maintaining quasi-equilibrium

I have computed the terms on the right-hand side of the strict MLM continuity equation for every day of the 50-day period, once per day. It may be straightforwardly shown from the model equations that in the numerical model, the shallow water PV obeys the equation

$$\frac{dP}{dt} = P(\tau_e h)^{-1}(h - h_e) + \nu h^{-1}\nabla^6\zeta \tag{6.6}$$

So, for a range of PV values spanning the model northern hemisphere, each term on the RHS of

$$\frac{\partial m}{\partial t}|_P = -\frac{\partial}{\partial P}\int\int_{A(P,t)}[\tau_e^{-1}P(h-h_e) + \nu\nabla^6\zeta]dA + \tau_e^{-1}\int\int_{A(P,t)}(h-h_e)dA$$

has been computed. The hyperdiffusion term was dumped directly from the model's timestepping routine, rather than being diagnosed from the gridded vorticity field. Area integrals have been computed simply by summing the quantity $()a_e^2 cos\phi\Delta\phi\Delta\lambda$ over all points with PV greater than the desired value, where () is the integrand, $a_e$ the radius of the earth, $\phi$ the latitude, and $\Delta\phi$ and $\Delta\lambda$ are the grid spacings in longitude and latitude, which are both one degree. Having obtained these integrals for various values of $P$, the derivatives with respect to $P$ are calculated by centered finite differences. The computations were repeated at different resolutions in $P$ and found to be convergent, that is, insensitive to this resolution for the range we tested.

Having computed the daily values for the three contributions to the mass tendency, I computed a "predicted" change to the total mass inside each contour over the entire period by integrating in time, that is, multiplying each daily tendency by one day and summing them. This was also done without including the hyperdiffusion term, in order to allow this term to be estimated as a residual as an alternative to its direct computation, for reasons made clear below. The "predicted" mass change can be compared to the actual change in mass over the period, the mass at the end minus that at the beginning, computed directly using only the fields of height and PV on the initial and final days. The result of this comparison is shown for both runs in figure 6-1. The $x$ axis is PV, rather than equivalent latitude. This means that relative to physical space, the coordinate is strongly stretched for

PV values greater than about $1 \times 10^{-8} m^{-1} s^{-1}$ and contracted for lower values; this is useful for our purposes since we are interested in the region at and just above $1 \times 10^{-8} m^{-1} s^{-1}$. The agreement between "predicted" and true mass changes is very good at T42. At T85 it is less good, presumably because our technique for computing the area integrals has trouble with the hyperdiffusion term, which varies at the grid scale in this case. The calculation was repeated for intervals of different duration (i.e. other than 50 days) and similar agreement was obtained. Of course this is no great achievement, since the "predicted" value is not really predicted but computed directly from known model quantities, but it gives us confidence in our procedure (e.g. once per day sampling) and numerics, except for the hyperdiffusion term at T85 as noted above. However, this term can still be reliably estimated as the difference between the actual mass change and that computed from the MLM mass budget without including hyperdiffusion. This calculation, displayed more clearly in figures 6-5 and 6-6, shows that the hyperdiffusion term is small, though not zero, in the main vortex edge region, which I define here as $1.5 \times 10^{-8} m^{-1} s^{-1} < P < 3.5 \times 10^{-8} m^{-1} s^{-1}$. In the outermost vortex edge and surf zone, defined here as $P < 1.0 \times 10^{-8} m^{-1} s^{-1}$, the hyperdiffusion term becomes much larger. In between there is a transition region, which I will call the "entrainment zone" ($1.0 \times 10^{-8} m^{-1} s^{-1} < P < 1.5 \times 10^{-8} m^{-1} s^{-1}$). Of the other two terms (the two on the RHS of (6.3)), the area integral of the height forcing is everywhere negligible compared to the boundary term.

In the main vortex edge region, figure 6-1 shows that the integrated net "thermal" forcing of PV along PV contours is very close to zero. This is due not to this term's being locally very small everywhere, but to almost exact cancellation between regions of positive and negative thermal forcing. A contour plot of the thermal forcing on a typical model day is shown in figure 6-2, superimposed on the PV field for that day. In the zonal mean, the thermal PV forcing is everywhere positive throughout the extratropics, but the vortex leans towards the region of negative forcing so that averaged over the main vortex the forcing is approximately zero. The inadequacy of zonal mean diagnostics for capturing essential aspects of polar vortex behavior is highlighted by this. It should be made clear that in the real polar vortex, there may not be large regions of negative PV forcing. In the real atmosphere, there appears to be radiative cooling throughout the polar vortex, presumably driven at any given level by descent forced by the breaking of Rossby or gravity waves at higher levels. Depending partly on the vertical structure of the cooling, this may be expected to increase the PV forcing in the positive sense throughout the vortex in the middle and lower stratosphere. No analog to this effect is present in the shallow-water model.

The fact that the hyperdiffusive contribution to the PV budget is quite similar in both cases, throughout most of the domain and despite a factor of 50 difference in the coefficient, suggests that the magnitude of this term is determined not by the coefficient but by the large-scale dynamics, which are effectively inviscid. This is reminiscent of the famous result in statistically steady 3D turbulence that the energy dissipation is independent of the viscosity. Given this, we may go further, and hypothesize that the contribution of this term should not even depend on the form of the small-scale dissipation, but only on its being weak, scale-selective (so that the large-scale dynamics are in fact effectively inviscid), and qualitatively similar in its behavior to a diffusion operator (that is, $\nabla^2$). This is often implicitly assumed in many sorts of modeling studies.

## 6.4 Local gradient reversal

The central component of the LGR method is a simple pattern recognition algorithm which identifies narrow filaments or small blobs in a tracer field which is otherwise smooth in at least one coordinate direction. These can then be removed, analogously to contour surgery. The assumption, again, is that these features are dynamically passive, and due to be sheared out to ever smaller scales until finally being dissipated. A detailed discussion of an earlier version of the algorithm is given in SPW.

Consider a cartesian domain with coordinates x, y, and grid points evenly spaced and equal in both dimensions, $\Delta x = \Delta y$. Then the tracer field $q(x, y)$ is represented discretely as $q(i, j)$ where i and j are the indices of the two coordinates, $x = i\Delta x$, $y = j\Delta y$. Those points which satisfy either

$$q(i,j) > q(i - n_{long}, j) + \delta q \quad AND \quad q(i,j) > q(i + n_{long}, j) + \delta q$$

or

$$q(i,j) > q(i, j - n_{lat}) + \delta q \quad AND \quad q(i,j) > q(i, j + n_{lat}) + \delta q$$

are considered agents of "outward", or equatorward wave breaking, again using the northern hemisphere convention that the tracer increases towards the pole. The criteria for inward breaking are defined analogously. $n_{lat}$, $n_{long}$, and $\delta q$ are adjustable parameters, though the former two are related (as discussed below) so that together they are equivalent to one free parameter. This defines a minimum "cutoff" spatial scale, features smaller than which are considered destined for dissipation. $\delta q$ defines a minimum "amplitude" for a filament to be considered, so that very small fluctuations in $q$ do not trigger the algorithm.

In SPW, an equal area grid was used for the LGR calculations. Here, a standard lat-long grid is used, which requires some minor modifications to the criteria. Most importantly, $n_{lat}$ and $n_{long}$ are not in general equal, due to the convergence of the meridians. Rather, we choose

$$n_{long} = int[n_{lat}/cos(\phi)]$$

where $\phi$ is the latitude, and int() represents rounding off downward to the nearest integer. $n_{lat}$is held fixed. Though it makes little difference in practice, we also adjust the threshold $\delta q$ in the longitudinal direction. That is, rather than applying the criteria exactly as stated above, we use separate values for $\delta q$ in the latitudinal and longitudinal criteria:

$$\delta q_{long} = \delta q_{lat} n_{long} [n_{lat}/cos(\phi)]^{-1}$$

where the subscripts on the $\delta q$'s have analogous meanings as for the $n$'s. This means that the critical *gradient* is the same in the two directions, while the critical scale over which this gradient must exist is nearly but not exactly the same. $\delta q_{lat}$ (hereafter called simply $\delta q$) and $n_{lat}$ are then the explicit free parameters in the scheme, to which sensitivity must be tested.

In the implementation used in SPW, "transport" was computed simply by counting the points which satisfied the criteria. Here I have made the transport computation more sophisticated, as mentioned above. One locates the points satisfying the LGR criteria as before; in the discussion that follows these will be referred to as the "center points". Then, however, one generates a smoothed tracer field by applying a smoothing procedure to the field in a region extending for $n$ grid points in both latitude and longitude from each center point. The smoothing procedure averages the PV values of the center point with those of

its neighbors to make the region locally homogeneous or nearly so.

In this study I have used only $n_{lat} = 2$ and $n_{lat} = 4$, for fields on a $1 \times 1$ lat-long grid. This amounts to assuming that when features reach a scale between $200 - 450$ km, they become dynamically passive and destined for dissipation. Further, the truncation and smoothing have been applied only to outward breaking filaments, that is, features with PV greater than their environments. This is because we are interested only in the region near the vortex edge, and a careful survey of the model output from the entire period revealed that no inward wave breaking events occurred there. An event would be considered inward breaking if a thin region of low-PV material were drawn into the vortex so that it was surrounded by large regions of higher PV. Various versions of the smoothing procedure have been tested on a number of instantaneous model PV fields. These tests have shown that the only small-scale negative PV anomalies (as defined by the LGR criteria) in the vortex edge region occur when small regions of surf zone air are pinched between the main vortex edge and a narrow vortex filament that is being drawn around an anticyclone. It is inconsistent to call this inward transport at the same time that the formation of the (outward-breaking) filament itself is considered outward transport. Lacking justification for making a more complex scheme to deal with this ambiguity, we simply make the consistent choice that a filament formed in an outward breaking event is considered an agent of pure outward transport, that is, we smooth only positive PV anomalies. The algorithm needs further modification to handle a situation in which filaments wrap tightly around the vortex.

In the smoothing procedure, each point involved is smoothed according to

$$q(i,j) \rightarrow 0.5q(i,j) + 0.125[q(i+1,j) + q(i-1,j) + q(i,j+1) + q(i,j-1)]$$

where $i, j$ are now the indices of each point being affected by the smoothing. The smoothing is applied to each point less than $n_{lat}$ points away from the center point in latitude, longitude, and along the diagonals. The truncation is applied once, but the smoothing is applied three times, after which applying it any additional times would make little or no difference since the region has been fairly well homogenized by this point. Note that the smoothing is applied over the same number of *points* in latitude and longitude, in apparent conflict with the way the LGR criteria are applied. This simplification is made because first, most of the points satisfying the criteria do so with respect to latitude rather than longitude anyway (that is, the filaments tend to be oriented quasi-zonally), and second, the filamentation events we will study occur within a restricted range of latitudes. Because of this the spatial inhomogeneity in the smoothing is slight in practice. The smoothing does, however, become more anisotropic at higher latitudes, due to the convergence of the meridians.

## 6.5 Transport calculations using RDF and LGR

### 6.5.1 Procedure

For both model runs, we applied a spectral filter to the PV and wind fields, producing fields truncated at T15 for the 50-day period of interest. The purpose of this is to remove the small-scale structure so that it can be regenerated by RDF calculations: it also has the effect of making the fields comparable in resolution to observationally derived fields. We then performed 10 5-day RDF runs, spanning the period, using the filtered fields as input. We applied the two-point and four-point LGR smoothing to the results.

Figure 6-3 shows the output from a 5-day RDF run ending on day 10, while figure 6-4 shows the same PV field after processing by the two-point smoother with $\delta q = .125 \times 10^{-8} m^{-1} s^{-1}$. Again only positive (outward breaking at the vortex edge) filaments have been smoothed. The only effect on the region of interest can be seen by examining the vortex filament in the upper right quadrant of the unsmoothed plot, which is partially removed in the smoothed plot.

The difference in mass enclosed by a given PV contour in the smoothed and unsmoothed fields may be treated as an estimate of the amount of mass contained in small-scale structures into which that contour has been deformed over the period. Call this the "LGR estimated transport". If we assume that

1. the small-scale features identified by LGR are due to be sheared out and eventually dissipated by the hyperdiffusion, and

2. in a time average, the integral of the hyperdiffusion around filamented PV contours is dominated by the hyperdiffusive dissipation of filaments, so that the effect of the hyperdiffusion along the rest of the contour is negligible,

then we may take the time integral of the LGR estimated transport over a long period as an estimate of the time integral of the hyperdiffusion term over the same period.

### 6.5.2 Results

Figure 6-5 compares the mass change due to the hyperdiffusion, computed as the residual of the total mass change (the solid curve in figure 6-1) with the terms other than hyperdiffusion (pluses), with the estimate obtained as described above using the two-point LGR filter (solid curves). Figure 6-6 shows the same quantities, but using the four-point filter. In both plots, three values of $\delta q$ (where here $q = P$), spanning a factor of four in this parameter, are shown.

In the main vortex edge region ($1.5 \times 10^{-8} m^{-1} s^{-1} < P < 3.5 \times 10^{-8} m^{-1} s^{-1}$), where the hyperdiffusive contribution to the MLM mass budget is small, the LGR technique also estimates a transport which is very small (or zero), regardless of the choice of $\delta q$ or $n_{lat}$. This is simply because to a first approximation no filaments are formed from the contours in this region. This is important, as it says that where LGR (and presumably, by extension, CACG) indicates the presence of a strong transport barrier, there really is one, and a very robust one in this case, with transport across it of a few percent per month or less.

One notices that, while both are close to zero, there does tend to be a small, consistent difference between the hyperdiffusive contribution to the MLM mass budget and the LGR estimated transport in the main vortex edge region. This is due to the small direct effect of hyperdiffusion on the smooth, un-filamented contours in this region, which the LGR technique cannot capture. This transport is in any case a model artifact, since the hyperdiffusion is one; what controls transport across the un-filamented part of the real vortex edge is not well understood.

In the very outermost vortex edge and surf zone ($P < 1.0 \times 10^{-8} m^{-1} s^{-1}$), the LGR and MLM results sometimes agree and sometimes disagree; in any case the results are sensitive to $\delta q$ and $n_{lat}$ for the parameter ranges chosen, hence not clearly robust. This region is essentially a "filament graveyard", where previously created filaments and small vortices undergo various processes such as decay and merger. Since this region consists primarily of leftover small-scale structure, it is not surprising that the LGR technique, which estimates the rate at which small-scale structure is produced from large-scale structure, does not give

robust results here. The lack of a clean scale separation between vortex and filaments means that which structures are diagnosed as small-scale depends strongly on the details of the definition.

In the entrainment zone (PV between about 1.0 and 1.5 PVU) the LGR and MLM results agree quite well in general. This is a significant result. It means that the LGR technique, which is an ad-hoc but plausible way to estimate the rate of vortex stripping, produces estimates which correspond closely to a rigorous and well-defined macroscopic vortex property: the *degree of radiative disequilibrium.* Remember that the hyperdiffusion is the only process in the model which can balance any non-cancellation of the terms involving the mass source, which is the shallow-water analog of radiative heating. LGR has been shown to produce good estimates of the hyperdiffusion's effect on the mass budget, hence the extent of the non-cancellation among the terms involving the mass source. Since this is balanced by vortex stripping, which leads to the eventual dissipation of filaments, the final conclusion is that the residual of the mass source terms (degree of radiative disequilibrium), the vortex stripping itself (which LGR estimates) and the filament dissipation are all, in a statistically steady state, the same quantity.

### 6.5.3 Sensitivity to duration of calculation

Though it is true that, as we have discussed, the only explicit parameters in our LGR filtering scheme are $n_{lat}$ and $\delta q$, there is really one more parameter in our entire transport calculation method whose role requires some scrutiny. For a given resolution of the grid for the RDF calculations, the duration of the RDF calculation must be considered. The RDF technique, like CA, is truly nondiffusive, but unlike CA it is not absolutely conservative. With either RDF or CA, as the duration of a calculation is extended, finer and finer scale features develop. In the case of RDF, however, features smaller than the grid spacing in any dimension cannot be resolved. As the run's duration is increased for a fixed initial time, the small-scale features eventually simply disappear. For a given grid spacing, then, the optimal run duration would be long enough to allow some fine scale structure to develop, but short enough that that structure does not disappear.

To test the LGR procedure's sensitivity to the run duration, I repeated the calculations for both model runs and both values of $n_{lat}$ using five runs of ten days each spanning the period, instead of ten runs of five days each. The results, shown in figures 6-7 and 6-8 are somewhat smaller transport estimates than in the five-day run case, and are smaller than the MLM results by up to a factor of two. Note, interestingly, that the slope of the curve in the entainment zone is still correct, but the curve is displaced to slightly lower PV. Note also that the depth of the minimum appears to be fairly robustly and accurately predicted in this case. It is not clear whether this last result is an accident, or whether the technique can, if applied subject to some *a priori* constraints on the parameters, reliably predict this minimum in general; it did so in the five-day runs with the two-point filter (figure 6-5) but not with the four-point filter (figure 6-6).

To investigate the issue of duration depencence further, I took a single ten day period, divided it up into periods of periods of one, two, three, five, and ten days' duration, and repeated the LGR calculations in each case (for the three-day periods, the calculations span nine days rather than ten; the results have been multiplied by 10/9 to compensate). The results of this are shown in figure 6-9. The six curves represent the three values of $\delta q$ used in the earlier calculations, with larger transport corresponding to smaller $\delta q$, and the two values of $n_{lat}$ (see figure caption). A ten day period is not long enough that we

can reasonably expect the LGR calculation to agree with the hyperdiffusive contribution to the MLM mass budget over this period. Rather than having a verifiably correct value to which to compare, one ought simply to look for something like convergence, at least in an asymptotic sense. For durations of one or two days, the result is strongly dependent on $\delta q$. At ten days, this dependence is largely gone. However, between five and ten days one begins to see a clear downward trend in the computed transport. This may be expected to continue at longer durations, as filaments have more time to vanish entirely during the course of a single RDF run. The optimal parameter choices are those for which the dependence on parameters is the weakest. At a run duration of ten days, the dependences on $n_{lat}$ and $\delta q$ are smallest, but the dependence on duration itself appears to come closest to having a stationary point at five days, particular if the results for the highest value of $\delta q$ ($0.25 \times 10^{-8} m^{-1} s^{-1}$) are discarded. The latter are outliers in most of the calculations done, in any case; the dependence on $\delta q$ appears minimized — indeed, appears nearly to vanish in the entrainment zone — for the higher two values of $\delta q$ (0.125 and $0.063 \times 10^{-8} m^{-1} s^{-1}$), which still differ by a factor of two from one another. Note, also, that the dependence on $n_{lat}$ and $\delta q$ appears to be smaller in an average over a larger number of five-day RDF runs (see figures 6-5 and 6-6) than in this case where the total period length is only ten days.

One can conclude from these results that there is an optimal range of RDF run duration, centered somewhere between five and ten days, within which the results are relatively insensitive to duration and to the other two parameters. This conclusion has been obtained independently of the comparisons to the MLM results, by considering only the dependence of the LGR results on parameter variations. It can, then, legitimately be viewed as showing that the good agreement between the MLM and LGR calculations is robust, rather than being a consequence of an arbitrary parameter choice.

In general, it is plausible to expect that the optimal duration will scale roughly as the advective time scale (here, a parcel in the jet core makes a complete circuit around the pole in roughly four days) since this is closely related to the time scale for planetary wave breaking.

### 6.5.4 Comparison to CACG

PWP used the CACG technique (see section 4.2.2) to estimate transport across the vortex edge, which they defined as the $1.125 \times 10^{-8} m^{-1} s^{-1}$ PV contour. They computed transport every 10 days of their 100 day run, which is identical to the first 100 days of our T42 run. We have computed transport above for days 80-130. For comparison to PWP's results, we also used LGR to compute transport across the $1.125 \times 10^{-8} m^{-1} s^{-1}$ contour between days 60-100. For this comparison we did calculations of 10 days' duration, and unlike in the calculations presented above, we did *not* perform any spectral truncation or other processing on the wind or PV fields prior to the RDF runs, since PWP did not do so. Again following their procedure, we performed our transport calculations in terms of area rather than mass. This tends to decrease the transport by about 20% when expressed as a fraction of the main vortex, since the fluid is thicker in the outer vortex edge (where the filaments come from) than in the interior.

