PLANETARY WAVE ACTIVITY IN THE STRATOSPHERE

Li Dingmin

DOCTOR OF PHILOSOPHY

University of Edinburgh

1991

EX BIBL. UNIV. EDINBURGEN.

DECLARATION

I hereby declare that the work described in this thesis is my own unless otherwise stated in the text and the thesis has been composed and typed by myself.

Contents

ACKNOWLEDGMENT

I respectfully acknowledge the help of Dr. R. S. Harwood, my supervisor, who has kept a watchful eye over the whole progress of my study and whose support and interest in my thesis has encouraged and helped me considerably. I would also like to thank Dr. C. N. Duncan, also my supervisor, for help and useful discussions.

I express my gratitude to the Meteorology Department of Edinburgh University for providing generously the facilities for my research. I wish to thank all the staff in the department for their friendly and helpful attitude and all the assistance they provided during the period of my study. Particularly I wish to thank Brien Cameron for his kindly help.

My study and research was finacially supported by a grant from Chinese Acdemia Scinica, I express my gratitude for the support. The thanks also go to the U.K. Oversea Research Student Award for the partly support.

Finally, I extend my fondest thanks to my wife, Li Yimay, for her fully support, encouragement and patience during the course of this work.

Abstract

This thesis presents a diagnostic study of some aspects of stratospheric dynamics, from seeking a better approximation of winds using satellite derived data to the dynamics of wave activity and wave-mean flow interaction with the aid of newly developed theory.

An approximation scheme named "quasi-geostrophic wind", is tested and its merits discussed by comparison and contrast with some currently common ~~used~~ methods, viz. geostrophic, linear balance, nonlinear balance winds.

The dynamical behaviour of the stratosphere is diagnosed, using LIMS data, in the form of the definition of wave activity density given by Andrews(1987) and in terms of the generalized Eliassen-Palm flux theorem. As the theory is strictly applicable only to small amplitude waves, the study is mainly concentrated on a period of low wave activity in late Autumn.

There is good agreement in the changes of terms in the generalized Eliassen-Palm flux theorem. Time-latitude plots of Eliassen-Palm flux divergence, wave activity density and its rate of change, together with wave dissipation show a day to day coherence. A good qualitative agreement is also found between the change of zonal mean wind and wave activity in accordance with the theorem. The balance among the terms in the theorem is only approximate, an indication of the importance

of nonlinear effects neglected in the linear premise. The balance is improved when the nonconservative term is included.

An analogous expression to the generalized Eliassen-Palm flux theorem but for chemical tracers is derived theoretically and a corresponding wave activity defined in terms of chemical tracers such as ozone is presented. This analogous expression shows the relation between the tracer transport and the wave transience; after some simplifications this gives a novel method of estimating the transport coefficients. In the middle atmosphere good agreement is found between wave activity defined in terms of ozone mixing ratio with that defined in terms of potential vorticity. Results of an attempt to estimate transport coefficients by tracer wave transiences and the corresponding parameterized flux are also presented and the contribution due to dispersion is discussed. The estimation by this method is found capable of reflecting the large effects of wave dispersion in late winter. The transport coefficients from dispersion of fluid parcels alone are not fully satisfactory due to the appearance of a large area of negative coefficients in the monthly mean plots which may be caused by the limit of linear theory, the effects of chemical eddies (sources and sinks) or by the errors in the data that may hinder our estimation of wave transience.

Chapter 1

Introduction

Stratospheric dynamics are believed to be dominated by planetary waves propagating from below. The Eliassen–Palm (hereafter E–P) flux and generalized E–P flux theorems (i.e. E–P flux divergence theorem) express important features of propagation. The theorem shows explicitly the dependence of E-P flux divergence on the physical properties of the wave, viz. transience and nonconservative effects. Wave activity, which is the key idea in these theorems, is very hard to diagnose because its fundamental definition originally involves Lagrangian parcel displacements from some undisturbed basic or 'initial' state and determining the parcel displacements by constructing trajectories is practically extremely difficult. However Andrews (1987) recasted this definition and the related definition for nonconservative effects and expressed them by Eulerian quantities in various coordinates and for different approximations. The main thrust of this thesis is to

investigate how these ideas stand up in practice, the possibility of directly applying the E-P flux theorem in the diagnostic studies of stratospheric dynamics using satellite derived data and the implication of these ideas to tracer transport.