Table 6.1 shows the results of this comparison. Outward transport across the contour for each 10-day period is shown, as well as the total over the entire 40-day period. Results are shown for the same values of $\delta q$ as used above, as well as two different values of the maximum surgery cutoff scale (the relevant parameter in the coarse- graining technique

| —LGR vs. CA Comparison— | | | | | |
|---|---|---|---|---|---|
| Day | LGR (0.063) | LGR (0.125) | LGR (0.250) | CA (0.10) | CA (0.05) |
| 60-70 | 1.1 | 1.0 | 0.6 | 1.0 | 0.9 |
| 70-80 | 4.0 | 3.6 | 2.4 | 4.5 | 3.9 |
| 80-90 | 3.4 | 2.4 | 0.7 | 3.2 | 2.6 |
| 90-100 | 0.3 | 0.1 | 0.1 | 0.2 | 0.2 |
| Total | 8.8 | 7.1 | 3.8 | 8.9 | 7.6 |

Table 6.1: Comparison of transport calculations using the LGR technique vs. those using the CACG technique, as done by PWP. The numbers in parentheses correspond to $\delta q$ (in units of $10^{-8}m^{-1}s^{-1}$ for the LGR results, and to $dm$ (the dimensionless maximum surgery scale) for the CA results. Transport is expressed as a percentage of the area enclosed by the contour on day 70.

used by PWP) for the CACG results[1]. For the LGR results the two-point smoother was used; results using the four-point smoother (not shown) are not significantly different.

Excepting the LGR results with $\delta q = 0.25 \times 10^{-8}$ (which value also yielded the least satisfactory agreement with the MLM results), the agreement between the LGR and CACG results is remarkably good. This is consistent with the idea that because of the clear scale separation between the vortex and the filaments, where the distinction between the two is drawn is relatively insensitive to the precise criteria used to draw it. It also means that the comparison to the MLM calculations can be viewed as providing concrete support for the coarse-graining technique as much as for LGR.

## 6.6 Discussion

In this chapter, the balance maintaining the model's polar vortex - surf zone structure has been quantitatively diagnosed in detail. The processes which enter the balance are the thermal or radiative mass source, the hyperdiffusion, and the quasi-inviscid large-scale dynamics, which mediates between these two nonconservative processes.

It was proved that if an inviscid shallow-water atmosphere is in an MLM steady state, then the mass source enclosed by PV contours must vanish, which is the shallow-water analog of MLM radiative equilibrium. This result extends earlier results which made the much stronger assumption of steady flow fields, and clarifies the relationship between the different processes in the shallow water model. Any departure from MLM radiative equilibrium must be related to the direct dissipation of enstrophy, which is a generalized measure of the "viscousness" of the atmosphere or some region of it.

I diagnosed the contributions to the strict MLM mass budget from all (two) nonconservative model processes for PV contours in and near the vortex edge region. The main vortex edge region was shown to be near MLM radiative equilibrium, meaning that the integral of the mass source over the area enclosed by the PV contours vanishes. This in turn implies that this region is effectively close to inviscid. The vanishing of the integrated mass source is not due to the source's being very small everywhere, but rather due to the fact that large regions of mass source and sink exist within the contours, which nearly cancel in the area

[1]The CACG results, some of which were not included in PWP, were kindly provided by Darryn Waugh.

(or contour) integral. In contrast, the outermost vortex edge - surf zone region is far from radiative equilibrium, implying that this region is effectively much more viscous. This is clearly due to the prevalence of small-scale structures, which are more readily subject to dissipation.

It was shown that the net hyperdiffusive contribution to the mass budget, even in regions where that contribution is large, is only weakly dependent on the value of the hyperdiffusivity. This offers some support for the hypothesis that in a time averaged sense, it does not particularly matter how vortex filaments are dissipated, as long as the dissipation is sufficiently weak and scale-selective that it does not directly affect the dynamically active large scales. To be sure, a more thorough test of this hypothesis would involve varying the form as well as the coefficient of the dissipative operator. Ideally, one would like to compare the results of a $\nabla^6$ operator to the more physically plausible $\nabla^2$. However, much higher resolution than T85 is required in order to run the model with a $\nabla^2$ diffusion and still obtain an inertial range, that is, vortex filaments. The limited computational resources available for this study, along with time limitations, precluded this step.

The LGR technique was used to estimate the transport across the contours due to filamentation events. This technique incorporates the assumptions that filaments are due eventually to be dissipated once they reach a minimum scale, regardless of the details of the dissipative process, and that the small-scale dissipation contributes to the MLM mass budget primarily by dissipating these filaments. Under these assumptions, the time-averaged net transport across the contours due to the small-scale dissipation can be estimated simply by estimating the mass contained in filaments produced during a given period of time. The LGR filter makes use of a particular definition of a "filament"; the assumption is that at least in a certain regime the results will not be sensitive to the precise definition.

The transport computed by the technique was shown to agree with the the hyperdiffusion term in the MLM budget, except in the outermost vortex edge and surf zone. In the main vortex edge, the LGR technique showed essentially zero transport; this is in near agreement with the MLM results which show this region to be near radiative equilibrium. There was, however, a small hyperdiffusive transport in this region, which could not be reliably estimated by LGR due to its being generally unrelated to filament production[2]. In the entrainment zone, the LGR estimated transport agreed remarkably well with the MLM hyperdiffusion term. The LGR results in the main vortex edge and entrainment zone are insensitive to variations in the scale cutoff parameter $n_{lat}$ and the amplitude threshold $\delta q$.

The integral constraint on a hypothetical inviscid polar vortex shows that the shallow water polar vortex studied here can depart from MLM radiative equilibrium only because of the presence of the hyperdiffusion. The magnitude of the hyperdiffusion term — the effective viscousness — in the MLM mass budget may thus be thought of as a quantitative measure of the *degree of radiative disequilibrium.* This is a well-defined, macroscopic, physically meaningful quantity. The numerical results discussed above show that the LGR technique can provide a reliable estimate of this quantity over much of the region of interest. This shows that the effective viscousness, or radiative disequilibrium, despite its being ostensibly a direct measure of nonconservative effects, *is in fact determined by quasi-inviscid, large-scale dynamics.* It also provides direct support for both the LGR and (indirectly) the CACG method, and guidance in interpreting results from these methods. At present, to

[2]There were a few RDF calculations which did show filamentation in the main vortex edge, which is why the LGR estimated transport is not always identically zero there. This filamentation was never visible in the raw model output, presumably having been suppressed by the hyperdiffusion.

my knowledge, no analogous technique has been tested in this or any comparable way.

In the surf zone, the LGR results sometimes disagree substantially with the MLM results, and are very sensitive to the parameters. This is not surprising, as the technique estimates the rate at which small-scale structure is *produced*, while the surf zone consists primarily of small-scale structure which is left over from previous wave breaking events. The clean scale separation between dynamically active and passive structures is lacking in this region.

The LGR technique was applied to low-resolution versions of the model wind and PV fields, and requires no knowledge of the small-scale dissipation. Hence the technique may be applied as is to observations. The only important difference between observationally derived fields and the smoothed model fields used here is the presence of noise in the former. The effect of noise has not been directly addressed here, but the results presented by SPW (using a precursor of the technique developed here) suggest that LGR and similar techniques (i.e. CACG) will be less sensitive to modest amounts of noise in the data than the direct approach or contour crossing, described in chapter 4.

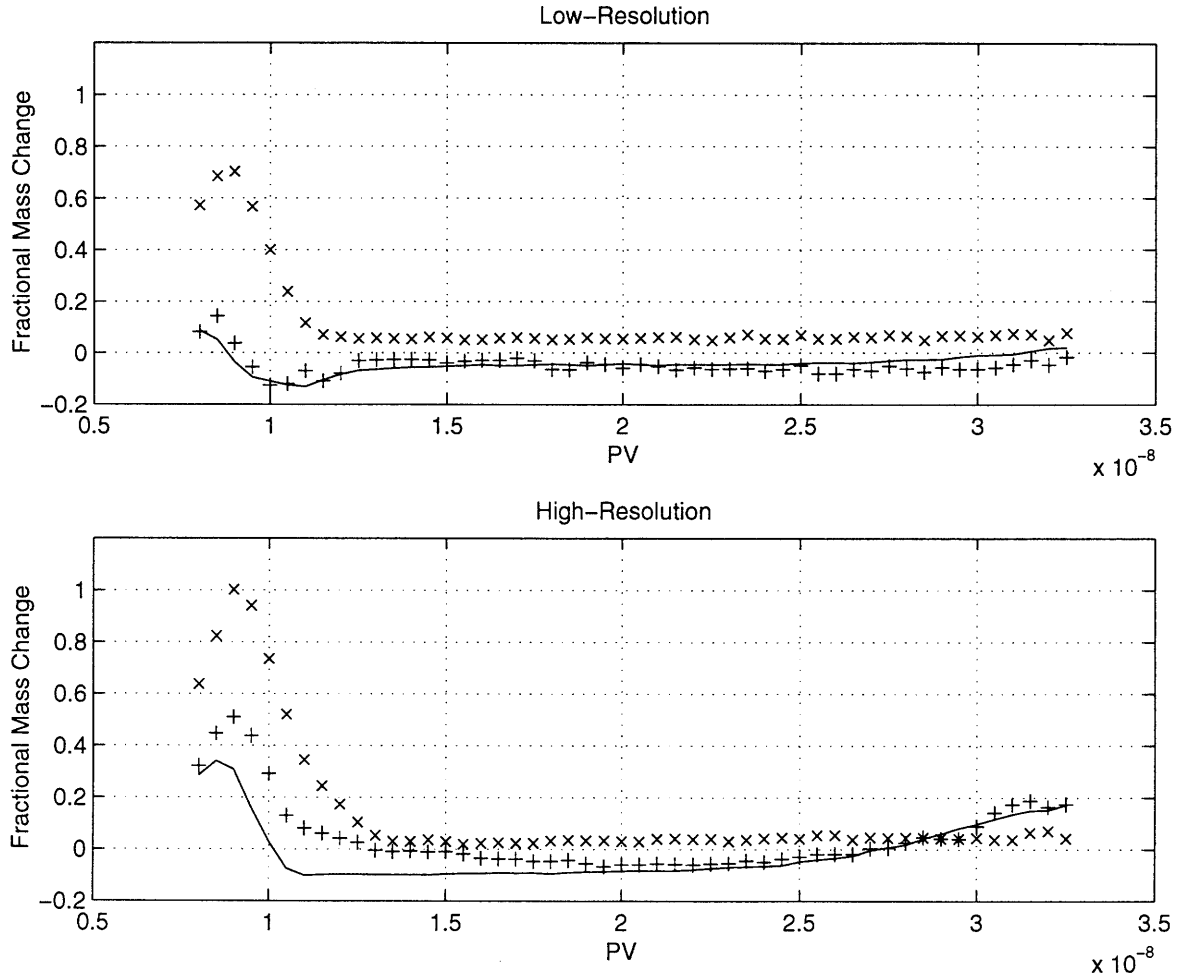


Figure 6-1: Comparison of true change in mass enclosed by PV contours (expressed as a fraction of average mass enclosed by the same contours) between days 80 and 130, and that computed by MLM diagnostics, for the both the low-resolution (T42) and high-resolution (T85) runs. Solid curve is the actual change, pluses the MLM result, and x's the MLM result without including the hyperdiffusion term.

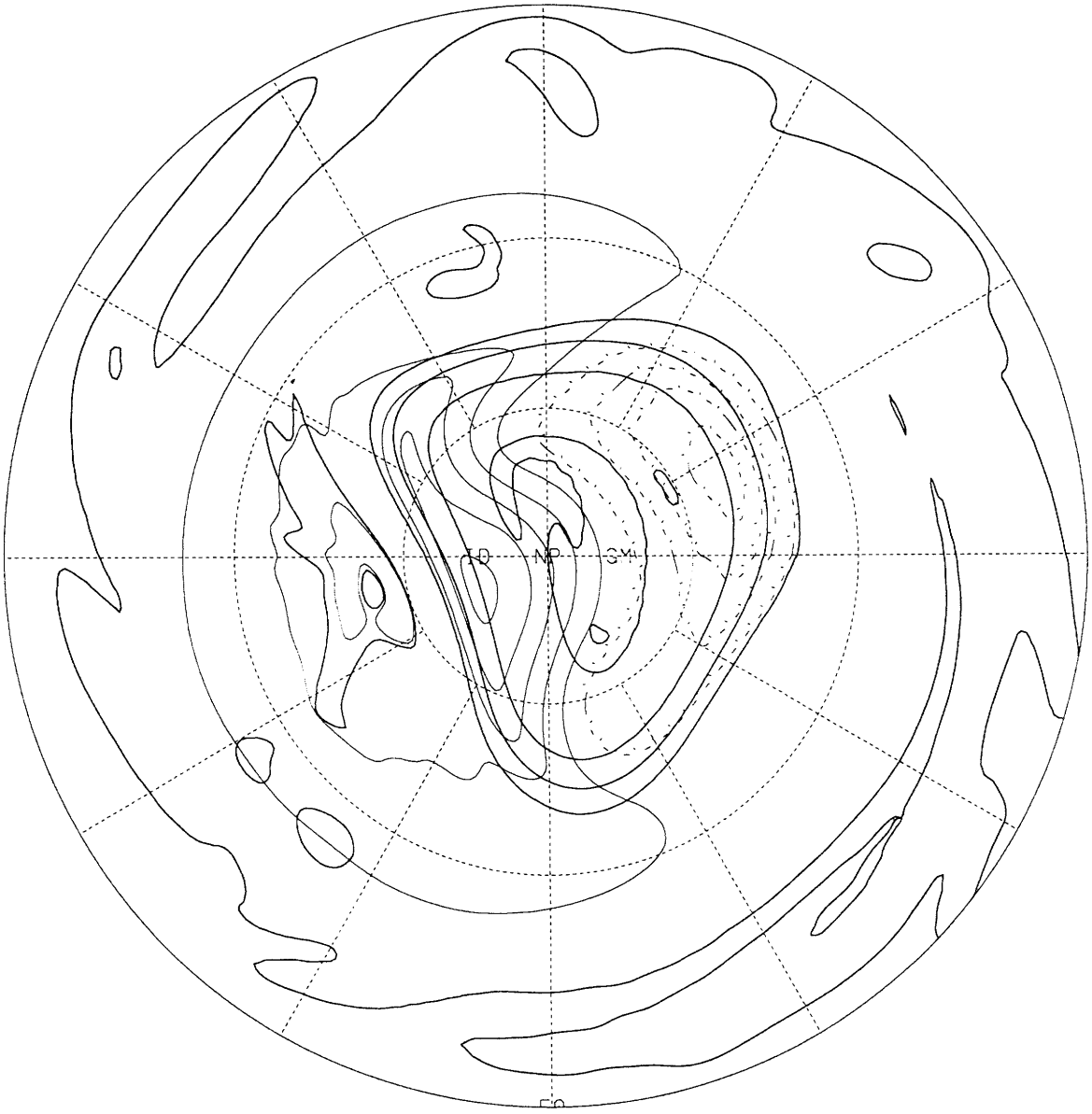

Figure 6-2: Thermal forcing of PV (light contours, negative values dashed) superimposed on PV (heavy contours) on day 100 of the high-resolution run. Forcing contour interval is $2 \times 10^{-15} m^{-1} s^{-2}$, PV contour interval is $1 \times 10^{-8} m^{-1} s^{-1}$.

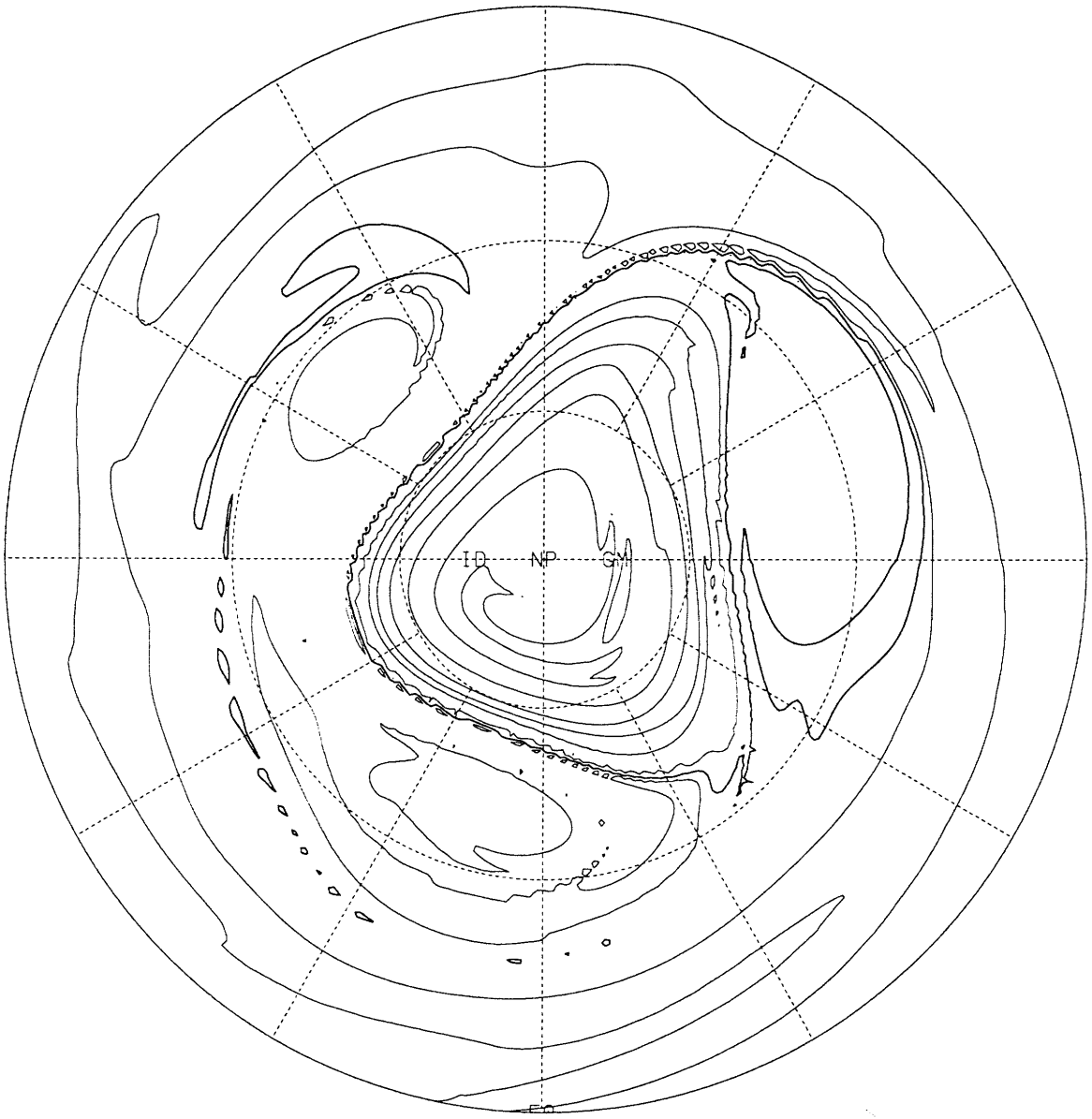


Figure 6-3: PV field resulting from a 5-day RDF run ending on day 10 of the period (day 90 of the simulation). Contour interval is $5 \times 10^{-9} m^{-1} s^{-1}$, heavy contour is $1 \times 10^{-8} m^{-1} s^{-1}$.

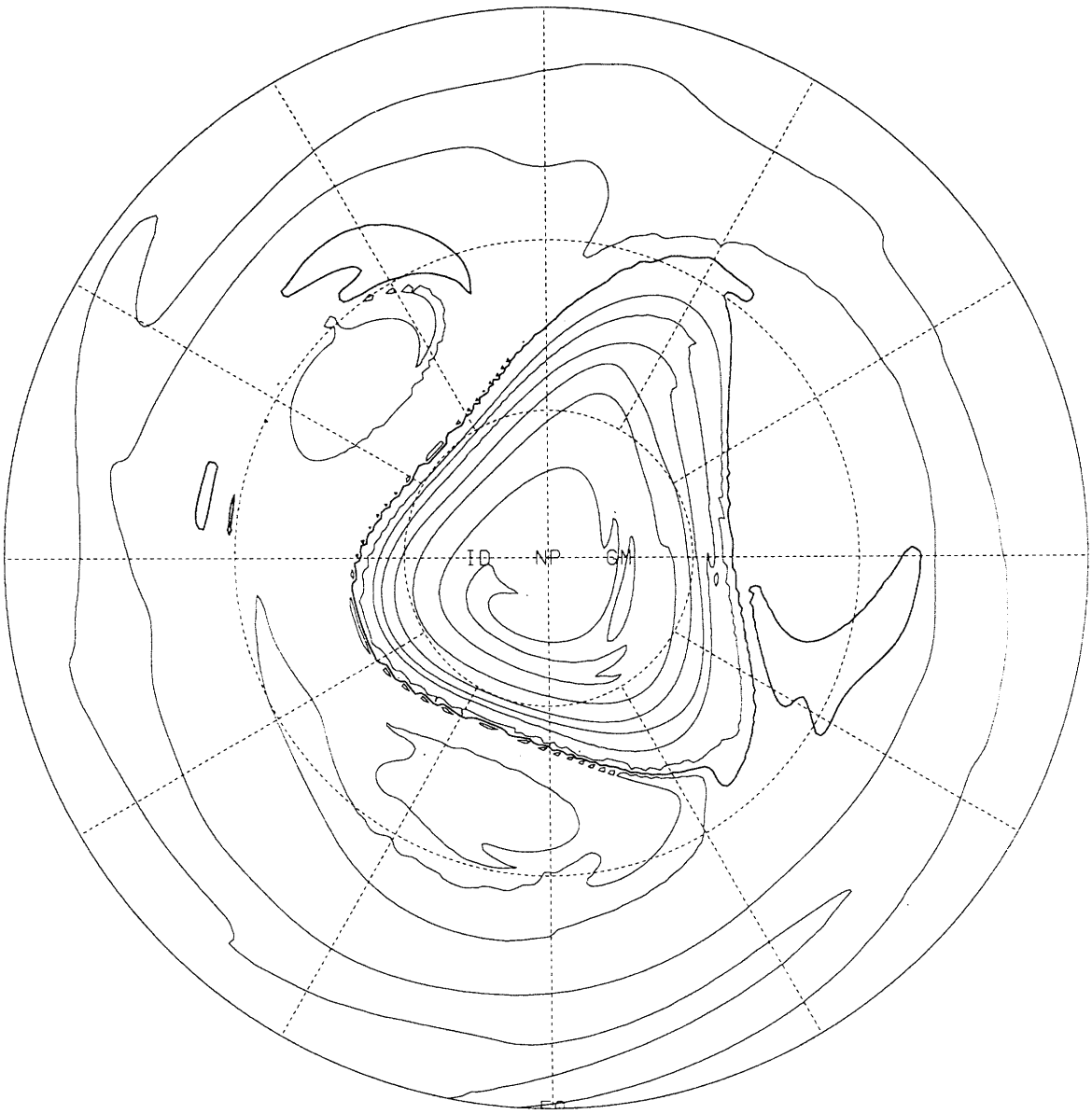


Figure 6-4: PV field resulting from a 5-day RDF run ending on day 10 of the period (day 90 of the simulation), after application of the 2-point LGR filter (see text). Contour interval is $5 \times 10^{-9} m^{-1} s^{-1}$, heavy contour is $1 \times 10^{-8} m^{-1} s^{-1}$.

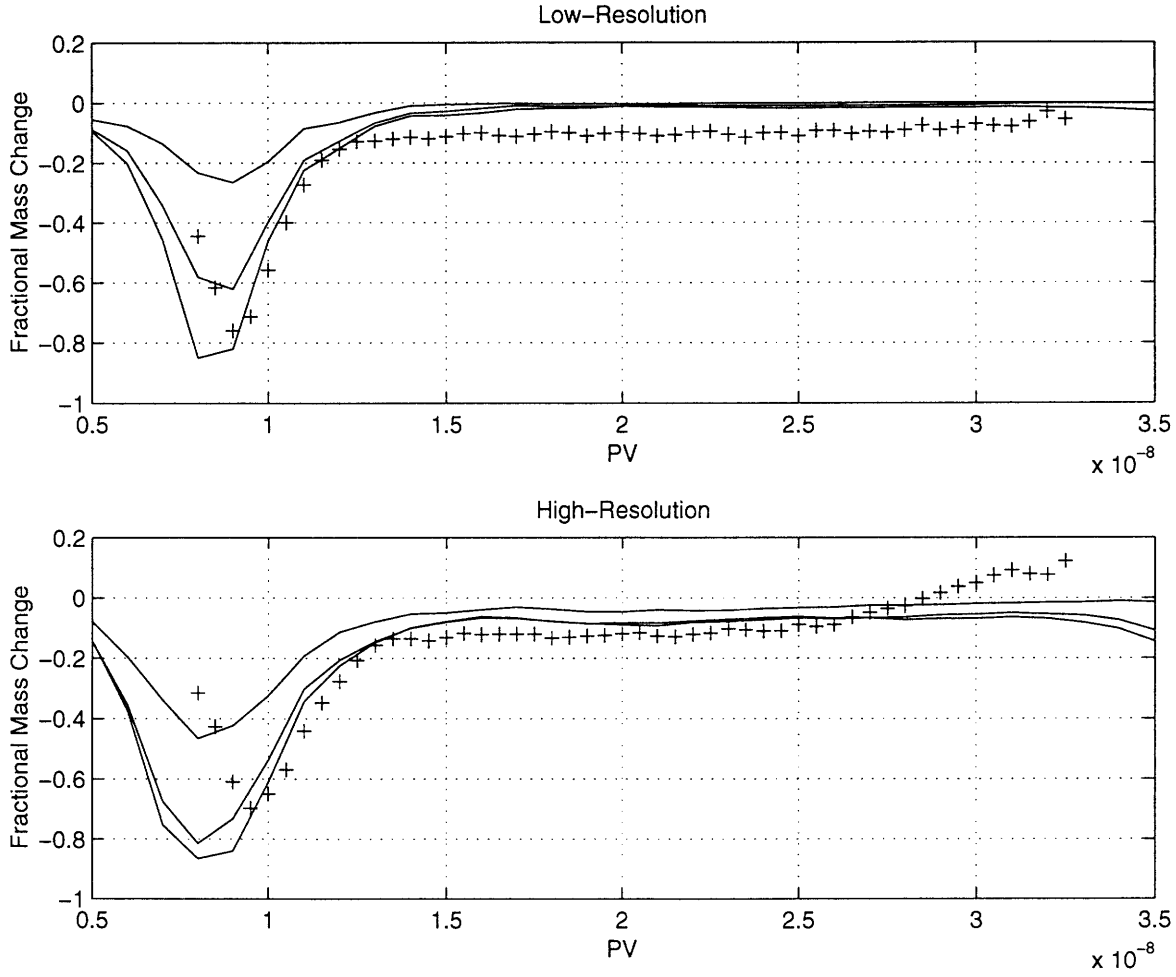


Figure 6-5: Mass transported across PV contours in filamentation events, estimated by the two-point LGR filter (solid curves) and the hyperdiffusion term in MLM budget (pluses). The three solid curves represent three values of the parameter $\delta q$. those being 0.063, 0.125, and $0.250 \times 10^{-8} m^{-1} s^{-1}$.

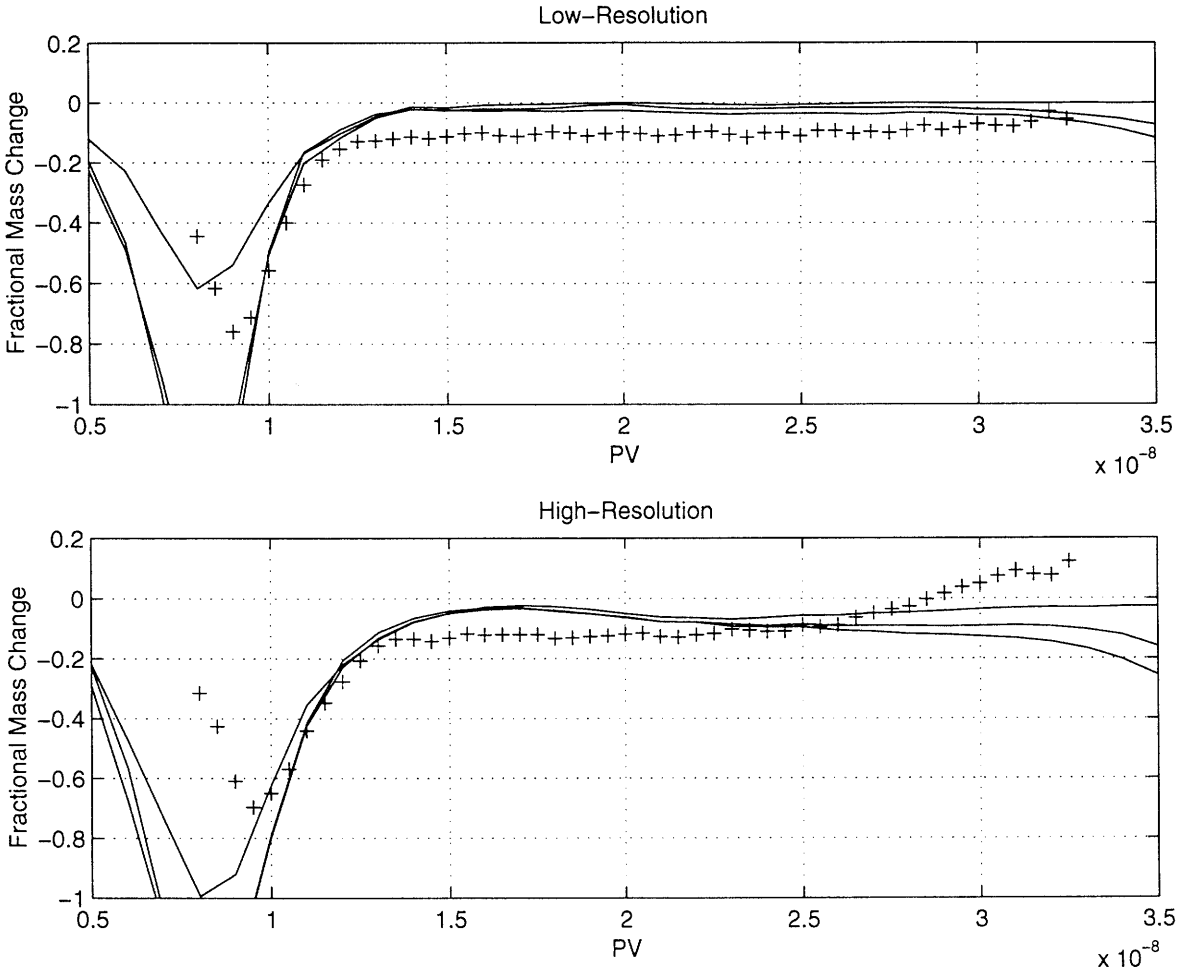


Figure 6-6: As in figure 6-5, but using the four-point LGR filter.

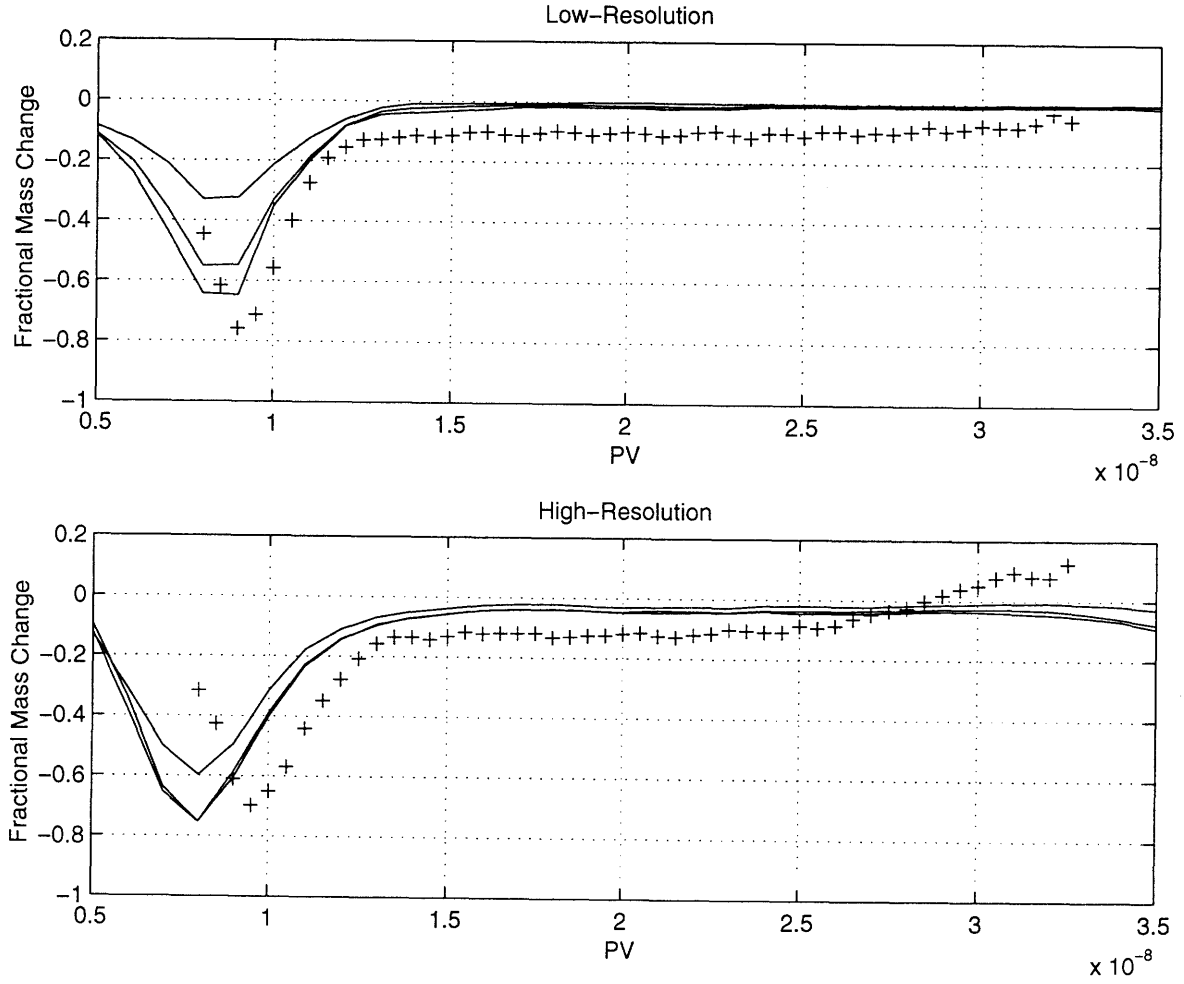


Figure 6-7: As in figure 6-5, but with RDF run durations of 10 rather than 5 days.

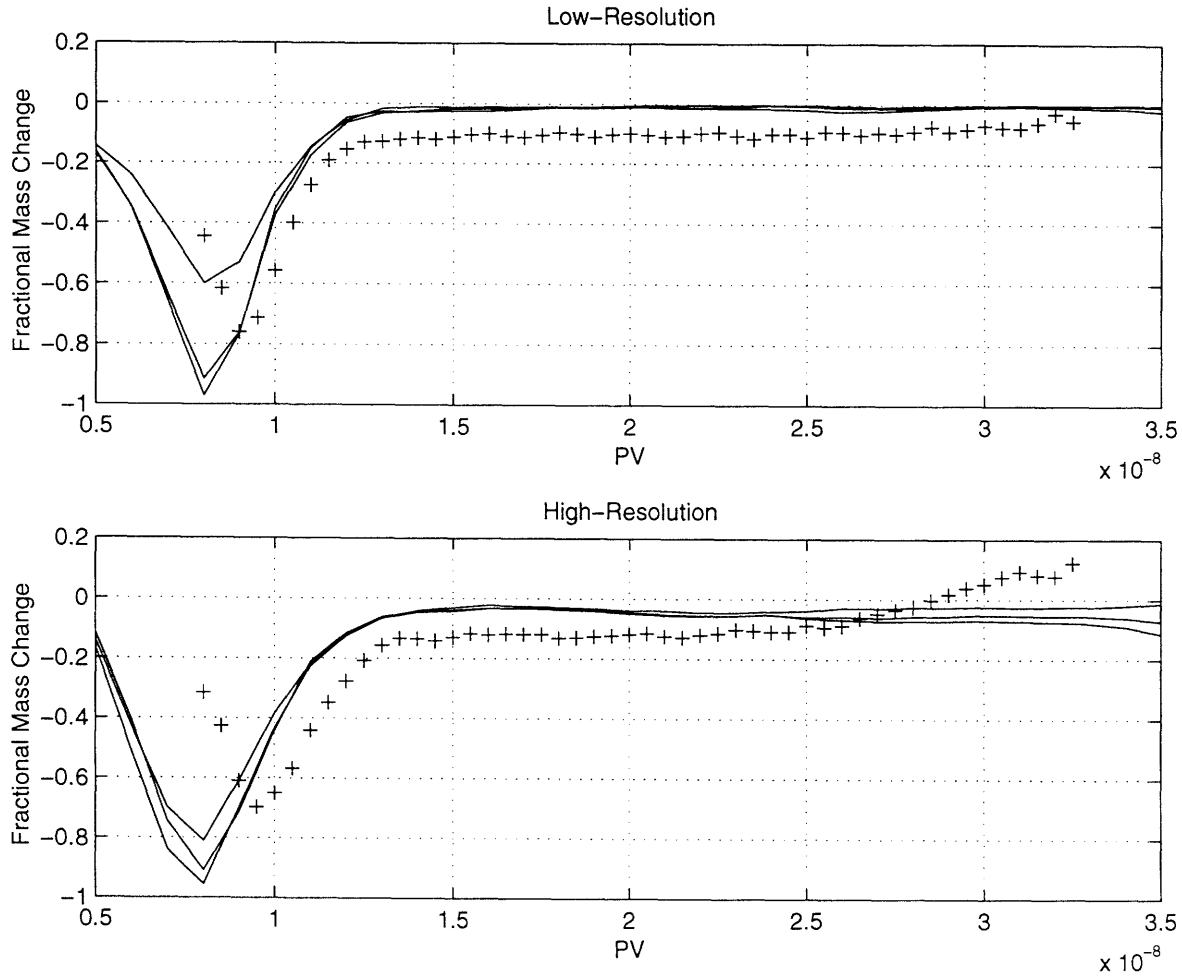


Figure 6-8: As in figure 6-6, but with RDF run durations of 10 rather than 5 days.

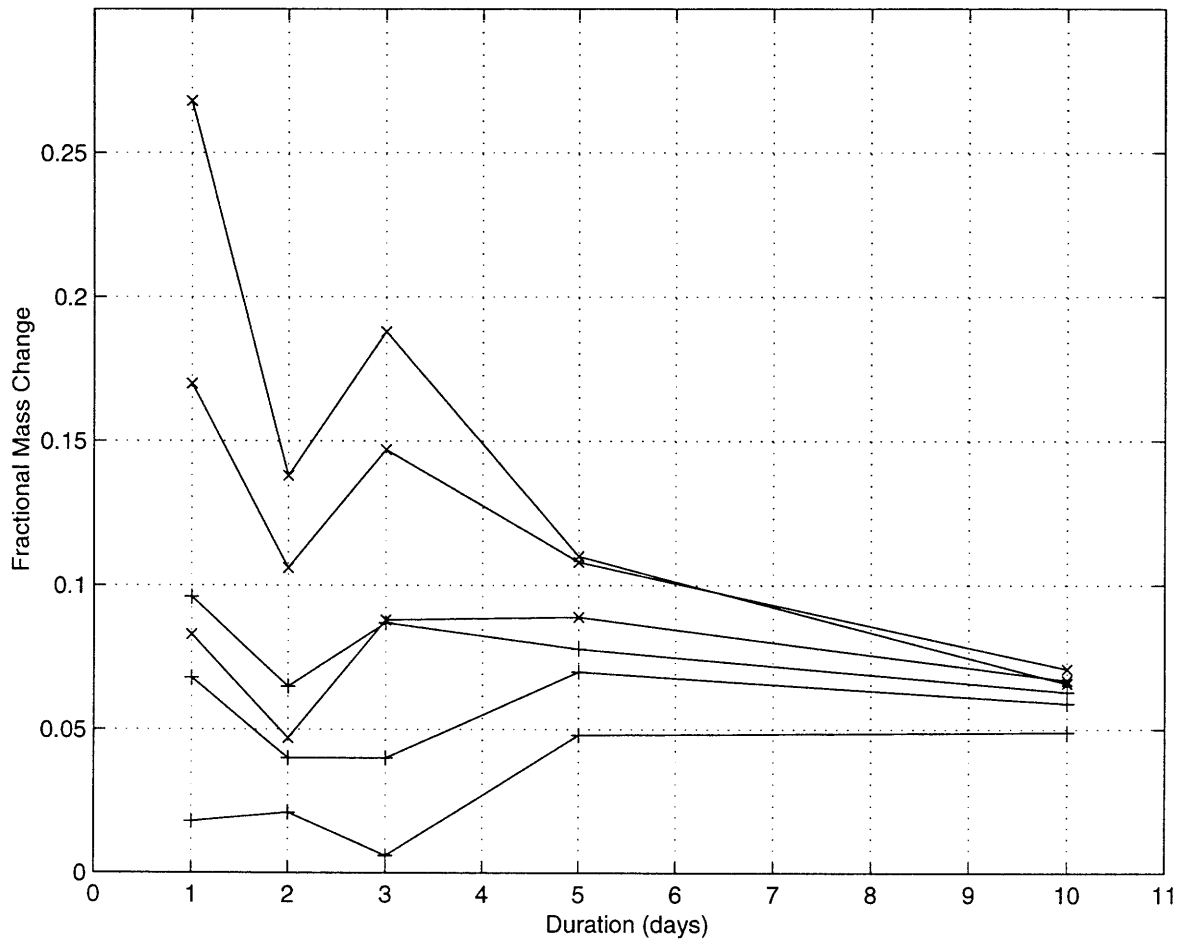


Figure 6-9: Transport across the $1.1 \times 10^{-8} m^{-1} s^{-1}$ contour over a single ten day period, computed by the LGR technique with RDF run durations varying from one to ten days. The six curves represent the three different values of $\delta q$ used in the preceding figures, and two values of $n_{lat}$; $n_{lat} = 2$ is denoted by pluses, $n_{lat} = 4$ by x's. The transport is represented as a fraction of the total mass enclosed by the contour.