As the scope of this investigation is mainly confined to the diagnostic aspects of middle atmosphere dynamics, an up-to-date review of theoretical understandings of middle atmosphere dynamics which are closely related to the present study will be given in chapter 2. One important purpose of this thesis is to find out to what extent the theory, which has some unavoidable assumptions, for instance linearity, can represent the real processes found from satellite derived data.

Geostrophic winds have been used as a matter of course for a long time in stratospheric studies. However, since the work by Robinson (1986) the errors caused by this approximation have attracted much attention in middle atmospheric studies and several modified methods have been derived. Chapter 3 will compare and contrast several schemes for calculating stratospheric wind fields from height fields derived from satellite observed data. The schemes are also tested using height fields produced by numerical model and compared with the model output winds. The critical condition for each scheme to be a valid approximation is discussed. The diagnostic techniques involved in the whole investigation of this thesis are also included in the begining of chapter 3.

There has been growing emphasis in studies of atmospheric processes and tracer transport upon the usefulness and importance of analysing the isentropic distribution of Rossby-Ertel potential vorticity. The reasons for this emphasis lie firstly in the invertibility of the potential vorticity field to give the other dynamical fields describing balanced motion, and secondly in the behaviour of potential vorticity as a material conservative tracer in the absence of diabatic heating, friction and other forces. The direct observational evidence for the existence of stratospheric wave breaking was provided by isentropic maps of potential vorticity (IPV) (McIntyre and Palmer, 1984; Leovy, 1984) and the 'main vortex, surf-zone' mechanics was also based on the IPV maps. The study by Butchart and Remsberg (1986) of changes of areas enclosed by isopleths of potential vorticity in isentropic surfaces gives a diagnostic method to estimate the net effect of non-conservative processes and/or irreversible mixing to unresolvable scales. However this idea relies on the assumption of nondivergence of horizontal wind fields on isentropic surfaces. The divergence of horizontal wind may have significant influences on the area change. Chapter 4 presents an investigation of the possible effect of wind divergence on the change of the area.

In Chapter 5, as the main part of present investigation and the natural development of previous chapters, we present in detail our application of the generalized E-P flux theorem to satellite-derived observational data. Since the division of the

atmosphere into the 'mean state' and 'wave' deviation remains as an important approach to atmospheric research, the value of the generalized E-P flux theorem is readily apparent; it sets the links amongst wave transience, wave activity and transport and nonconservative sources and sinks. Through the appearance of E-P flux divergence as the sole force exerted by waves on zonal mean, it also reflects the effect of waves on the mean state. There are three main parts in the chapter. In the first part we give the introduction to the theorem and describe the feasibility of applying it to satellite observational data. The second part stresses the behaviour of stratospheric dynamical processes in terms of the the theorem. The coherence among the terms in the theorem itself, such as the evolution of wave activity, the correspondence between wave activity and E-P flux divergence, the relation between zonal mean state and the balance of the theorem in the real atmosphere is discussed. In the third part, we try to extend the application to the wave activity defined by the ozone mixing ratio in the place of potential vorticity. The theoretical basis of doing this is described in chapter 6. This both shows the relation between ozone and potential vorticity and acts as a check for our calculations. The brief conclusion of this chapter is also in the end of the part.