# Chapter 7

# Shallow Water Results II: Transilient Matrices

In this chapter, transilient matrices are formed from the shallow water model flow. These are then used to construct a one-dimensional model of the evolution of an idealized "chemical tracer" in the full two-dimensional model. The tracer is acted on by a relaxation to a latitude-dependent basic state ("chemistry") as well as transport. The 1D model is constructed in a PV-based equivalent latitude coordinate, with the transport performed by the transilient matrix. The results of the 1D model are compared with the results of the 2D model, and used to determine the permeability of the vortex edge to tracers. The latter determination then allows a complete description of the Lagrangian circulation, including the relationship between parcel trajectories and PV contours, to be constructed.

## 7.1 Construction of the transilient matrix

The transilient matrix was constructed from the results of RDF calculations, using the straightforward "contour crossing" technique (SPW), as follows.

Each RDF run was performed for a period of $n$ days, where $n = 1 - 10$. Designate the first day as day 0 and the last as day $n$. On day 0 and day $n$, the PV field was first divided into bins spaced equally in *mass*. The equivalent latitude used for labeling the bins, however, was computed with respect to area. The locations of the bins changed very little from day to day, in either PV or equivalent latitude, as the model was near a steady state in the MLM coordinate.

Each particle in the RDF run was assigned a final and initial PV value by interpolating the model PV field to its position on days 0 and $n$. These PV values then determine the final and initial equivalent latitude bins as described above. Each parcel was given a mass value determined by the grid box area times the instantaneous model layer thickness at its location, both evaluated at the final time (the grid box area is a function of latitude since the RDF grid used was a constant latitude-longitude grid). The transilient matrix element $C_{jj'}$, where $j'$ denotes the initial bin and $j$ the final bin, was then computed as the total mass of all parcels in bin $j$ which had come from bin $j'$, divided by the total mass of all parcels in bin $j$.

By construction, the columns of the matrices so constructed did sum to 1, since all parcels in bin $j$ at the final time had come from somewhere. However, the rows were not constrained to sum to 1. This is because the parcels were irregularly distributed on day 0

(remember that the advection is done backwards starting from a regular grid) and there was no constraint which forced the amount of mass represented by the parcels in a given bin to be the same on days 0 and $n$. If the parcel density is high enough, experience has shown that mass conservation is only weakly violated by the RDF technique for flow regimes which are nondivergent or nearly so. In regions of substantial divergence, however, it is more strongly violated. The transilient matrices obtained from the described procedure were found to have rows with sums noticeably different from 1 in the subtropical regions, due to the relatively large divergence there.

Transilient matrices were computed in this manner for each $n$ day period of the longer 50 day period, where in this case $n = 5$ or 10. All of these were then averaged to produce a single matrix representative of the mixing occurring during a hypothetical average $n$-day period. (In the case of $n = 1$, 10 matrices were computed, for 1-day periods spaced at intervals of 5 days, and the average computed from these, as opposed to computing 50 matrices). An algorithm attributed by R. Stull to J. Petersen (Stull 1995) was then applied to this averaged matrix to force the matrix to conserve mass. This algorithm has the advantage of preserving elements which are zero. Since the subtropics exchanged zero material directly with the polar vortex during any $n$-day period, the large adjustments which the algorithm made to the subtropical regions of the matrix did not affect the polar vortex edge region. The latter region fortunately nearly conserved mass to start with, and so was little affected by the algorithm. It should be kept in mind that the matrix representation of the model's subtropical region may contain significant errors due to the algorithm's adjustments, but fortunately the polar vortex edge is of primary interest here.

The matrices for time steps of 1, 5, and 10 days are shown in figures 7-1, 7-2, and 7-3, respectively. Note in all three the almost complete absence of mixing in most of the southern hemisphere, as well as in the polar vortex edge region. The northern midlatitude surf zone is clearly visible as a region of strong mixing. The northern subtropical edge appears as a mixing barrier, but only a weak one; however the distortion of the results in this region by the mass conservation algorithm makes this quantitatively uncertain.

The distribution of eigenvalues of the 5-day transilient matrix is shown in figure 7-4. Note that there are is a substantial range of small values, indicating the existence of many slowly-decaying eigenvectors. The few small negative eigenvalues are due to very slight mass nonconservation by the transilient matrix (after adjustment by J. Petersen's algorithm).

## 7.2 1D Model

### 7.2.1 Model formulation

The 1D model consists of the tracer evolution equation

$$a(j,t+\Delta t) = \sum_{j'} C_{jj'}(\Delta t)a(j',t) + \frac{(a_e(j)-a(j,t))\Delta t}{\tau_a} + \frac{a(j,t)(h_e(j)-h(j))\Delta t}{h(j)\tau_h}$$

where $j$ represents the index of a given equivalent latitude bin, and $C_{jj'}$ is an element of the transilient matrix for time step $\Delta t$ (equal to $n$ days). This equation results from simply averaging the nonconservative terms in the full equation for the tracer as discussed below, and by representing all transport processes by the transilient matrix.

$a_e(j)$ is the average chemical equilibrium tracer value for bin $j$. It is important to realize that this is dependent on the configuration of the PV contours defining bin $j$, since the latter are far from being coincident with latitude circles and the actual chemical term in the full model is dependent on true, as opposed to equivalent, latitude. $a_e(j)$ must be computed from the full 2D model fields. This is done by computing the average (true) latitude of the material in bin $j$ over the 50-day period. The average is mass-weighted, and taken over space (that is, the entire region in longitude and latitude enclosed by the PV contours bounding bin $j$) as well as time. The same is true of $h_e(j)$, except that here the latitudinal profile of $h_e$ is of course taken into account, since it is not simply equal to the latitude as $a_e$ is.

### 7.2.2 Comparisons of the 1D and 2D models

The 1D model is compared to the 2D model as follows. The mean profile of tracer $a$ in equivalent latitude space is computed on days 1 and 50 from the shallow-water data. The profile from day 1 is input to the 1D model, which is then run for a number of time steps equivalent to 50 days, where the time step is set as appropriate for the transilient matrix. The result is then compared to the profile from the 2D model on day 50. The results of this comparison are shown in figure 7-5.

The figure shows that the 1D model and 2D models are both near steady state, and that apart from some drift of the 1D model away from the 2D model, particularly in the subtropics, the two steady states are close to each other. Accepting for the moment the proposition that these are nontrivial steady states, the comparison suggests that the 1D model is doing a reasonably good job of capturing the 2D model's behavior. For reasons discussed above, the disagreement in the subtropics is not surprising. The results would seem to offer validation for the transilient matrix computation, justifying further effort spent on analyzing the details of the matrix' properties. However, the figure also shows the result one obtains by assuming a steady state with no mixing at all, $a = a_e$ (so the effect of the height source is neglected as well; this term is in fact negligible compared to the others). This reproduces the 2D model's behavior about as well as the full 1D model does. The somewhat disturbing suggestion is that transport has no role in determining the model's steady state tracer profile vs. equivalent latitude.

Figure 7-6 shows an initial value experiment carried out with the 1D model. The initial profile is simply taken to be $a = \phi_e$, and neither the chemical nor height relaxation terms are included, so the model consists solely of repeated multiplication by the transilient matrix. It is seen that the profile reaches a state similar to the 2D model's steady state within only a few time steps. There can be no steady state other than $a = const$ for nontrivial $\mathbf{C}$ and no nonconservative terms, so this state continues to be mixed towards homogeneity, but nonetheless the character of $\mathbf{C}$ is seen to at least qualitatively represent the 2D model's behavior well.

The difficulty, apparently, is that both the mixing and the chemistry "want" to produce approximately the same state, so the competition between the two processes is weak, making the 1D experiments a weak test of the accuracy of the transilient matrix. More formally, the chemical equilibrium state $a_e(j)$ projects predominantly on the slowly decaying eigenvectors of $\mathbf{C}$. The structures of the eigenvectors are not shown, as it requires many of them to construct $a_e(j)$ and so the decomposition is not particularly enlightening — remember that there are many with small eigenvalues. Physically, this means that where the mixing is strong, PV and latitude are almost entirely uncorrelated so that $a_e(j)$ has no gradient with

respect to equivalent latitude, and the mixing has no effect. Where PV and latitude are strongly correlated, as in the vortex edge region, the mixing is very weak. This difficulty arises because $a_e(j)$ is sensitive to the structure of the PV field, and thus is influenced by the same mixing process which the transilient matrix represents. *For chemistry dependent on latitude*, the roles of transport and chemistry are not cleanly separated in this formulation.

### 7.2.3 Sensitivity to time step

As a consequence, the 1D model is insensitive to parameters which we might in general expect to play an important role. In particular, the time step has been varied from 1 day to 10 days, with virtually no effect on the results. In general one expects that in order for the model to behave self consistently, the time step should be of the same order as the large eddy turnover time. One does not expect the self-consistent behavior to extend over an order of magnitude in this parameter. The actual matrices are not consistent, in that for $\Delta t = 1$ day, for example

$$\mathcal{C}(10\Delta t) \neq [\mathcal{C}(\Delta t)]^{10}$$

This is demonstrated by figure 7-7. Because of the complexity inherent in comparing two $100 \times 100$ matrices, what is shown here are the diagonal elements of the ten-day matrix, the five-day matrix raised to the second power, and the one-day matrix raised to the tenth power. Roughly, the smaller the diagonal element at a given equivalent latitude, the stronger the mixing across the corresponding PV contour, and vice versa. Interestingly, in well-mixed regions the diagonal elements appear to be independent of time step, but in regions of weak mixing the dependence is fairly strong between one and five days. However, between five and ten days the dependence is much weaker, suggesting that the optimal time step is in this vicinity, which was (presumably not by coincidence) also the optimal duration for the LGR calculations in the preceding chapter.

Despite their differences, all the matrices mix relatively strongly in the surf zone and relatively weakly in the vortex edge region. As a result, the differences made in the 1D model's behavior by the different time step choices are small for the comparisons to the 2D model, even between one and five days.

This prevents investigation of certain issues of interest. The results shown in figures 7-5 and 7-6 do indicate that the 5-day transilient matrix represents the model's transport in a reasonably correct way. In fact, one presumes that if it were very wrong, the 1D model would no longer be so insensitive to it. For example, if the matrix indicated overly strong mixing in the polar vortex edge region, the profiles with and without mixing would not look so similar. However, the results do prove insensitive to variations in time step. Using the 1-day or 10-day transilient matrix (and changing the time step accordingly) makes virtually no difference to the results shown in figure 7-5. This is because the structure is the same regardless of time step; there is strong mixing in the surf zone, where $a_e$ is nearly independent of equivalent latitude, and weak mixing in the vortex edge region.

### 7.2.4 Potential for evaluating diffusion model

One initial motivation for undertaking the research reported in this chapter was to study the details of the transilient matrix' structure. This could provide quantitative means for evaluating various forms of simpler parameterizations. For example, although the matrix obtained using, say, a 5-day time step is not tridiagonal, Flierl (personal communication) suggests that one could attempt to find the tridiagonal matrix which is closest to it, in some

meaningful sense, by solving an inverse problem. The problem could be formulated as

$$\mathbf{T}^n + \mathbf{E} = \mathbf{C}$$

where $\mathbf{T}$ is a tridiagonal matrix, $\mathbf{E}$ is an error matrix to be minimized using some appropriate norm, and $n$ is a scalar exponent (note that a tridiagonal matrix raised to a power greater than unity is in general no longer tridiagonal). In the present context, a tridiagonal matrix would be equivalent to a profile of eddy diffusivity as a function of equivalent latitude, which we might label $K_{\phi_e\phi_e}(\phi_e)$. The smallness of the minimum attainable value of the chosen norm for $\mathbf{E}$ would be an indication of the appropriateness of the diffusion model for the flow under consideration.

The difficulty with carrying out this procedure is that the results will be dependent on detailed aspects of the transilient matrix structure. For example, assume we are given a transilient matrix which mixes isotropically; that is, all rows and columns have identical profiles relative to the diagonal, except near the boundaries where isotropy cannot be maintained. One naturally would try to fit this by a tridiagonal matrix representing a constant diffusivity, $K_{\phi_e\phi_e}(\phi_e)$. Raising this matrix to a higher power would result in a matrix whose row elements (far from the boundaries) would have Gaussian profiles centered on the diagonal. The minimum attainable value of the norm of $\mathbf{E}$ for an observed transilient matrix would depend on the departures of its rows from Gaussian-ness.

In the present results, details at this level of subtlety change as the time step is varied. In order to determine how well a diffusivity profile can match the 2D flow's mixing behavior, one needs an ability to show that a particular choice of time step (or a range of time steps smaller than an order of magnitude) is more correct than others. Because of the insensitivity to this parameter discussed above, this is impossible by the present method. Consistency checks such as that shown in figure 7-7 do provide some guidance for picking the time step as well as increased confidence in the overall results, but more detailed comparisons (not shown) including all the off-diagonal elements suggest that features more subtle than the gross permeability of the vortex edge are also less robust.

### 7.2.5 Direction for possible future work

One might be able to avoid this problem by performing a different experiment, in which the "chemistry" term is made proportional to equivalent latitude rather than true latitude. One might attempt to justify such an experiment on physical grounds by arguing that for many species, the apparent source seen by an isentropic layer is due to cross-isentropic advection rather than chemistry *per se*. Since air parcels in the vortex region follow orbits whose projections on isentropic surfaces are closer to PV contours than to latitude circles, one could argue that the descent averaged over PV contours is more relevant than the zonally averaged descent.

With this choice, one may expect that the 1D model will be more sensitive to the time step, enabling better validation of a particular choice for this parameter, hence more justification for analyzing the transilient matrix' structure in detail. This work requires some further development to the (2D) shallow water model to implement a tracer nonconservative term dependent on equivalent latitude, and is left to future investigations.

### 7.2.6 Equivalent latitude 2D models

The chemistry-mixing degeneracy described above has played, in part, the role of an annoying technical difficulty. The reason for it is interesting, however, and beyond this it has implications for possible future stratospheric research projects. While 3D chemical transport models are becoming more practical for stratospheric research, and may eventually supplant their 2D (latitude-height) counterparts entirely, at present this does not seem imminent. As long as 2D models are destined to be used for the forseeable future, one may imagine expending some effort to improve them. An obvious weakness of current 2D models is their zonally averaged formulation. An equivalent latitude formulation, based on contours of PV (or possibly a chemical tracer) would have certain advantages, which hopefully are clear from the discussion in the beginning of chapter 4. The results of this chapter, however, point out a difficulty with the equivalent latitude formulation which is not present in the zonally averaged formulation. That is that incoming solar radiation, which is crucially important to stratospheric chemistry, is expressed quite naturally as a function of latitude and height. An equivalent latitude formulation would require some way of representing the average solar radiation falling on a zonally asymmetric contour, without computing it explicitly by first performing a 3D radiative calculation and then averaging. The intrinsic nonlinearity of radiative transfer would further complicate such an endeavor.

Relatedly, in the surf zone, where PV contours are extremely convoluted and often multiply connected, the equivalent latitude representation loses much of its intuitive physical meaning. The results here demonstrate the consequences of this explicitly. One could imagine using a coordinate which switches smoothly between equivalent latitude in the vortex region and true latitude at lower latitudes, such as the one devised by Norton (1994). This, on the other hand, would introduce a new level of arbitrariness and complexity, both to the model formulation itself and to any comparison of model and observations.

## 7.3 The Lagrangian circulation

### 7.3.1 Permeability of the vortex edge

Despite the limited justification for detailed analysis of the structure of the transilient matrix, the matrix can still be used as a piece of independent information on the permeability of the vortex edge. As such it is complementary to the results presented in chapter 6. In particular, the matrix offers direct information about tracer fluxes, as opposed to net mass fluxes. The net flux of tracer across a boundary may be expected to be less sensitive to the detailed structure of the matrix than the inverse problem discussed above, because the flux depends on sums over many matrix elements (see equation 3.11).

Figures 7-8 and 7-9 show the results of initial value experiments using the 1D model with mixing alone. The initial conditions are a "step" tracer distribution, with tracer equal to 1 everywhere poleward of some equivalent latitude, and zero elsewhere. It may be seen that material leaks across the polar vortex edge only very slowly. The precise rate depends on the contour chosen, but values are not more than a few percent per month for contours in the main vortex edge region. The 5-day transilient matrix was used to obtain these results, but figure 7-7 suggests that using the 10-day matrix, for example, would not change the results greatly.

When combined with the results of chapter 6, this information allows us to construct a comprehensive picture of the Lagrangian circulation in the model's extratropics.

### 7.3.2 The characteristics of parcel trajectories

It should be kept in mind that MLM diagnostics are, as suggested by the word *modified*, not truly Lagrangian. We made the connection in chapter 6 between the net effects, integrated around PV contours, of the thermal and hyperdiffusive PV source terms. However, it is in principle possible that there could be a substantial large-scale mass flux *through* the vortex even if these integrals vanished, as long as there were no net mass convergence *into* the vortex. Parcels could enter the vortex through regions of positive PV forcing and leave through regions of negative forcing (see figure 6-2). This would imply large tracer fluxes across the vortex edge, which, however, would be invisible to MLM diagnostics. Note that this scenario would imply large secular PV changes following the parcels making the cross-vortex journey.

The alternative picture is one in which parcels in the main vortex edge region simply follow orbits which are approximately close to PV contours, but oscillate slightly about the latter as they pass through the regions of positive and negative PV forcing. This picture is more consistent with basic notions of quasi-balanced, quasi-conservative flow, but the fully Lagrangian contour crossing results presented in this chapter confirm it directly, by showing that the vortex exchanges very little air with its exterior. In this picture, the contours in the entrainment zone are continually being pushed equatorward by the substantial thermal PV forcing there. However, these contours are then drawn out into filaments in wave breaking events. These filaments are then dissipated, and the net effect is that the contours recede poleward again. In the main vortex edge, where the thermal forcing and hyperdiffusion both nearly vanish, parcels simply orbit repeatedly as described above, providing a robust barrier between the vortex interior and the surf zone. The surf zone is a region of rapid chaotic mixing, but PV contrasts are small there.

The contour crossing calculations have thus proved useful. They have added a fully Lagrangian perspective, and made clear the relationship between tracer and mass fluxes, confirming the view of the Lagrangian circulation which was intuitively expected. It is important to keep in mind that the contour crossing approach is only sure to be useful when analyzing numerical model output, as opposed to observational analyses, since it is quite sensitive to noise (SPW). This highlights an advantage of using numerical models as proxies to understand transport in the real atmosphere. By admitting the application of diagnostics too delicate to be used on observations, model data sets allow a wider range of diagnostics to be applied, yielding a more comprehensive understanding of the model flow's transport properties. To the extent that the model is believed to represent reality, this offers an enhanced view into the real flow's behavior, of a sort which would otherwise be unavailable.

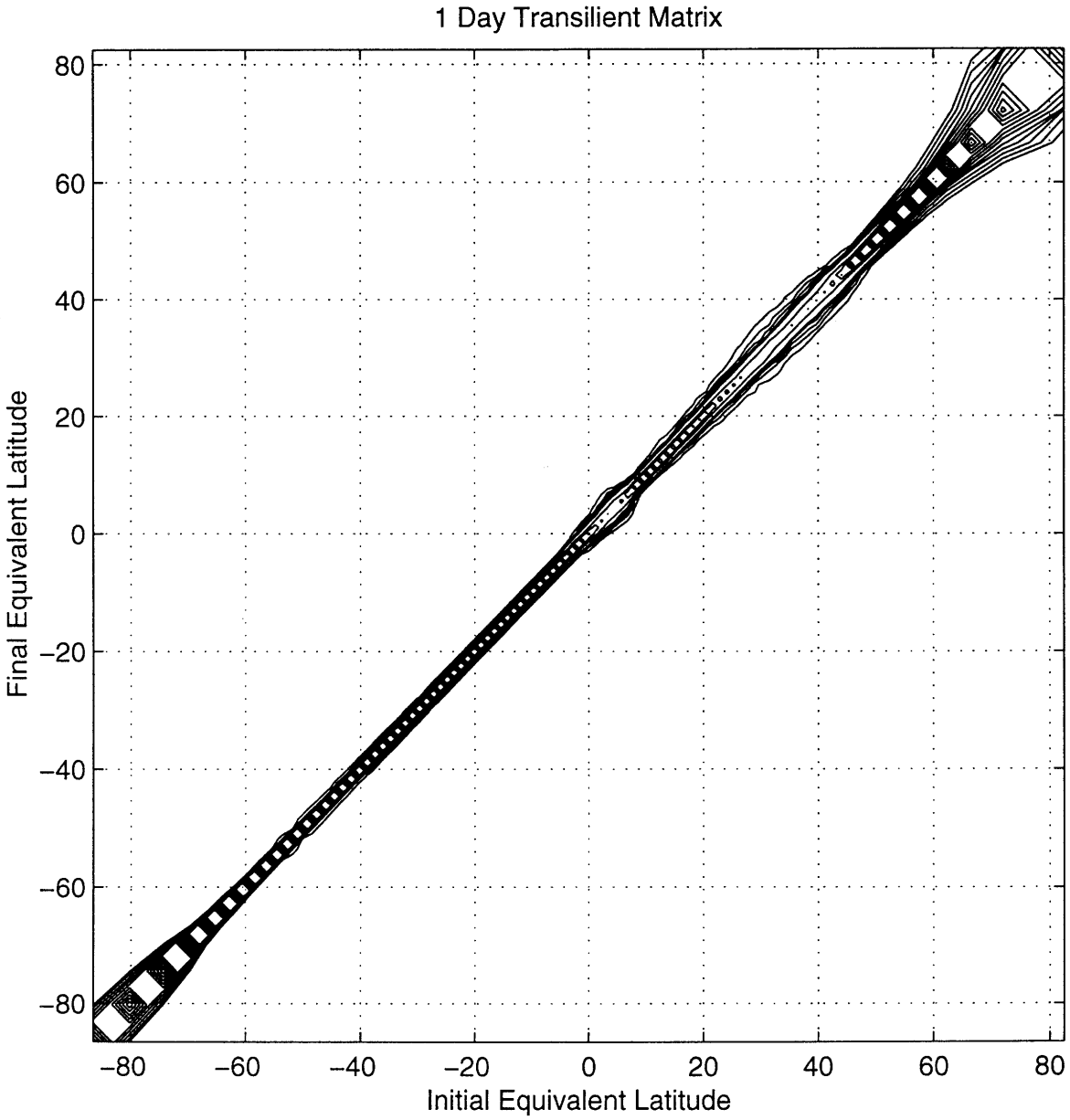


Figure 7-1: Contour plot of 1-day transilient matrix, constructed from contour crossing calculations using shallow water model data. Contour interval is variable: it is 0.02 between the values of $0.0 - 0.1$, and 0.1 between the values of $0.1 - 1$.

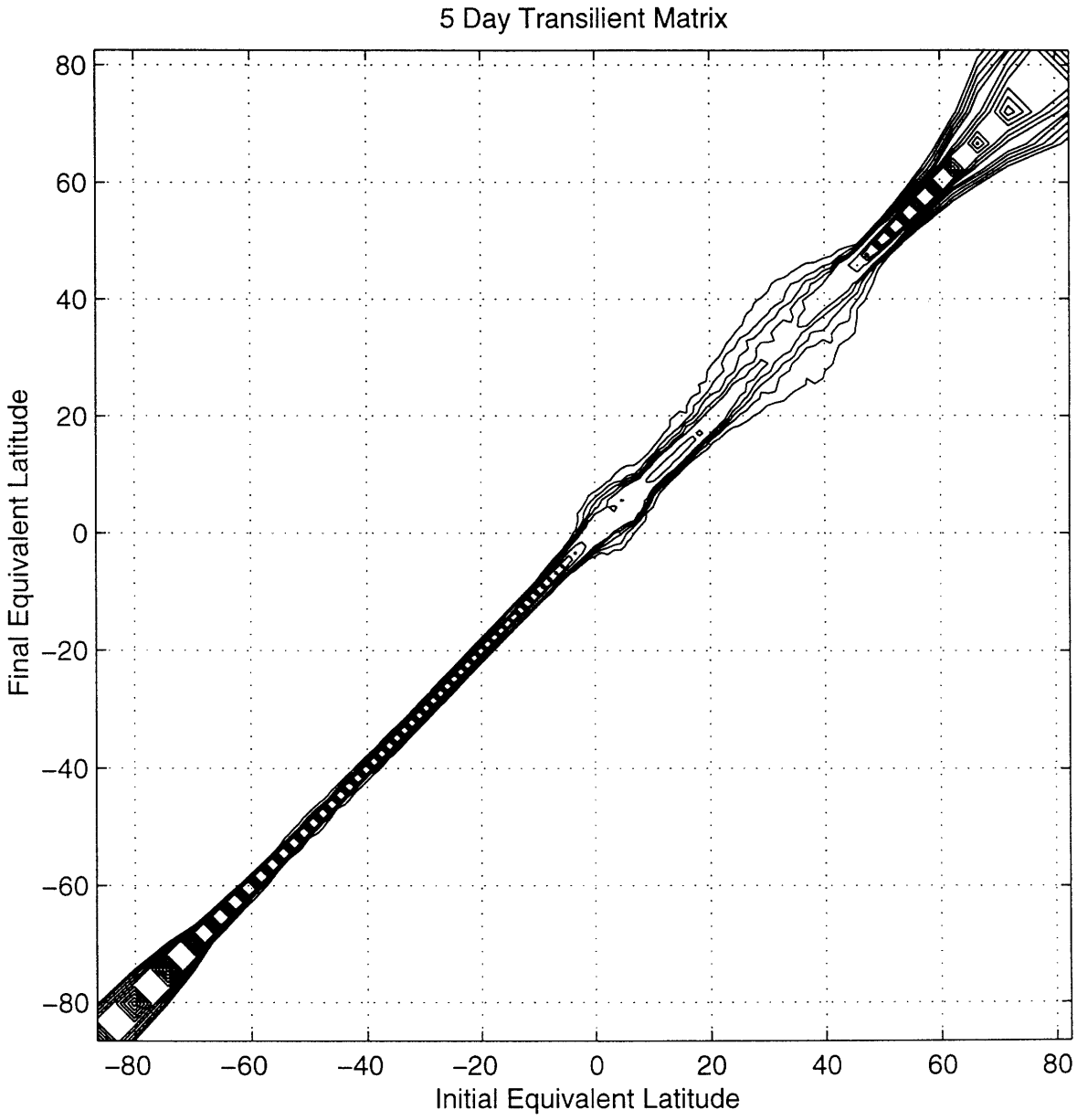


Figure 7-2: 5-day transilient matrix, constructed from contour crossing calculations using shallow water model data. Contour interval as in figure 7-1.

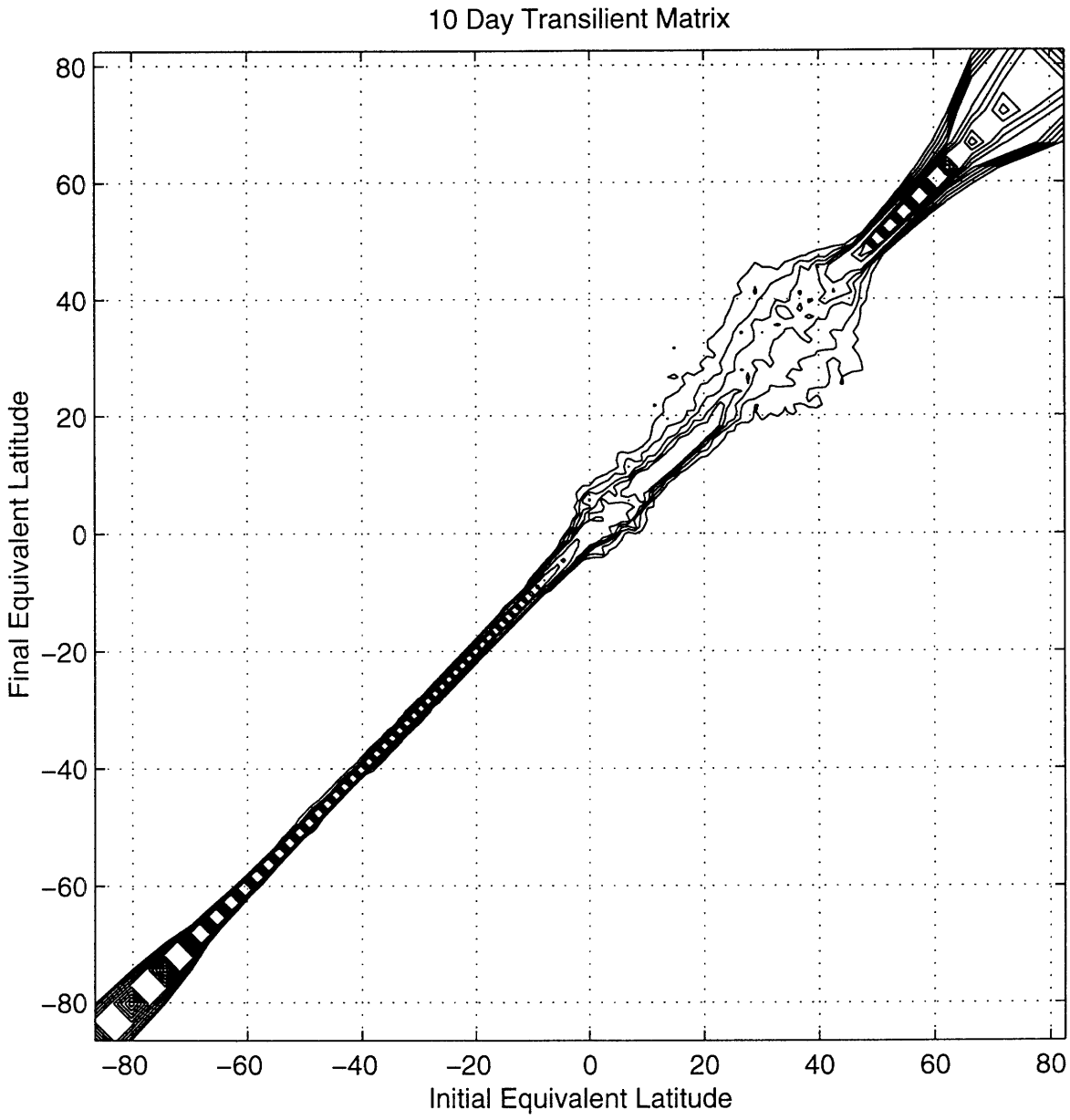


Figure 7-3: 10-day transilient matrix, constructed from contour crossing calculations using shallow water model data. Contour interval as in figure 7-1.

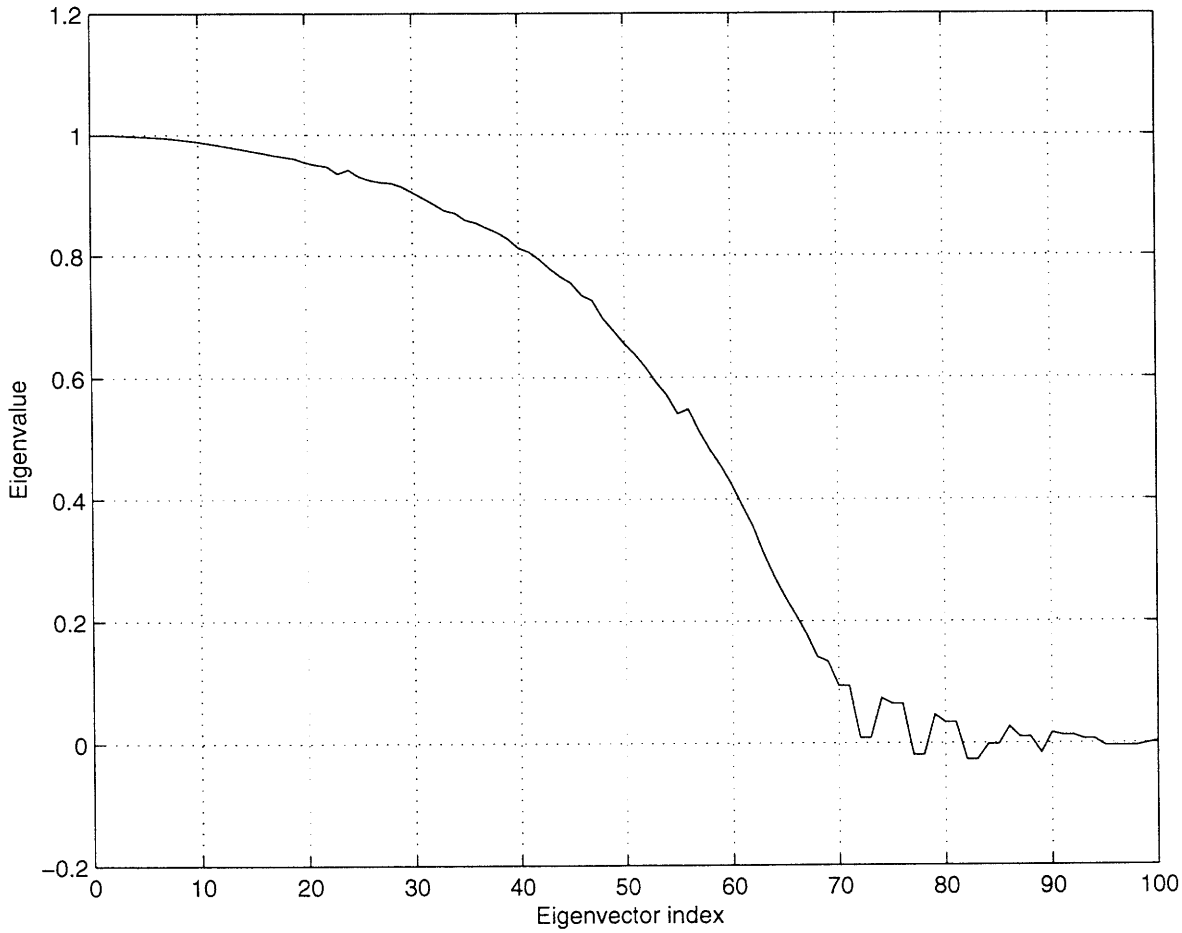


Figure 7-4: Eigenvalues of 5-day transilient matrix ($y$ axis), ordered from highest to lowest (along the $x$ axis).

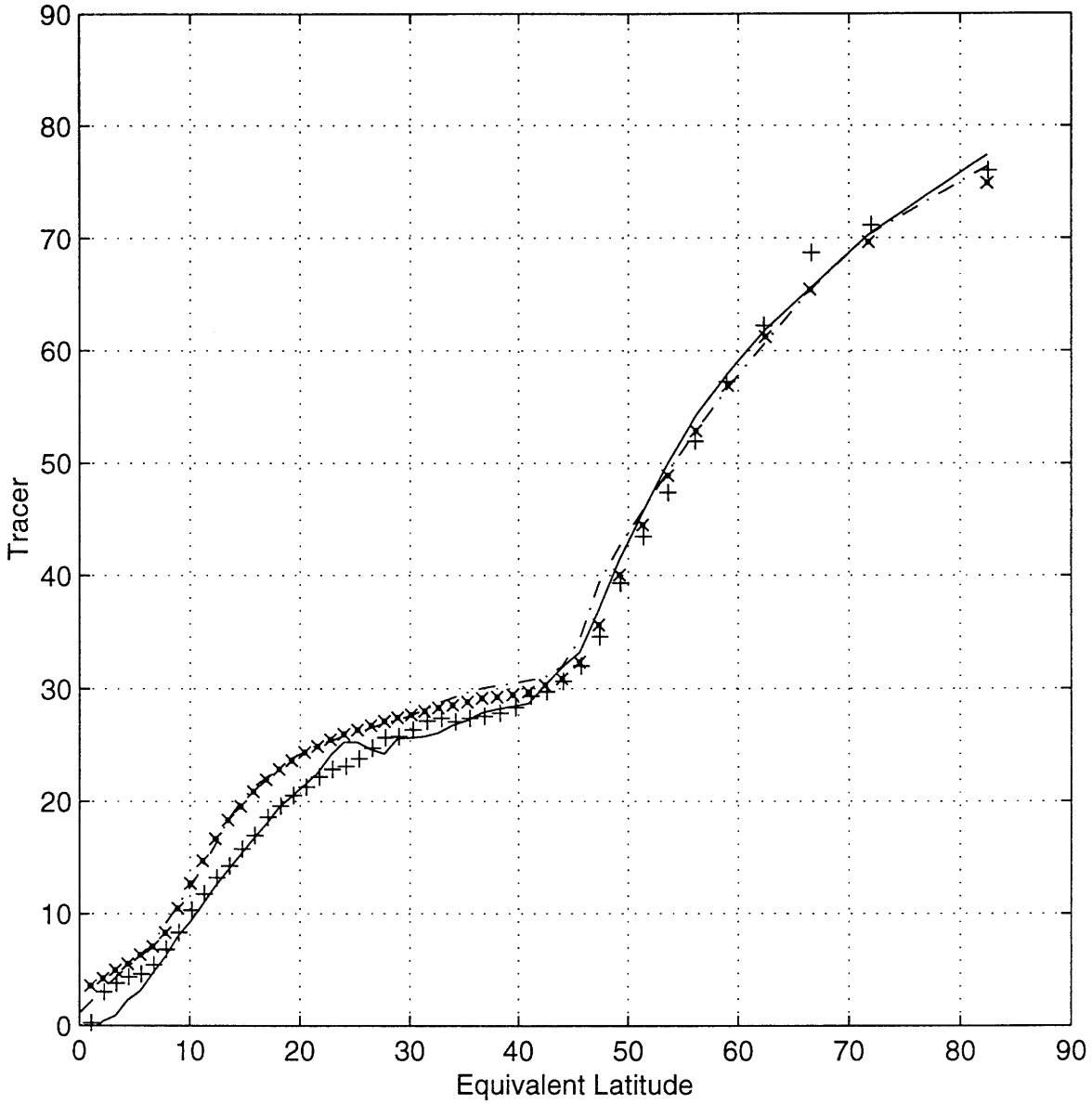


Figure 7-5: Comparison of the 1D and 2D models. Solid curve is the initial tracer profile as a function of latitude, taken from the 2D model. Pluses represent the 2D model's final tracer profile. The x's are the final tracer profile from the 1D model, while the dot-dash curve is the result from the 1D model when mixing is turned off ("chemical equilibrium").

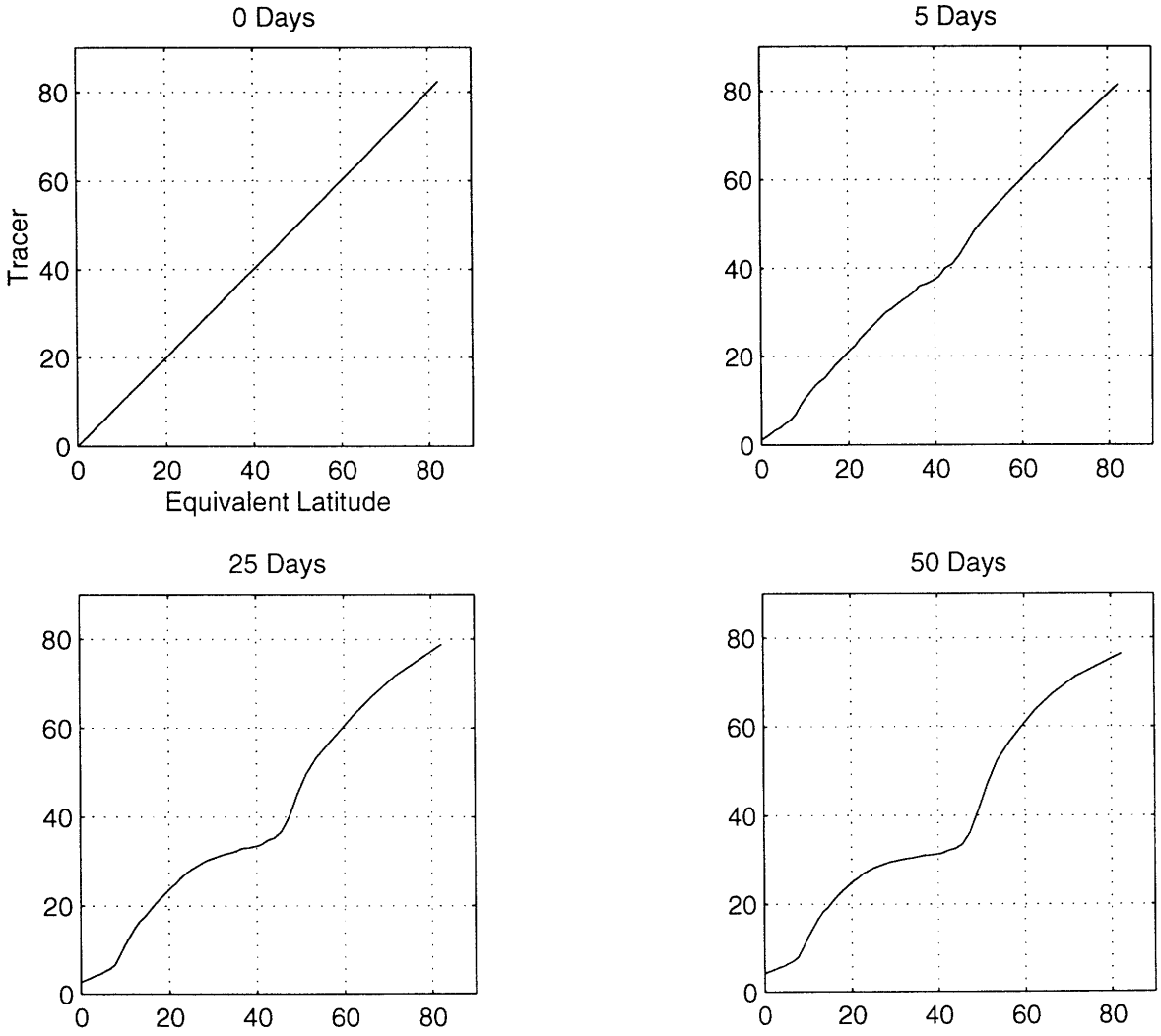


Figure 7-6: Initial value calculation using the 1-D model, with the chemistry term turned off. The initial condition has the tracer equal to the equivalent latitude (top left panel). Results are shown after 5, 25, and 50 days, corresponding to 1, 5, and 10 applications of the 5-day transilient matrix.