As mentioned above, potential vorticity has properties analogous to those of a material conserved chemical tracer under certain conditions. There should exist, therefore, some similarities between potential vorticity and chemical tracers in the

dynamical and transport processes, hence in the theoretical expressions such as equations and theorems describing those processes. Chapter 6 attempts to derive an analogous expression of generalized Eliassen-Palm flux theorem and a similar measure of wave activity in terms of chemical mixing ratio in the place of potential vorticity. As a result a version of wave activity density defined using mixing ratio is given and after simplification a way of calculating transport coefficients from wave transience is described. Finally some results of applying this idea to observational data to obtain values of the coefficients are shown and discussed.

Finally chapter 7 gives the summary and discussions of this thesis.

Chapter 2

The Current Understanding of The Stratospheric Dynamics

2.1 Planetary wave theory

Because the winter stratosphere is dominated by the waves with largest wave length (wave number 1 to 3), the theoretical understanding of stratospheric dynamics is closely related to the understanding of planetary waves, their formation, propagation, dissipation, breaking and other effects. As this thesis will deal with such issues as wave-mean flow interaction, wave effects on transport etc., the theory of planetary waves will be briefly reviewed in this section.

2.1.1 Planetary wave propagation

Apart from the existence of free travelling planetary waves, most planetary waves in the middle atmosphere appear to propagate upward from the forcing regions in the troposphere. Charney and Drazin (1961) were the first to discuss in detail the vertical propagation of planetary waves. They used a β -plane, linearized, quasi-geostrophic model and assuming the zonal-mean wind changes only with height and that the waves were steady, they found the criterion for propagation:

$$0 < \bar{u} - c < \bar{u}_c \equiv \beta(k^2 + l^2 + f_o^2/4H^2N^2)^{-1} \tag{2.1}$$

where $\bar{u}$ and c are zonal mean wind and phase speed respectively, k, l are wave numbers along longitude and latitude direction, f_o Coriolis parameter and H, scale height; N, buoyancy frequency. For wave with zero phase speed, i.e. stationary with respect to the ground, we have $0 < \bar{u} < \bar{u}_c$. This states that Rossby waves can propagate vertically in a window where the zonal mean wind is westerly (eastward) and not stronger than $\bar{u}_c$. Since $\bar{u}_c$ is approximately inversely proportional to the square of wave number, only those waves with the smallest wave number or largest wave length are able to penetrate to the stratosphere. This agrees broadly with the observation that stationary disturbances tend to be composed predominately of Fourier component of wave number 1, 2 and 3 in the winter westerlies and tend to be absent in the easterlies and in the summer when easterlies prevail.

Dickinson (1968a,b, 1969a) extended Charney and Drazin's model to spherical geometry and to include effects of latitudinal variation of the zonal mean wind. He confirmed the results of Charney and Drazin and showed that the zonal wind variation was of significant importance, apparently giving two types of wave guides located between pole and strong westerly jet and between the jet and the zero wind line (critical line) in subtropical area. He further included the diabatic effect by introducing linear cooling in his model and showed that the cooling results in attenuation of planetary waves in weak westerly flows. Simmons later (1974) found the waves are not necessarily refracted away by the strong jet and the result of Dickinson's is due to his omission of the first and second derivatives of zonal-mean wind with respect to latitude.

Detailed quantitative investigation of vertical propagation of stationary planetary wave in the stratosphere began with the work of Matsuno (1970). He concentrated on the hypothesis that the stationary planetary waves in the stratosphere of Northern Hemisphere winter are forced from the troposphere and therefore imposed 500mb height based on observation as a lower boundary in his linearized model. He used a linearized quasi-geostrophic potential vorticity equation in spherical geometry and modified it to include an ageostrophic term, called an "isallobaric" contribution to the northward disturbance wind. This term ensures energetic consistency and permitted creation of the generalized Eliassen-Palm flux theorem in

the form which will be discussed further below (near equation 2.11). Taking a basic flow varying both with latitude and height and using certain boundary conditions, he solved the equation for stationary wave amplitude and phase numerically. By analogy with the theory of acoustics or optics he introduced a "refractive index" n_s, defined by:

$$n_s^2 = \frac{\bar{q}_\phi}{a\bar{u}} - \frac{s^2}{a^2\cos^2\phi} - \frac{f^2}{4N^2H^2} \tag{2.2}$$

Wave propagation is favoured where n_s^2 is large and positive, and is inhibited where n_s^2 is negative. This refractive index is similar to that of Dickinson (1968b) but with definition of the gradient of quasi-geostrophic potential vorticity, $\bar{q}_\phi$ being different. The results from Matsuno's model are in quite good agreement with observations though the computed wave-2 geopotential amplitude decays more rapidly with height than observed, possibly because his $\bar{u}$ fields were from a different occassion than his wave observations.

Matsuno also used a diagnostic of wave-energy flux in the meridional plane $(\rho_o\overline{v'\Phi'}, \rho_o\overline{w'\Phi'})$, which for stationary, conservative, linear waves equals to the product of zonal mean wind and Eliassen-Palm flux. The pattern of this vector from his calculation shows characteristic upward and equatorward orientation (see figure 2.1) in the observed planetary wave Eliassen-Palm flux.

Later workers (e.g. Simmons, 1974; Schoeberl and Geller, 1977; Lin, 1982) using

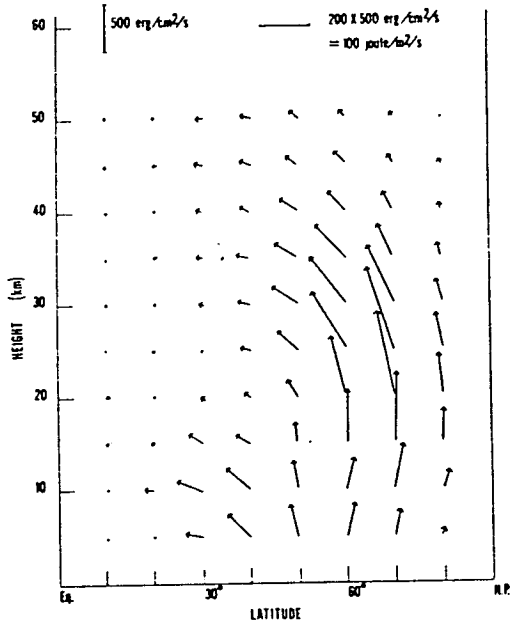


Figure 2.1: Distribution of the wave-energy flow in meridional plane associated with the wave 1. [From Matsuno 1970]

various and more sophisticated models, for instance including friction and diabatic heating etc., confirmed Matsuno's results and found that the vertical propagation of planetary waves into the stratosphere is sensitive to the zonal-mean wind structure, especially its second derivative with respect to latitude and the position of latitude of polar night jet. They also showed that only wave 1 and 2 components can propagate into the stratosphere, in qualitative agreement with the Charney and Drazin criterion, and that the inclusion of thermal forcing in their models rather than topographic effect alone degrades the agreement between model result and observation. Nonlinear wave effects may play a role in the process and this remain to be investigated.

2.1.2 Planetary wave breaking

A substantial contribution in the last decade to the planetary wave theory was the identification of planetary wave breaking phenomena from stratospheric satellite data by McIntyre and Palmer (1983,1984). In their case study, they used isentropic maps of Ertel potential vorticity together with isobaric charts of geopotential height from January and February 1979 to argue and identify the planetary wave breaking phenomenon and discuss its dynamics. The concept of wave breaking is defined as irreversible deformation of material contours and surfaces, rather than a simple undulation back and forth as linear theory assumed. The Ertel potential

vorticity, as will be discussed later, is quasi-material (or conservative) in isentropic surface and thus was used by the two authors to visualize the breaking process.