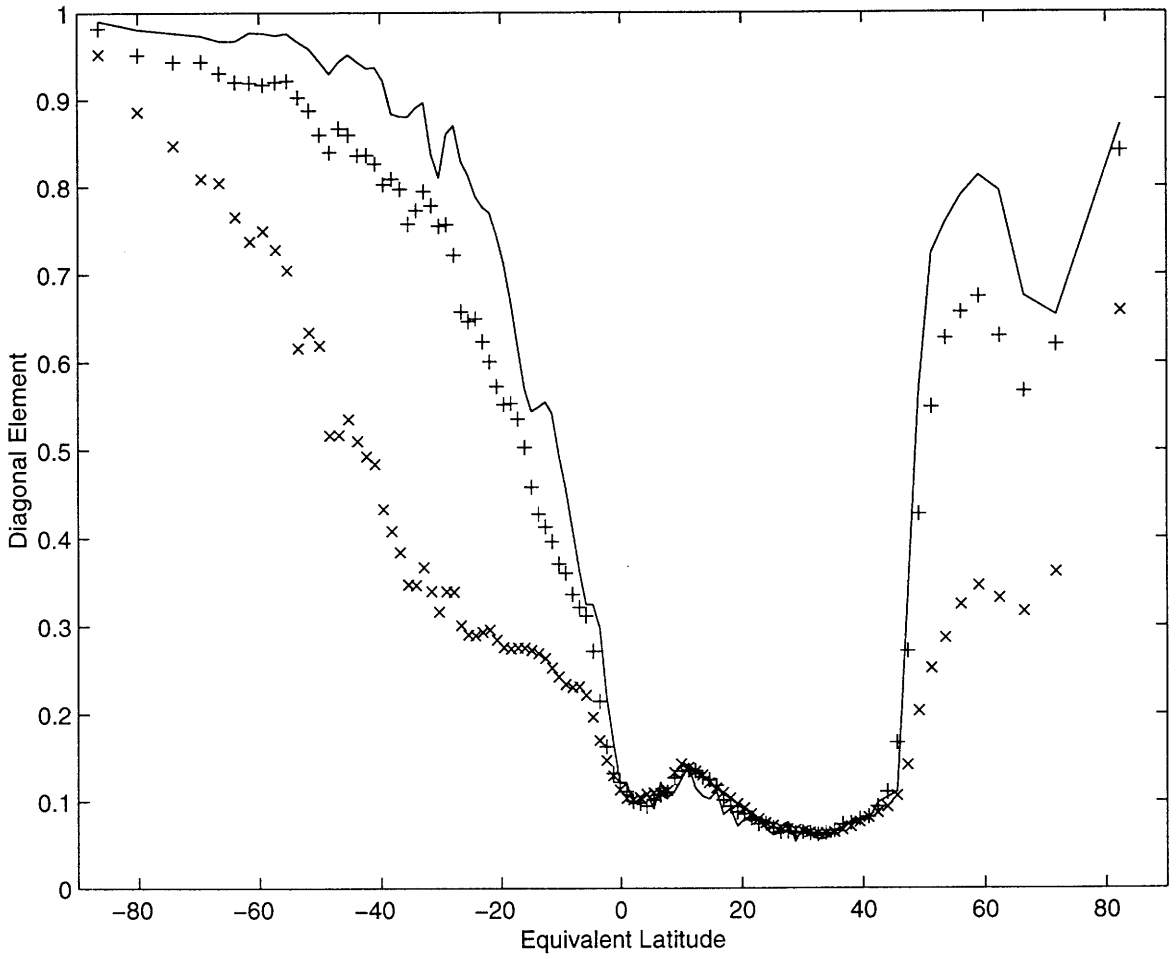


Figure 7-7: Comparison of the diagonal elements of the matrices corresponding to the 10-day transilient matrix (solid curve), 5-day transilient matrix squared (pluses), and 1-day transilient matrix raised to the tenth power (x's).

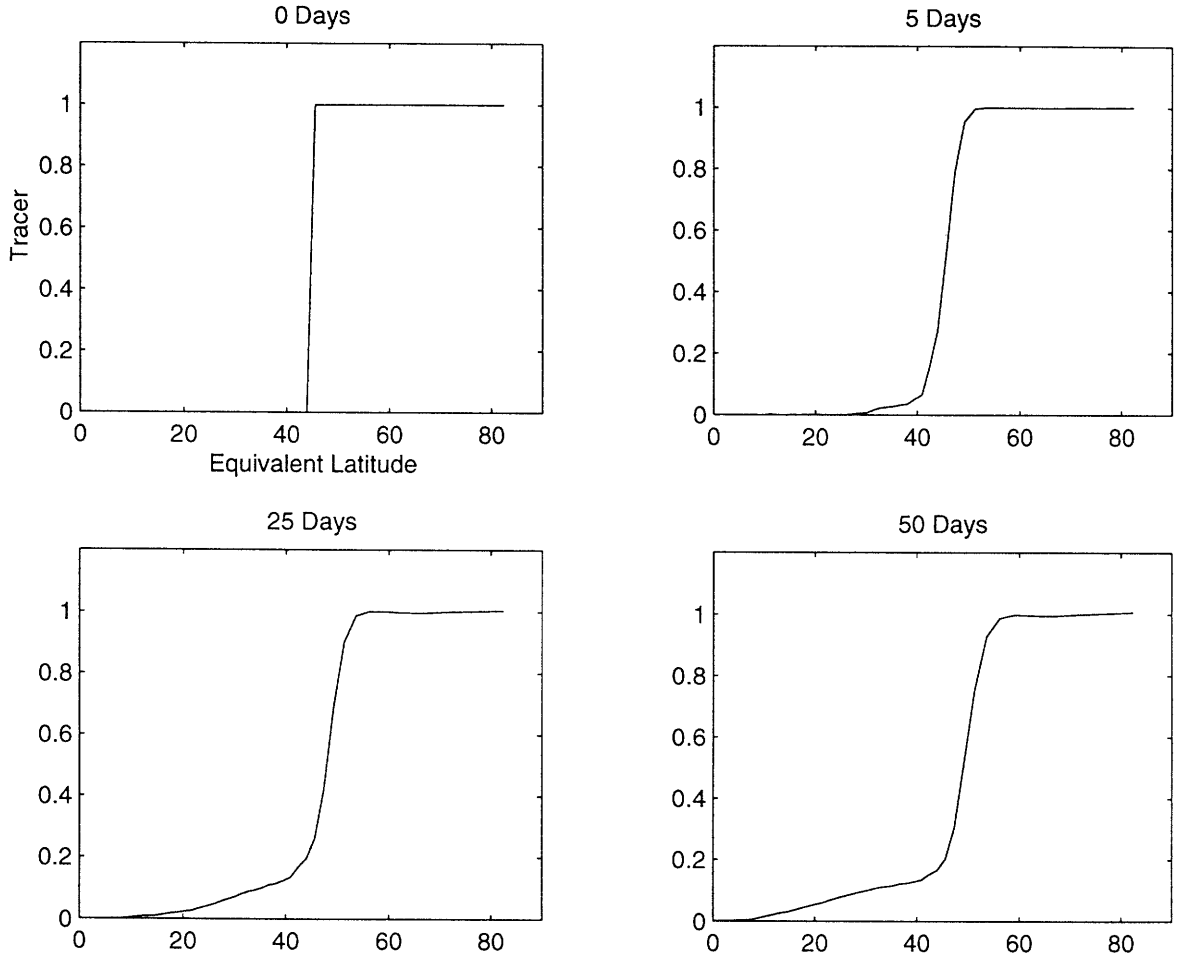


Figure 7-8: Initial value experiment with the 1-D model using the 5-day transilient matrix and no chemistry term. The initial condition is a step function in equivalent latitude (upper left panel). The edge of the initial step is chosen just outside the vortex edge; note the rapid mixing of the most equatorward part of the tracer distribution.

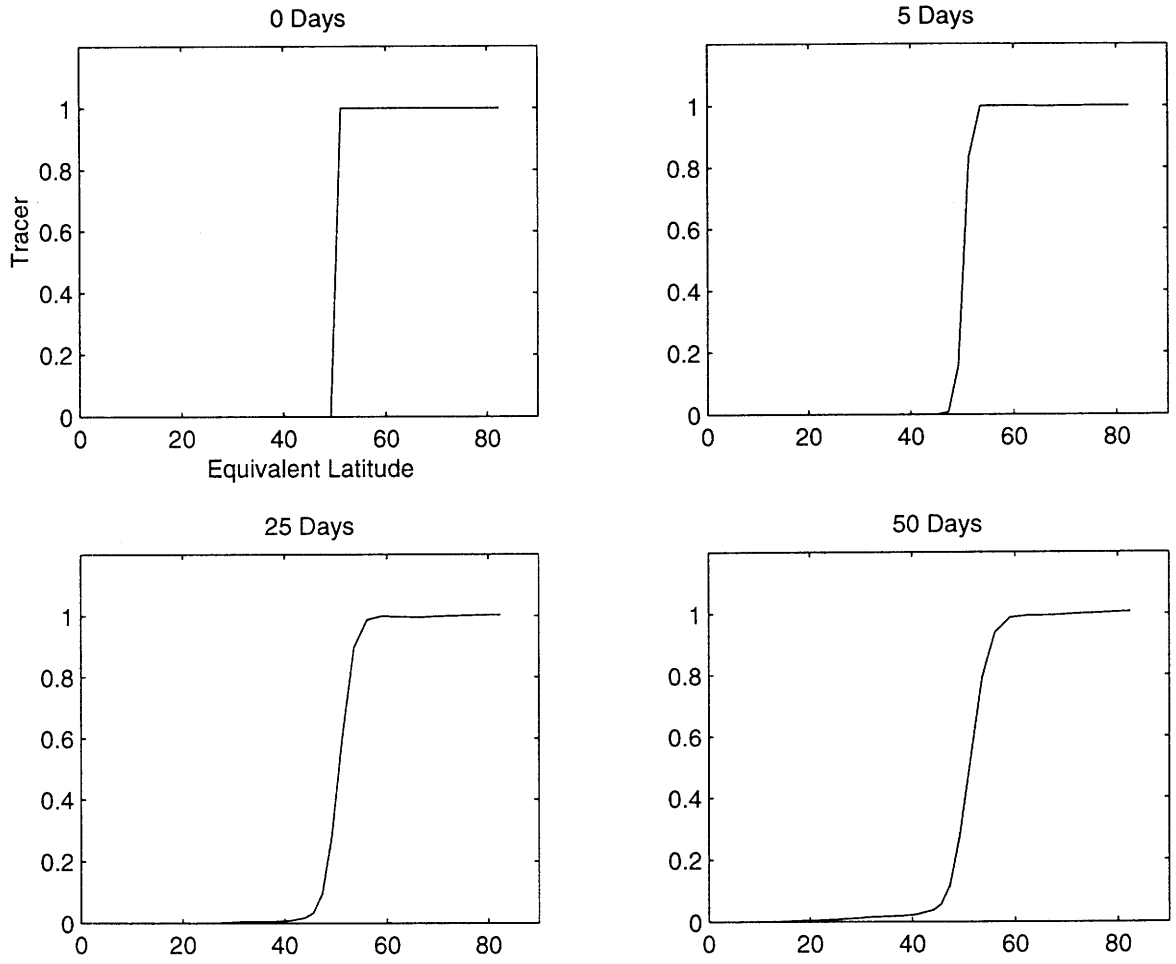


Figure 7-9: As in figure 7-8, but the initial tracer distribution now does not overlap at all with the surf zone; note the much slower leakage of material to low latitudes.

# Chapter 8

# Summary, Discussion, and Final Remarks

This chapter first summarizes the most significant results of the previous ones. Following this, I add a few final remarks to put the results in a broader context.

## 8.1 Summary and discussion

Chapter 2 reviewed some basic facts about the stratosphere.

Chapter 3 discussed the general problem of choosing an appropriate transport model to represent mixing by a given 2D flow. The limitations of the eddy diffusion model were pointed out. Alternate descriptions were introduced and their advantages and limitations explored. It was shown that the transilient matrix is the most general representation of mixing which is both local in time and can be formulated purely as an operator on the mean tracer profile. It was also pointed out that the flux due to mixing by a transilient matrix is always downgradient if the tracer profile is monotonic. Downgradient transport is thus a much more general phenomenon than is diffusion. Diffusion is a special case, corresponding to a tridiagonal transilient matrix, or a matrix which can be expressed as a tridiagonal matrix raised to a power.

The "relaxation to the average" (RA) model was derived two different ways, first as an approximation to the transilient matrix, and then heuristically, from an idealized model of mixing by eddies which quickly traverse the domain in $y$ while being narrow and widely spaced in $x$. This model bears some resemblance to vertical transport by deep moist tropical convection[1], as well as horizontal transport by breaking stratospheric planetary waves. The RA model is the simplest model of nonlocal mixing by large eddies which retains a finite mixing time scale. It is a generalization of the familiar "mixed layer" model, which assumes infinitely fast mixing. It was shown that the RA model is not equivalent to diffusion, either formally or practically.

The implied conclusion is that diffusion is a suboptimal representation of stratospheric mixing. This is not a new conclusion, but the results which here lead up to it hopefully make the argument behind it somewhat more coherent. However, the extent to which diffusion

[1]This assumes substantial detrainment of air at all levels in cloud updrafts. In fact, deep convection often has a "bottom-up/top-down" character, in which the opposite extremes of the domain exchange material without greatly influencing the interior.

can in practice be tuned to give good results is still not entirely clear, nor are the detailed pros and cons of using any particular alternate representation in a 2D (latitude-height) chemical transport model.

Models such as the "pipe model" of Wofsy et al. (1997) do, however, seem to be an alternative worth pursuing. In such models, the atmosphere is divided into regions which are then treated as internally well-mixed in the horizontal. Transport between the regions is accomplished by mean advection, the horizontal component of which is unidirectional at any given vertical level, and a bidirectional mixing rate. These two processes determine all property fluxes between the regions. One application of such a model is to treat the mixing rate as an unknown to be determined by fixing all other factors in the model. Otherwise, it must be specified, and one would ideally like to diagnose it from observations. This makes a connection between chapter 3 and chapters 4 and 6, which addressed the problem of determining the transports across internal surfaces from given flow fields, the latter chapter focusing on the polar vortex edge. For example, a pipe model type approach has been used to infer ozone depletion in the Arctic vortex by Knudsen et al. (1997), who used the contour crossing method to obtain the inward (i.e. poleward) flux of ozone across the vortex edge. Although their contour crossing calculations of the inward and outward transports of *mass* across the vortex edge appeared to be substantially contaminated by noise, for their purpose the results were relatively insensitive to this because of the small ozone contrast across the vortex edge region (B. Knudsen, personal communication). One can imagine applications which would be more sensitive to the mixing rate, justifying a greater expenditure of effort to diagnose it accurately.

Chapter 4 reviewed various techniques which have been used to diagnose the transport across moving internal boundaries, usually tracer isosurfaces or contours. The most direct approaches tend to be very sensitive to noise in the given tracer fields, so that they are of questionable utility for many applications involving observational data sets. MLM diagnostics are theoretically exact, but using them to obtain quantitative transport rates requires information which is unavailable from observations at present and for the forseeable future. Both the direct approaches and MLM diagnostics can be very useful for diagnosing numerical model output, as was done in chapters 6 and 7. Indirect advection-based approaches, such as CACG and LGR, are less sensitive to noise than the direct approaches, and do not require observationally unavailable information as MLM does, but have the disadvantage that they are *ad hoc* rather than being rigorously derivable from physical laws. The technique of lobe dynamics can be used to compute transport across manifolds, rather than tracer isosurfaces. This technique has a rigorous mathematical basis, but has only very recently begun to be applied to aperiodic flows.

Chapter 5 described the shallow water model, the boundary conditions, forcings, and other parameters used for the runs to be diagnosed, and techniques used to process its output. The model represents a single layer of the stratosphere, subject to thermal relaxation and topographic forcing. Purely 3D effects, such as descent (and therefore diabatic cooling) driven by wave dissipation at higher levels, are of course absent.

Chapter 6 investigated the balances maintaining the shallow-water polar vortex. The information obtained in this investigation was used to test the LGR method, and indirectly, the CACG method as well, of computing the rate of material transport across the outer vortex edge due to filamentation.

It was proved that in an inviscid shallow-water atmosphere with an arbitrary mass source, if the mass enclosed by PV contours is steady in time, then the integral of the mass source over the area enclosed by the contours must vanish. This has previously been proved

for the special case of steady flow fields, but the present result removes this restriction. This is a significant generalization, since real atmospheric flow fields are never steady. It must be pointed out that this result does not clearly generalize to three dimensions in the presence of vertical gradients of absolute vorticity, though it may still hold approximately in some situations[2]. Whether an exact, useful three-dimensional generalization exists is a subject of ongoing investigation.

The MLM mass budget was diagnosed directly from the model output, for PV contours in and near the vortex edge region. Since the model contains a hyperdiffusion term to dissipate small-scale enstrophy, it is not inviscid. The presence of the hyperdiffusion term allows departures from the MLM radiative equilibrium which would be required in an inviscid steady state. The hyperdiffusion term was found to be small (though not zero) in the main vortex edge region where the contours are unfilamented. The main vortex is thus close to MLM radiative equilibrium. The hyperdiffusion term becomes large in the outermost vortex edge region where filamentation is common, balancing a significant net thermal forcing of PV in this region. The vortex edge can thus be divided into a nearly inviscid main region and an effectively viscous outermost region, the transition between which is marked by a thin entrainment zone. The nature of the balances in these regions is quite insensitive to a factor of 50 change in the model's hyperdiffusion coefficient.

The LGR technique was used to estimate the rate of mass removal from the vortex edge due to filamentation. In the entrainment zone (and sometimes in the outermost vortex edge and surf zone) agrees well with the hyperdiffusion term. The technique is thus able to reliably estimate the extent of the departure from radiative equilibrium, a basic and physically meaningful quantity. The technique also agrees well with the CACG technique. The results here provide direct support for these two techniques, which had previously been defensible only by physical plausibility arguments. More importantly, they support the view that the balance between the two nonconservative processes in the model (thermal forcing and small-scale dissipation) is governed directly by the large-scale dynamics, which are themselves quasi-inviscid.

The LGR technique is unable to reliably diagnose the small departure from radiative equilibrium which occurs in the main, unfilamented vortex edge region. This quantity is, however, something of a model artifact, presumably being directly affected by the hyperdiffusion. What process determines the transport, surely very small, across the unfilamented portion of the real vortex edge is not clear at the present time.

In chapter 7, the contour crossing technique was used to construct the transilient matrix for the shallow-water model flow in its statistical equilibrium state. This matrix was used as the basis for a 1D chemical transport model, meant to mimic the 2D model's behavior. A passive tracer, acted on by a highly idealized, latitude-dependent "chemistry" was added to the 2D model, and the 1D model was used to reproduce the evolution (or steady state) of this tracer. Because the 1D model was formulated in an equivalent latitude coordinate based on PV, the latitude-dependent chemistry term was averaged over the PV contours in the 2D model to provide the proper forcing for the 1D model. This led to a somewhat degenerate situation in which the mixing and chemistry competed only weakly. While the 1D model reproduced basic aspects of the 2D model's behavior (more precisely, its steady state) fairly well, the degeneracy rendered this a weak test of the accuracy of the detailed structure of the transilient matrix. This implied a lack of justification for a detailed investigation

[2]For example, one expects it to hold at the level of the jet core at the stratopause, where vertical gradients of absolute vorticity are small.

into how well the matrix could be mimicked by a tridiagonal (diffusion) matrix raised to some power. It also raised issues relevant to any possible future attempts to construct 2D chemical-transport models in equivalent latitude space.