Three principal figures from McIntyre and Palmer's study are showed in figure 2.2. Figure 2.2a is the geopotential height on 10mb on 26 January 1979. In correspondence to the strong off-pole cyclonic vortex over Scandinavia, a high potential vorticity area (as shaded) (called by McIntyre and Palmer the "main vortex") appeared in figure 2.2b with a region of substantial cross-stream gradients of potential vorticity. As the dynamical restoring mechanism, whereby waves can propagate westward relative to the stream and remain stationary in the polar vortex, depends on the existence of a cross-stream gradient of Ertel potential vorticity in the isentropic surface (part of this gradient is due to beta-effect) in the linear theory, the appearance of this structure ('main vortex') is able to support the Rossby wave propagation.

Outside the heavily shaded region, the gradient is very weak, less wave-like and highly nonlinear. The later map (figure 2.2c) shows the long thin tongue of high potential vorticity has been pulled out from the 'main vortex' and been mixed quasi-horizontally and irreversibly into the surrounding region of weaker gradients. By analogy to wave-breaking in a beach, McIntyre and Palmer chose to call the outside region as 'surf-zone'. Along with the pulling of the tongue of higher potential vorticity from the main vortex, the lower potential vorticity from the

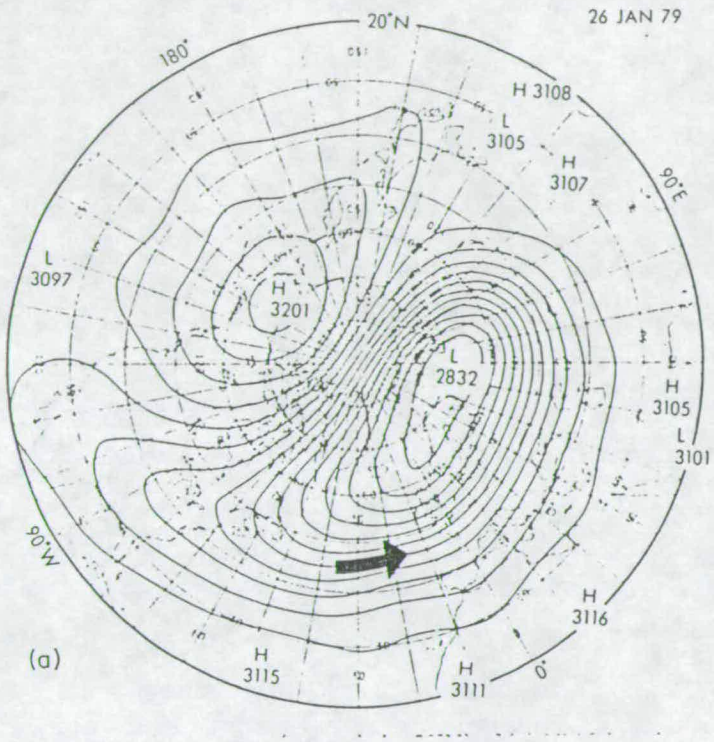


Figure 2.2: Polar stereographic maps (outer circle: $20°N$) of (a) NMC-based analysis of the 10mb geopotential height (decameters) on January 26, 1979. (b) Coarse-grain estimate of Ertel's potential vorticity divided by gH/p_0 (where H=6.5km and $p_0 = 1000$mb) on the 850-K isentropic surface on January 26, 1979, in $10^{-4} K m^{-1} s^{-1}$. Values greater than 6 units are heavily shaded. The dashed circle shows the position of a local maximum of just 4 units. (c) As for (b), but for January 27,1979. [From McIntyre and Palmer (1984)]

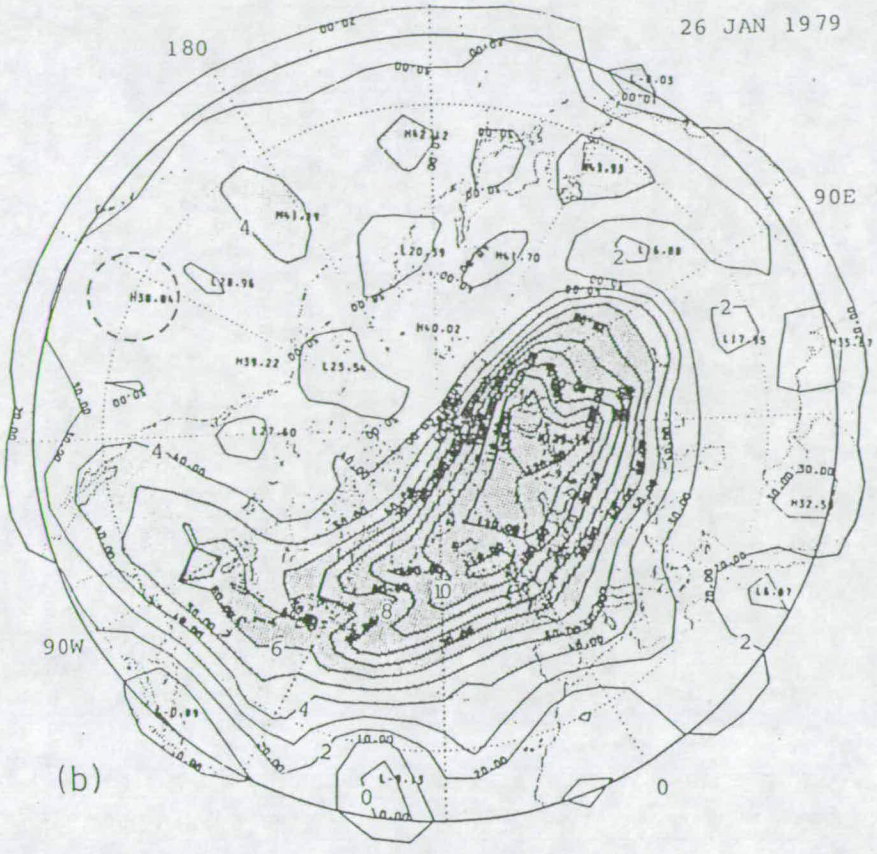


Fig. 2.1 *figure continues)*

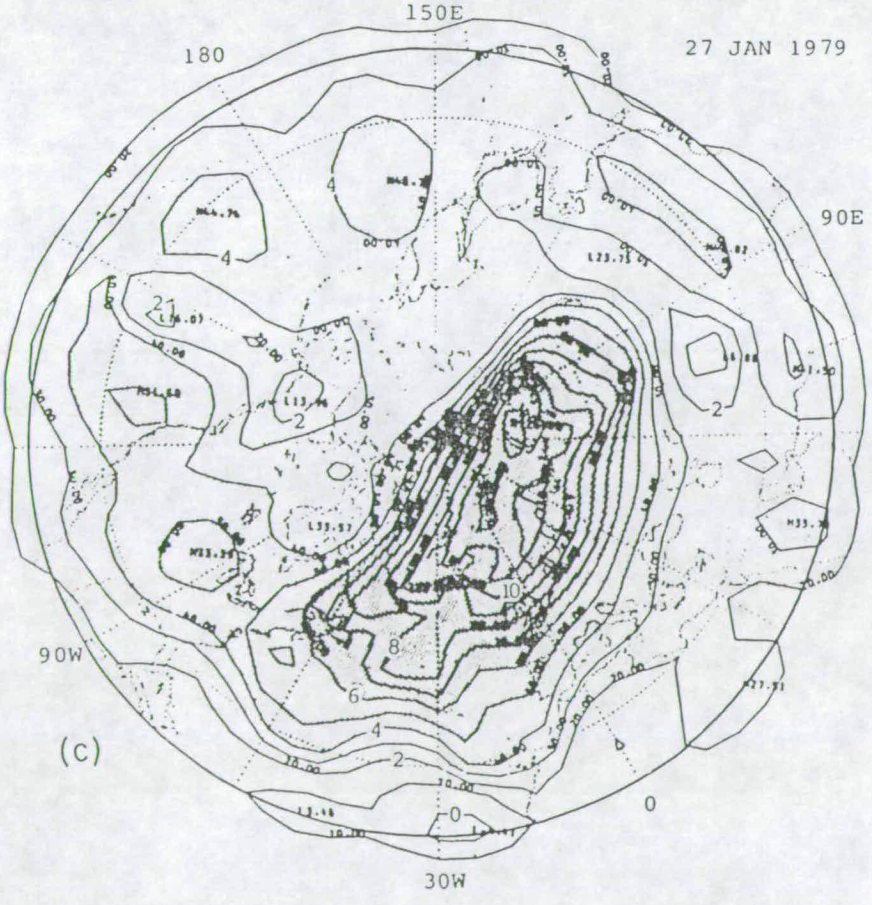


Fig2 . 1*(continued)*

lower latitude is also driven into the polar area, eating its way into the potential vorticity gradient at the edge of the main vortex, and systematically reducing the area best capable of supporting Rossby-wave propagation. This process is called erosion. The breaking planetary wave hypothesis was later supported by Leovy et al. (1985) (see also Clough et al. 1985) by studying the transport of ozone in the stratosphere where maps of ozone mixing ratio on 850K isentropic surface display the signature of similar breaking wave phenomenon though potential vorticity and ozone is negatively corelated. These collaborative works laid the foundation for firm recognition of planetary wave breaking events and its importance in the stratospheric dynamics.

In addition, since the net transport of tracers is closely related with the breaking of waves (as simple undulation of material lines causes no net transports), the understanding of planetary wave processes also have significance in transport studies. This will be discussed further in the last section of this chapter and in chapter 7.

As planetary wave breaking is a highly transient and nonlinear process, according to the generalized Eliassen-Palm flux theorem (see section 2.4), it also has great impacts on the zonal mean such as the dramatic changes of zonal mean flow during stratospheric warmings. This will be discussed in section 5 of this chapter.

2.2 The isentropic coordinate and Ertel potential vorticity

One of the features of the stratospheric dynamical study in the last decade is the emphasis and trend of use of isentropic surfaces ($\theta - surfaces$) as vertical coordinate. Three reasons form the foundation of this trend. The first and overall the most obvious and advantageous one is that the "isentropic vertical velocity", $D\theta/Dt$, equals to the diabatic heating rate Q and thus the vertical component of meridional circulation depends solely upon diabatic heating. The meridional circulation in θ-coordinates is called the "diabatic circulation" and considered important in understanding of meridional tracer transport (Dunkerton,1978;Andrews,1986).