However, the flux of tracer across the vortex edge, a more robust quantity in these calculations, was shown to be very small, further confirming the picture of a very isolated vortex developed in chapter 6. This enabled a comprehensive description of the Lagrangian circulation in the model's extratropics to be constructed.

## 8.2 Final remarks

### 8.2.1 Different dynamical regimes

It should be made clear that the numerical results in chapters 6 and 7 can be taken as representative only of a particular dynamical regime, in which the inner vortex remains robust and coherent. Because the observed vortex is in this regime most of the time, I have felt it worth the effort to characterize it in some detail. However, departures from this regime before the final spring warming are a well-documented and quasi-regular phenomenon in the Arctic. These departures are not only the much-studied "sudden warmings" but also less spectacular disturbances in which the vortex nonetheless breaks into multiple, dynamically active fragments (e.g. O'Neill et al. 1994).

During these events, one expects that there will not be a simple scale separation between the vortex and the filaments, due to the presence of intermediate-scale structures. Correspondingly, there will be no clear distinction between dynamically active and passive features; a secondary vortex of some intermediate size may survive for a while, may be sheared out and dissipated, or may possibly merge with another fragment. Additionally, the weakening (or complete destruction) of the main vortex removes much of the strong shear and strain which prevents filament rollup and active re-merger. These complications make the selection of a cutoff scale, beyond which features are assumed sheared out and eventually dissipated, more problematic, and violates a key assumption behind the LGR and coarse-graining techniques. Consequently, one expects that these techniques may not be useful for estimating the net transport due to major vortex fragmentations. When such events occur, it is possible — though not certain — that they contribute significantly to the seasonally averaged exchange of air between the vortex and midlatitudes. How to quantify this transport is an open problem.

### 8.2.2 Polar vortex as prototype transport problem

As stated at the outset, the goal of this work has been to understand stratospheric mixing. The subject has fundamental importance for being an essential component of a complete understanding of both the dynamics and chemistry of the stratosphere.

There has been, nonetheless, an additional motivation behind the present work, perhaps most clearly manifest in chapters 3 and 4, which is to improve the framework for understanding mixing in general. The mixing in the extratropical winter stratosphere can be viewed in some respects as a prototype of a certain sort of mixing problem. As such, it is appropriate for developing ideas and techniques which may then be — perhaps — portable to other problems of this sort.

The extratropical stratosphere is a good prototype with which to start for several reasons. From a dynamical standpoint, it is relatively simple. It is close to inviscid, and

diabatic heating, besides being generally weak, is essentially due to radiation alone, a relatively well-understood physical process. The flow is strongly dominated by the largest scales, which are reasonably well observed despite the coarseness of the observational network. The large-scale flow dynamics are well described, at least qualitatively, by balance models such as quasi-geostrophy, providing a strong theoretical basis which aids greatly in understanding some aspects of the mixing problem. More subtle properties also conspire to simplify the problem; the dominance of the polar vortex in the wintertime extratropical flow, and the consequent scale separation between the vortex and the rest of the structures in the PV field, provide a good example. Finally, the importance of stratospheric ozone to life on earth provides motivation, not to mention material resources, for working on stratospheric mixing.

Of the other problems which resemble the extratropical stratospheric one in various respects, all appear more difficult, for one reason or another.

The mixing between the tropical and extratropical stratospheres is one important example. Some valuable information has been gained by studies which have approached the problem through diagnostics of tracer-tracer correlations measured *in situ* by aircraft (e.g. Volk et al. 1996; Minschwaner et al. 1996). However, as noted previously, this is in some sense solving an inverse problem. To the extent that the forward problem of diagnosing the mixing directly from observed wind and tracer fields is tractable, one would like to pursue this approach as well. A major difficulty is that observational analyses in the stratosphere depend heavily on satellite measurements, satellites measure temperature, and due to the weakness of balance constraints in the tropics, temperature measurements alone do not lead to good analyses of winds. As a result, stratospheric wind analyses appear unreliable for tropical transport calculations, as noted by Waugh (1996).

The mixing across the extratropical tropopause is another problem of great current interest. While some aspects of extratropical stratosphere-troposphere exchange probably do not depend on the details of mixing processes near the tropopause (Holton et al. 1995), the chemistry of the extratropical upper troposphere and lowermost stratosphere surely do. Again, this problem is harder than the stratospheric one, for several reasons. Phase changes of water can directly complicate the thermodynamics, and the presence of clouds makes understanding of the role of radiation more difficult. (Remember that, since the tropopause appears to be best defined as a PV isosurface, a good understanding of the nonconservative processes which can change PV is essential.) The relative shortness of diabatic time scales in the troposphere may make layerwise two-dimensional thinking less useful, although eddy time scales can be faster than in the stratosphere as well, partially mitigating this. Even considering the dry, adiabatic component of the dynamics in isolation, the range of space and time scales in the flow field is wider than in the stratosphere. Rollup of filaments into small secondary vortices is common (Appenzeller et al. 1996), and presumably these can re-merge with the main "vortex" (which in this case is the stratosphere).

Given these difficulties, a multifaceted, deliberate approach is surely necessary if a reliable, quantitative understanding is to be forthcoming. One can imagine applying techniques like those used in this thesis, but with careful adaptations to the particular nature of the problem. Continuous shuttling back and forth between models and observations is required. Observations will always be lacking in some respects. In many cases, only by testing a diagnostic technique on output from a numerical model can one be sure that the technique is, at least in principle, capable of producing meaningful results. Then, one must understand the failings of the available observational data sets, and how these affect the diagnostic techniques. Lastly, one should be able to produce from these diagnostics some measures of

the mixing which are usable in models which do not represent the mixing process explicitly. At every stage, any opportunity for constraining the problem by the application of solid theoretical results should be seized. In this thesis and in SPW, we have tried to demonstrate this approach for the extratropical winter stratospheric mixing problem.

# Appendix A

# Trajectory methods

This appendix describes various aspects of several Lagrangian trajectory algorithms which have been used in stratospheric studies.

## A.1 Reverse domain filling

The reverse domain filling (RDF) technique (Sutton et al. 1994; Yang 1995; Schoeberl and Newman 1995) may be viewed as a nondiffusive offline advection scheme which can be applied for a predetermined, finite time period. Given a set of wind fields covering some particular time period, and a tracer field at the beginning of that period, RDF allows one to generate a high-resolution tracer field at the end of the period which, in general, includes structure at smaller scales than were present in the initial field.

RDF works as follows. One initializes a large number of Lagrangian trajectory calculations with a set of hypothetical point parcels located at the nodes of a regular grid. One then performs the trajectory calculations backwards in time, starting at the time at which the high-resolution field is desired, and ending at some earlier time. At that earlier time, the parcels are given tracer values by interpolating the tracer field, which is given at the earlier time, to the parcels' positions at that time. Finally, the parcels are placed back on the nodes where their trajectories were initialized, but labeled by the tracer values they have been given as a result of the calculation. One is then left with a numerical representation of a tracer field on a regular grid. The RDF calculations discussed in chapters 6 and 7 were performed on a $1^\circ \times 1^\circ$ latitude-longitude grid.

## A.2 Forward domain filling

Alternatively, one could initialize the parcels on a regular grid and run the calculations forwards in time; this is the standard, or "forward", domain filling technique (Fisher et al. 1993). In this case, however, at the end of the calculation one is left with parcels spaced irregularly, making further analysis difficult. The reason for doing RDF is to avoid this difficulty.

## A.3 Contour advection

RDF is closely related to contour advection (CA) (Waugh and Plumb 1994: Norton 1994). This technique advects (in forward time) a set of one-dimensional contours imbedded in

two-dimensional space. Each contour is represented by a set of closely spaced points. A "renoding" algorithm, developed for use in the *contour dynamics* modeling technique (e.g. Dritschel 1989a), adds and removes points as needed to maintain the desired resolution along the contour. The advantage of contour advection over RDF is that it allows essentially infinite resolution of fine-scale structure, as represented by the contours, at modest computational cost. The disadvantage is that it yields no information about the tracer field away from the chosen contours. By choosing enough contours, one can reproduce the field at any desired resolution, but this increases the computational cost. The contour representation may be more awkward than the grid representation for some sorts of post-processing. It should also be noted that RDF generalizes straightforwardly to 3D, while CA does not[1].

Comparisons of CA and RDF have been performed by Yang (1995), Schoeberl and Newman (1995), and SPW. These show that the two techniques yield results which are identical down to the resolution of the RDF grid.

## A.4 Trajectory algorithms and validation

The trajectories themselves are computed from the given wind fields; typically these are interpolated linearly in time to allow a time step for the trajectory calculations which is smaller than the interval at which the winds are given. In our code, a fourth-order Runge-Kutta algorithm is used for the advection itself. This is probably the most common choice, though others are of course possible.

Several studies of the accuracy of stratospheric trajectory calculations have been performed, including Knudsen and Carver (1994), Morris et al. (1995), and Knudsen et al. (1996). The first and third included comparisons to observed balloon trajectories, while the second investigated (among other things) the differences which result from using wind analyses from different meteorological centers. These studies have shown that errors in individual stratospheric trajectories generally increase quickly, so that after only a few days they are generally substantial. This appears to be due primarily to errors in the wind fields, rather than to the trajectory algorithms themselves. Vertical interpolation error can be a significant problem when the number of vertical levels in a given height range is at the lower end of the range typical of current analyses (Knudsen et al. 1996). The isentropic approximation appears to be a good one for periods of up to 7-10 days (Morris et al. 1995).

Much of the error in individual trajectories is due to strong dispersion in the local mean flow direction (Knudsen and Carver 1994; Morris et al. 1996), which is also usually approximately parallel to tracer contours. For this reason, ensembles containing large numbers of trajectories (such as are involved in CA or RDF calculations) seem to do fairly well at identifying filaments and locating them in the cross-flow direction, despite relatively large total (i.e. net scalar distance) errors in individual trajectories. Comparisons of fine-scale structures generated by various sorts of advection calculations to high-resolution *in situ* tracer observations have been performed by Waugh et al. (1994) Plumb et al. (1994), Orsolini et al. (1995), Newman et al. (1996), and Orsolini et al. (1997). In many cases these show remarkably good agreement, though there are also cases where agreement is poor. On balance, however, it is clear that these methods provide some definite skill in predicting

[1]The obstacle to developing a 3D "surface advection" algorithm is the need for a renoding scheme to add and remove points as necessary to preserve curvature. This task is much more difficult in 3D than 2D, and a satisfactory solution has not been obtained to date.

the development of small-scale tracer structures from coarse large-scale data in the winter extratropical stratosphere.

This is an impressive, and perhaps even counterintuitive result. The reason the advection calculations work as well as they do is clearly the downscale nature of the enstrophy cascade in 2D turbulence, and the fact that this downscale cascade is to some extent spectrally *nonlocal*, and so governed by flow structures at the largest scale.

# Appendix B

# Relationship between different integral constraints

This appendix explains the relationship between the results of Schneider (1987), McIntyre and Norton (1990), Juckes (1987), and the main result of section 6.2.

First write the inviscid, shallow-water vorticity equation in flux form

$$\frac{\partial \zeta_a}{\partial t} + \nabla \cdot (\mathbf{u}\zeta_a) = 0$$

Then perform an area integral of the entire equation over the area enclosed by some contour, for the moment unspecified, and note that $\zeta_a$ is simply the layer thickness times the PV. The result is

$$\int\int \frac{\partial (hq)}{\partial t} dA + \int\int \nabla \cdot (\mathbf{u}hq) dA = 0 \tag{B.1}$$

Both Schneider (1987) and McIntyre and Norton (1990) proceeded by assuming a steady state, so that the first term in (B.1) vanishes. The difference between the two results is that Schneider let the contour be one of constant absolute vorticity, so that he obtained

$$\zeta_a \int\int \nabla \cdot \mathbf{u} dA = 0$$

implying that if the contour has a value other than zero, the integrated *flow divergence* $(\nabla \cdot \mathbf{u})$ must vanish. McIntyre and Norton let the contour be one of constant PV, so that they obtained

$$q \int\int \nabla \cdot (\mathbf{u}h) dA = 0$$

so that for a nonzero PV contour the integrated *mass flux divergence* $(\nabla \cdot ((\mathbf{u}h))$must vanish. Nominally, they derived their result in three dimensions, so that the thickness $h$ was replaced by the isentropic density $\sigma$. However, they made an additional assumption which has the same effect as assuming no vertical gradient of absolute vorticity, which makes their system formally identical to the shallow-water system.

One could also let the contour be one of constant thickness, in which case the result would be

$$h \int\int \nabla \cdot (\mathbf{u}q) dA = 0$$

or no net PV flux into or out of the contour. Another possibility which yields an interesting

result is to choose a streamfunction contour, as discussed by Sardeshmukh and Hoskins (1988).

The difference between all of these results and the one presented in section 6.2 is that the latter assumes only that the mass enclosed by a PV contour is steady; *the flow fields need not be.* The above results all assume a true steady state, that is, $\frac{\partial}{\partial t} = 0$ everywhere. Otherwise the first term in B.1 does not vanish. For a contour which is moving, taking the time derivative outside the integral in that term results in the introduction of a boundary term. The contribution of the present work is to show that this does not change the result. The results of Schneider (1987) and McIntyre and Norton (1990) are special cases of the significantly more general result proved in section 6.2.

Juckes (1987) proved a different, related result. He proved that under quasi-geostrophic scaling, in three dimensions, the net mass flux across a PV contour is smaller by a factor of order Rossby number than the zonal mean mass flux. Juckes did *not* assume a pure steady state, but rather only a steady state of the sort assumed in section 6.2, that is, with the time derivative outside the integral (before it is taken zero). This result does not tell us anything directly about the nature of the balance in an inviscid MLM steady state, since *all* the terms in the relevant equation are order Rossby number smaller than the zonal mean mass flux (which does not appear in the equation).

# Appendix C

# List of Acronyms

| Acronym | Stands for | Introduced in section |
|---|---|---|
| CA | Contour advection | 4.1.1 |
| CACG | Contour advection with coarse-graining | 4.2.2 |
| LGR | Local gradient reversal | 4.2.2 |
| MLM | Modified Lagrangian mean | 2.3.3 |
| PV | Potential vorticity | 1.1 |
| PWP | Polvani et al. (1995) | 2.4 |
| RA | Relaxation to the average | 3.5 |
| RDF | Reverse domain filling | 4.1.1 |
| SPW | Sobel et al. (1997) | 1.2 |

# References


Andrews, D. G., and M. E. McIntyre, 1976: Planetary waves in horizontal and vertical shear: the generalized Eliassen-Palm relation and the mean zonal acceleration, *J. Atmos. Sci.*, **33**, 2031-2048.

Andrews, D. G., and M. E. McIntyre, 1978: An exact theory of nonlinear waves on a Lagrangian-mean flow, *J. Fluid. Mech.*, **89**, 609-646.

Andrews, D. G., J. R. Holton, and C. B. Leovy, 1987: *Middle Atmosphere Dynamics*, Academic Press, 489 pp.

Appenzeller, C., H. C. Davies, and W. A. Norton, 1996: Fragmentation of stratospheric intrusions. *J. Geophys. Res.*, **101**, 1435-1456.

Arakawa, A., 1993: Closure assumptions in the cumulus parameterization problem, in *The Representation of Cumulus Convection in Numerical Models*, Meteorological Monograph **46**, K. A. Emanuel and D. J. Raymond, Eds., American Meteorological Society.

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

Bacmeister, J. T., M. R. Schoeberl, L. R. Lait, P. A. Newman, and B. Gary, 1990: Small-scale waves encountered during AASE, *Geophys. Res. Lett.*, **17**, 349-352.

Bacmeister, J. T., M. R. Schoeberl, M. Loewenstein, J. R. Podolske, S. E. Strahan, and R. K. Chan, 1992: An estimate of the relative magnitude of small-scale tracer fluxes, *Geophys. Res. Lett.*, **19**, 1101-1104.

Bacmeister, J. T., 1993: Mountain-wave drag in the stratosphere and mesosphere inferred from observed winds and a simple mountain-wave parameterization scheme, *J. Atmos. Sci.*, **50**, 377-399.

Bacmeister, J. T., P. A. Newman, B. L. Gary, and R. K. Chan, 1994: An algorithm for forecasting mountain wave-related turbulence in the stratosphere, *Wea. and Forecasting*, **9**, 241-253.

Bacmeister, J. T., S. D. Eckermann, P. A. Newman, L. Lait, K. R. Chan, M. Loewenstein, M. H. Proffitt, B. L. Gary, 1996: Stratospheric horizontal wavenumber spectra of winds, potential temperature, and atmospheric tracers observed by high-altitude aircraft, *J. Geophys. Res.*, **101**, 9441-9470.

Baldwin, M. P., and J. R. Holton, 1988: Climatology of the stratospheric polar vortex and planetary wave breaking, *J. Atmos. Sci.*, **45**, 1123-1142.

Balluch, M., and P. H. Haynes, 1997: Quantification of lower stratospheric mixing processes using aircraft data, *J. Geophys. Res.*, submitted.

Batchelor, G. K., and A. A. Townsend, 1956: Turbulent Diffusion, *Surveys in Mechanics*, 352-399, Cambridge U. Press, London.

Beigie, D., A. Leonard, and S. Wiggins, 1994: Invariant manifold templates for chaotic advection, *Chaos, Solitons & Fractals*, **4**, 749-868.

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.

Betts, A. K., 1986: A new convective adjustment scheme. I: Observational and theoretical basis, *Q. J. R. Meteorol. Soc.*, **112**, 667-691.

Betts, A. K., and M. J. Miller, 1986: A new convective adjustment scheme. II: Single column tests using GATE wave, BOMEX, ATEX, and arctic air-mass data sets, *Q. J. R. Meteorol. Soc.*, **112**, 693-709.

Bowman, K. P., 1993: Barotropic simulation of large-scale mixing in the Antarctic polar vortex, *J. Atmos. Sci.*, **50**, 2901-2914.

Brewer, A. W., 1949: Evidence for a world circulation provided by the measurements of helium and water vapor distribution in the stratosphere, *Q. J. R. Meteorol. Soc.*, **95**, 351-363.

Browning, G. L., J. J. Hack, and P. N. Swarztrauber, 1989: A comparison of three numerical methods for solving the shallow water equations on the sphere, *Mon. Wea. Rev.*, **117**, 1058-1075.

Butchart, N., and E. Remsberg, 1986: The area of the stratospheric polar vortex as a diagnostic for tracer transport on an isentropic surface, *J. Atmos. Sci*, **43**, 1319-1339.

Charney, J. G., and P. G. Drazin, 1961: Propagation of planetary-scale disturbances from the lower into the upper atmosphere, *J. Geophys. Res.*, **66**, 83-109.

Charney, J. G., and A. Eliassen, 1964: On the growth of the hurricane depression, *J. Atmos. Sci.*, **21**, 68-75.

Chen, P., 1994: The permeability of the antarctic vortex edge, *J. Geophys. Res.*, **99**, 20563-20571.

Dahlberg, S. P., and K. P. Bowman, 1994: Climatology of large-scale isentropic mixing in the arctic winter stratosphere from analyzed winds, *J. Geophys. Res.*, **99**, 20,585-20,599.

Dobson, G. M. B., 1956: Origin and distribution of the polyatomic molecules in the atmosphere. *Proc. R. Soc. London A*, **236**, 187-193.

Dritschel, D. G., 1989a: Contour dynamics and contour surgery: Numerical algorithms for extended, high-resolution modeling of vortex dynamics in two-dimensional, inviscid, incompressible flows, *Comp. Phys. Rep.*, **10**, 77-146.

Dritschel, D. G., 1989b: On the stabilization of a two-dimensional vortex strip by adverse shear, *J. Fluid Mech.*, **206**, 193-221.

Dritschel, D. G., and L. M. Polvani, 1992: The roll-up of vorticity strips on the surface of a sphere, *J. Fluid Mech.*, **234**, 47-69.

Duan, J., and S. Wiggins, 1996: Fluid exchange across a meandering jet with quasiperiodic variability, *J. Phys. Oceanogr.*, **26**, 1176-1188.

Dunkerton, T. S., and D. P. Delisi, 1985: Evolution of potential vorticity in the winter stratosphere of January-February 1979, *J. Geophys. Res.*, **91**, 1199-1208.

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.

Eluszkiewicz, J., R. A., Plumb, and N. Nakamura, 1995: Dynamics of wintertime stratospheric transport in the Geophysical Fluid Dynamics Laboratory SKYHI general circulation model, *J. Geophys. Res.*, **100**, 20883-20900.

Eluszkiewicz, J., D. Crisp, R. Zurek, L. Elson, E. Fishbein, L. Froidevaux, J. Waters, R. G. Grainger, A. Lambert, R. Harwood, and G. Peckham, 1996: Residual circulation in the stratosphere and lower mesosphere as diagnosed from Microwave Limb Sounder data, *J. Atmos. Sci.*, **53**, 217-240.

Emanuel, K. A., J. D. Neelin, and C. S. Bretherton, 1994: On large-scale circulations in convecting atmospheres, *Q. J. R. Meteorol. Soc.*, **120**, 1111-1143.

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.

Fisher, M., A. O'Neill, and R. Sutton, 1993: Rapid descent of mesospheric air into the stratospheric polar vortex, *Geophys. Res. Lett.*, **20**, 1267-1270.

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

Frisch, U., 1989: Lectures on turbulence and lattice gas hydrodynamics, *Lecture Notes on Turbulence: Lecture Notes from the NCAR-GTP Summer School of June 1987*, J. R. Herring and J. C. McWilliams, Eds., World Scientific, New Jersey.

Fung, I., K. Prentice, E. Matthews, J. Lerner, and G. Russell, 1983: Three-dimensional tracer model study of atmospheric $CO_2$: Response to seasonal exchanges with the terrestrial biosphere, *J. Geophys. Res.*, **88**, 1281-1294.

Garcia, R. R., 1990: Parameterization of planetary wave breaking in the middle atmosphere, *J. Atmos. Sci.*, **48**, 1405-1419.

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

Goody, R. M., and Y. L. Yung, 1989: *Atmospheric Radiation*, Oxford University Press, 519 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.

Hartmann, D. L., L. E. Heidt, M. Loewenstein, J. R. Podolske, J. F. Vedder, W. L. Starr, and S. E. Strahan, 1989: Transport into the south polar vortex in early spring, *J. Geophys. Res.*, **94**, 16779-16795.

Haynes, P. H., and W. E. Ward, 1993: The effect of realistic radiative transfer on potential vorticity structures, including the influence of background shear and strain, *J. Atmos. Sci.*, **50**, 3431-3453.

Haynes, P. H., and J. Anglade, 1997: The vertical-scale cascade in atmospheric tracers due to large-scale differential advection, *J. Atmos. Sci.*, **54**, 1121-1136.

Hochstadt, H., 1973: *Integral Equations*, John Wiley & Sons, 282p.

Hoerling, M. P., T. K. Schaack, and A. S. Lenzen, 1993: A global analysis of stratosphere-troposphere exchange, *Mon. Wea. Rev.*, **126**, 162-172.

Holton, J. R., P. H. Haynes, M. E. McIntyre, A. R. Douglass, R. B. Rood, L. Pfister, 1995: Stratosphere-troposphere exchange, *Rev. Geophys.*, **33**, 403-439.

Hoskins, B. J., M. E. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps, *Quart. J. Roy. Met. Soc.*, **111**, 877-946.

Juckes, M. N., 1987: *Studies of stratospheric dynamics*, Ph. D. thesis, University of Cambridge, England.

Juckes, M. N., and M. E. McIntyre, 1987: High-resolution one-layer model of breaking planetary waves in the stratosphere, *Nature*, **328**, 590-596.

Juckes, M. N., 1989: A shallow water model of the winter stratosphere, *J. Atmos. Sci.*, **46**, 2934-2955.

Juckes, M. N., 1996: Comments on "On the subtropical edge of the stratospheric surf zone", *J. Atmos. Sci.*, **53**, 3770-3771.

Knudsen, B. M., and G. D. Carver, 1994: Accuracy of the isentropic trajectories calculated for the EASOE campaign, *Geophys. Res. Lett.*, **21**, 1199-1202.

Knudsen, B. M., J. M. Rosen, N. T. Kjome, and A. T. Whitten, 1996: Comparison of analyzed stratospheric temperatures and calculated trajectories with long-duration balloon data, *J. Geophys. Res.*, **101**, 19137-19145.

Kraichnan, R. H., 1959: The structure of isotropic turbulence at very high Reynolds numbers. *J. Fluid Mech.*, **5**, 493-543.

Kraichnan, R. H., 1976: Eddy viscosity in two and three dimensions, *J. Atmos. Sci.*, **33**, 1521-1536.

Lamarque, J.-F., and P. G. Hess, 1994: Cross-tropopause mass exchange and potential vorticity budget in a simulated tropopause folding, *J. Atmos. Sci.*, **51**, 2246-2269.

Leovy, C. B., C-R. Sun, M. H. Hitchman, E. E. Remsberg, J. M. Russell, III, L. L. Gordley, J. C. Gille, and L. V. Lyjak, 1985: Transport of ozone in the middle stratosphere: evidence for planetary wave breaking, *J. Atmos. Sci.*, **42**, 230-244.

Lilly, D. K., 1989: Lecture notes, in *Lecture Notes on Turbulence: Lecture Notes from the NCAR-GTP Summer School of June 1987*, J. R. Herring and J. C. McWilliams, Eds., World Scientific, New Jersey.

Malhotra, N., and S. Wiggins, 1997: Geometric structures, lobe dynamics, and Lagrangian transport in flows with aperiodic time-dependence, with applications to Rossby wave flow, *J. Nonlinear Sci.*, accepted subject to minor revision.

Manney, G. L., R. W. Zurek, A. O'Neill, and R. Swinbank, 1994: On the motion of air through the stratospheric polar vortex, *J. Atmos. Sci.*, **51**, 2973-2994.

McIntyre, M. E., 1980: Towards a Lagrangian-mean description of stratospheric circulations and chemical transports, *Phil. Trans. R. Soc. Lond. A*, **296**, 129-148.

McIntyre, M. E., and T. N. Palmer, 1983: Breaking planetary waves in the stratosphere, *Nature*, **305**, 593-600.

McIntyre, M. E., and T. N. Palmer, 1984: Surf zone in the stratosphere, *J. Atm. and Terr. Phys.*, **46**, 825-849.

McIntyre, M. E., and T. N. Palmer, 1985: A note on the general concept of wave breaking for Rossby and gravity waves, *PAGEOPH*, **123**, 964-975.

McIntyre, M. E., and W. A. Norton, 1990: Dissipative wave-mean interactions and the transport of vorticity or potential vorticity, *J. Fluid Mech.*, **212**, 403-435.

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.

McIntyre, M. E., 1995: The stratospheric polar vortex and sub-vortex: fluid dynamics and midlatitude ozone loss, *Phil. Trans. R. Soc. Lond. A*, **352**, 227-240.

Miller, P. D., C. K. R. T. Jones, A. M. Rogerson, and L. J. Pratt, 1997: Quantifying transport in numerically generated velocity fields, submitted to *Physica D*.

Minschwaner, K., A. E. Dessler, J. W. Elkins, C. M. Volk, D. W. Fahey, M. Loewenstein, J. R. Podolske, A. E. Roche, and K. R. Chan, 1996: Bulk properties of isentropic mixing into the tropics in the lower stratosphere, *J. Geophys. Res.*, **101**, 9433-9439.

Morris, G. A., M. R. Schoeberl, L. C. Sparling, P. A. Newman, L. R. Lait, L. Elson, J. Waters, R. A. Suttie, A. Roche, J. Kumer, and J. M. Russell III, 1995: Trajectory mapping and applications to data from the Upper Atmosphere Research Satellite, *J. Geophys. Res.*, **100**, 16491-16505.

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.

Nakamura, N., 1996: Two-dimensional mixing, edge formation, and permeability diagnosed in an area coordinate, *J. Atmos. Sci.*, **53**, 1524-1537.

Nakamura, N., and J. Ma, 1996: Modified Lagrangian-mean diagnostics of the stratospheric polar vortices: part II. Nitrous oxide and seasonal barrier migration in Cryogenic Etalon Limb Array Spectrometer measurements and SKYHI general circulation model, *J. Geophys. Res.*, submitted.

Nash, E. R., P. A. Newman, J. E. Rosenfield, and M. R. Schoeberl,1996: An objective determination of the polar vortex using Ertel's potential vorticity, *J. Geophys. Res.*, **101**, 9471-9478.

Newman, P. A., 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.

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

Newman, P. A., L. R. Lait, L. R., M. R. Schoeberl, M. Seablom, L. Coy, R. Rood, R. Swinbank, M. Proffitt, M. Loewenstein, J. R. Podolske, J. W. Elkins, C. R. Webster, R. D. May, D. W. Fahey, G. S. Dutton, and K. R. Chan, 1996: Measurements of polar vortex air in the midlatitudes, *J. Geophys. Res.*, **101**, 12879-12891.

Ngan, K., and T. G. Shepherd, 1997: Chaotic mixing and transport in Rossby-wave critical layers, *J. Fluid Mech.*, **334**, 315-351.

Norton, W. A., 1994: Breaking Rossby waves in a model stratosphere diagnosed by a vortex-following coordinate system and a technique for advecting material contours. *J. Atmos. Sci.*, **51**, 654-673.

Norton, W. A., and M. P. Chipperfield, 1995: Quantification of the transport of chemically activated air from the northern hemisphere polar vortex, *J. Geophys. Res.*, **100**, 25817-25850.

O'Neill, A., and V. D. Pope, 1988: Simulations of linear and nonlinear disturbances in the stratosphere, *Q. J. R. Meteorol. Soc.*, **114**, 1063-1110.

O'Neill, A., W. L. Grose, V. D. Pope, H. Maclean, and R. Swinbank, 1994: Evolution of the stratosphere during northern winter 1991/92 as diagnosed from U. K. Meteorological Office analyses, *J. Atmos. Sci.*, **51**, 2800-2817.

Orsolini, Y., P. Simon, and D. Cariolle, 1995: Filamentation and layering of an idealized tracer by observed winds in the lower stratosphere, *Geophys. Res. Lett.*, **22**, 839-842.

Orsolini, Y., G. Hansen, U.-P. Hoppe, G. L. Manney, and K. H. Fricke, 1997: Dynamical modeling of wintertime lidar observations in the Arctic: ozone laminae and ozone depletion, *Q. J. R. Meteorol. Soc.*, **123**.

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

Ottino, J. M., 1990: Mixing, chaotic advection, and turbulence, *Annu. Rev. Fluid Mech.*, **22**, 207-253.

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 travelling Rossby waves, *Geophys. Astrophys. Fluid Dyn.*, **58**, 285-319.

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

Pierrehumbert, R. T., and H. Yang, 1993: Global chaotic mixing on an isentropic surface, *J. Atmos. Sci.*, **50**, 2462-2480.

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

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

Plumb., R. A., amd M. K. W. Ko, 1992: Interrelationships between mixing rations of long-lived stratospheric constituents, *J. Geophys. Res.*, **97**, 10145-10156.

Plumb, R. A., D. W. Waugh, R. J. Atkinson, P. A. Newman, L. R. Lait, M. R. Schoeberl, E. V. Browell, A. J. Simmons, and M. Loewenstein, 1994: Intrusions into the lower stratospheric arctic vortex during the winter of 1991/92, *J. Geophys. Res.*, **99**, 1089-1105.

Plumb, R. A., 1996: A "tropical pipe" model of stratospheric transport, *J. Geophys. Res.*, **101**, 3957-3972.

Polvani, L. M., and R. A. Plumb, 1992: Rossby wave breaking, microbreaking, filamentation, and secondary vortex formation: the dynamics of a perturbed vortex, *J. Atmos. Sci.*, **49**, 462-476.

Polvani, L. M., D. W. Waugh and R. A. Plumb, 1995: On the subtropical edge of the stratospheric surf zone, *J. Atmos. Sci.*, **52**, 1288-1309.

Polvani, L. M., D. W. Waugh and R. A. Plumb, 1996: Reply, *J. Atmos. Sci.*, **53**, 3772-3775.

Polvani, L. M., and R. Saravanan, 1997: The vertical structure of Rossby wave breaking in the polar wintertime stratosphere, *Tenth Conference on the Middle Atmosphere*, abstracts volume, American Meteorological Society.

Prandtl, L., 1925: Bericht über untersuchungen zur ausgebildeten turbulenz, *Z. angew. Math. Mech.*, **5**, 136.

Prather, M. J., 1986: Numerical advection by conservation of second-order moments, *J. Geophys. Res.*, **91**, 6671-6681.

Prather, M. J., and A. H. Jaffe, 1990: Global impact of the Antarctic ozone hole: chemical propagation, *J. Geophys. Res.*, **95**, 3473-3492.

Prather, M. J., and E. E. Remsberg, eds., 1993: *The Atmospheric Effects of Stratospheric Aircraft: Report of the 1992 Models and Measurements Workshop. Volume I — Workshop Objectives and Summary.* NASA Reference Pub. 1292, Vol. I.

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

Prather, M. J., 1996: Time scales in atmospheric chemistry: theory, GWPs for $CH_4$ and $CO$, and runaway growth, *Geophys. Res. Lett.*,

Raymond, D. J., 1994: Convective processes and tropical atmospheric circulations, *Q. J. R. Meteorol. Soc.*,**120**, 1431-1455.

Raymond, W. H., and R. B. Stull, 1990: Application of transilient turbulence theory to mesoscale numerical weather forecasting, *Mon. Wea. Rev.*, **118**, 2471-2499.

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

Rogerson, A. M., P. D. Miller, L. J. Pratt, C. K. R. T. Jones, and J. Biello, 1997: Chaotic mixing in a barotropic meandering jet. Manuscript submitted to *J. Phys. Oceanogr.*.

Rood, R. B., 1985: A critical analysis of the concept of planetary wave breaking, *PAGEOPH*, **123**, 733-755.

Rosenfield, J. E., M. R. Schoeberl, L. R. Lait, P. A. Newman, M. H. Proffitt, and K. K. Kelly, 1990: Radiative heating rates during the Airborne Arctic Stratospheric Experiment, *Geophys. Res. Lett.*, **17**, 345-348.

Rosenlof, K. H., and J. R. Holton, 1993: Estimates of the stratospheric residual circulation using the downward control principle, *J. Geophys. Res.*, **98**, 10465-10479.

Rosenlof, K. H., 1995: Seasonal cycle of the residual mean meridional circulation in the stratosphere, *J. Geophys. Res.*, **100**, 5173-5191.

Salby, M. L., R. R. Garcia. D. O'Sullivan, and P. Callaghan, 1990: The interaction of horizontal eddy transport and thermal drive in the stratosphere, *J. Atmos. Sci.*, **47**, 1647-1665.

Saravanan, R., and D. G. Dritschel, 1994: Three-dimensional quasi-geostrophic contour dynamics, with an application to stratospheric vortex dynamics, *Q. J. Roy. Met. Soc.*, **120**, 1267-1297.

Sardeshmukh, P. D., and B. J. Hoskins, 1988: The generation of global rotational flow by steady idealized tropical divergence, *J. Atmos. Sci.*, **45**, 1228-1251.

Schneider, E. K., 1987: A simplified model of the modified Hadley circulation, *J. Atmos. Sci.*, **44**, 3311-3328.

Schoeberl, M. R., and D. L. Hartmann, 1991: The dynamics of the stratospheric polar vortex and its relation to springtime ozone depletions, *Science*, **251**, 46-52.

Schoeberl, M. R. and P. A. Newman, 1995: A multiple level trajectory analysis of vortex filaments, *J. Geophys. Res.*, **52**, 25,801-25-816.

Siegmund, P. C., P. F. S. van Velthoven, and H. Kelder, 1996: Cross-tropopause transport in the extratropical northern winter hemisphere, diagnosed from high-resolution ECMWF data, *Q. J. R. Meteorol. Soc.*, **122**, 1921-1941.

Sobel, A. H., R. A. Plumb, and D. W. Waugh, 1997: Methods of calculating transport across the polar vortex edge, *J. Atmos. Sci.*, in press.

Sreenivasan, K. R., 1991: On local isotropy of passive scalars in turbulent shear flows, in *Turbulence and Stochastic Processes: Kolmogorov's Ideas 50 Years On*, J. C. R. Hunt, O. M. Phillips, and D. Williams, eds., London Royal Society.

Strahan, S. E., and J. D. Mahlman, 1994: Evaluation of the SKYHI general circulation model using aircraft $N_2O$ measurements. 2: Tracer variability and diabatic meridional circulation, *J. Geophys. Res.*, **99**, 10,319-10,332.

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

Stull, R. B., 1988: *Introduction to Boundary Layer Meteorology*, Kluwer Academic Publishers, 666p.

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

Stull, R. B., 1995: Transilient turbulence theory, in *The Planetary Boundary Layer and its Parameterization*. Notes from an NCAR Summer Colloquium, C. H. Moeng, ed., notes distributed by NCAR. Notes from Stull's lecture written up by A. Sobel and B. Hannegan.

Sutton, O. G., 1949: *Atmospheric Turbulence*, Methuen & Co. LTD., 107 pp.

Sutton, R. T., H. MacLean, R. Swinbank, A. O'Neill, F. W. Taylor, 1994: High-resolution stratospheric tracer fields estimated from satellite observations using lagrangian trajectory calculations, *J. Atmos. Sci.*, **51**, 2995-3005.

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

Taylor, G. I., 1935: Statistical theory of turbulence, part I, *Proc. Roy. Soc. A.*, **CLI**, 421-444.

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.

Tuck, A. F., 1989: Synoptic and chemical evolution of the Antarctic vortex in late winter and early spring, 1987, *J. Geophys. Res.*, **94**, 11,687-11,737.

Tuck, A. F., J. M. Russell, and J. E. Harries, 1993: Stratospheric dryness: antiphased dessication over Micronesia and Antarctica, *Geophys. Res. Lett.*, **20**, 1227-1230.

Volk, C. M., J. W. Elkins, D. W. Fahey, R. J. Salawitch, G. S. Dutton, J. M. Gilligan, M. H. Proffitt, M. Loewenstein, and P. R. Podolske, 1996: In-situ measurements constraining exchange between the tropics and middle latitudes in the lower stratosphere, *Science*, **272**, 1763-1768, 1996.

Wauben, W. M. F., R. Bintanja, P. F. J. van Velthoven, and H. Kelder, 1997: On the magnitude of transport out of the Antarctic polar vortex, *J. Geophys. Res.*, **102**, 1229-1238.

Waugh, D. W., and D. G. Dritschel, 1991: The stability of filamentary vorticity in two-dimensional geophysical vortex-dynamics models, *J. Fluid Mech.*, **231**, 575-598.

Waugh, D. W., 1992: The efficiency of symmetric vortex merger, *Phys. Fluid A*, **4**, 1745-1758.

Waugh, D. W., and R. A. Plumb, 1994: Contour advection with surgery: a technique for investigating finescale structure in tracer transport, *J. Atmos. Sci.*, **51**, 530-540.

Waugh, D. W., R. A. Plumb, R. J. Atkinson, M. R. Schoeberl, L. R. Lait, P. A. Newman, M. Lowenstein, D. W. Toohey, L. M. Avallone, C. R. Webster, and R. D. May, 1994a: Transport out of the lower stratospheric arctic vortex by Rossby wave breaking, *J. Geophys. Res.*, **99**, 1071-1088.

Waugh, D. W., L. M. Polvani, and R. A. Plumb, 1994b: Nonlinear, barotropic response to a localized topographic forcing: formation of a "tropical surf zone" and its effect on interhemispheric propagation, *J. Atmos. Sci.*, **51**, 1401-1416.

Waugh, D. W., 1996: Seasonal variation of isentropic transport out of the tropical stratosphere, *J. Geophys. Res.*, **101**, 4007-4023.

Waugh, D. W., R. A. Plumb, J. W. Elkins, D. W. Fahey, K. A. Boering, G. S. Dutton, E. Keim, R.-S. Gao, B. C. Daube, S. C. Wofsy, M. Loewenstein, J. R. Podolske, K. R. Chan, M. H. Proffitt, K. K. Kelly, P. A. Newman, and L. R. Lait, 1997: Mixing of polar vortex air into middle latitudes as revealed by tracer-tracer scatter plots, *J. Geophys. Res.*, in press.

Waugh, D. W., and D. G. Dritschel, 1997: Vertical propagation of waves on edges of stratospheric polar vortices, *Tenth Conference on the Middle Atmosphere*, American Meteorological Society.

Wei, M.-Y., 1987: A new formulation of the exchange of mass and trace constituents between the stratosphere and troposphere, *J. Atmos. Sci.*, **44**, 3079-3086.

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

Wofsy, S. C., J. Eluszkiewicz, and R. A. Plumb, 1997: Tracer distributions and perturbations due to HSCTs in an ultra-compact, fully-empirical model of stratospheric transport, *The 1997 Conference on the Atmospheric Effects of Aviation*, Virginia Beach, Jorge Scientific.

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

Yang, H., 1995: Three-dimensional transport of the Ertel potential vorticity and $N_2O$ in the GFDL SKYHI model, *J. Atmos. Sci.*, **52**, 1513-1528.