The second reason is that the isentropic coordinates have some of the advantages of generalized Lagrangian mean formalism without many of accompanying technical problems. For example for motions that are adiabatic, Q=0 and the vertical velocity vanishes, so that there is no motion across isentropic surfaces and the fluid parcels must remain on constant potential temperature surfaces. The problems of a 3-dimensional fluid trajectory is reduced to 2-dimensional in this case.

Many physical mechanisms will become much clearer if discussed in isentropic

instead of other coordinates. For instance, the phenomena of planetary wave breaking discussed in previous section, the particular property of Ertel potential vorticity found by Haynes and McIntyre (1987) that there can be no net transport of PV across any isentropic surface and that PV can neither be created nor destroyed within a layer bounded by two isentropic surfaces, all proceed in isentropic coordinates. Also the Charney-Drazin non-acceleration theorem is a finite amplitude result in this coordinate (Andrews, 1983; Tung, 1986). These advantages give the third reason for the now popular adoption of isentropic coordinates.

The main foundation underlying the use of IPV maps to study the dynamical processes in the atmosphere are reviewed by Hoskins et al. (1985) as two principles, namely the Lagrangian conservation principle and invertibility principle. The long history of concept of potential vorticity and its development are also reviewed there. The brief description of these principles and their relation to the present study are presented in chapter 4 and 6 below and will not be addressed in any detail here.

2.3 The Eliassen-Palm flux

In the past decade, the improvement of our understanding of stratospheric dynamics has benefited remarkably from the introduction and application of the

dynamical concept of Eliassen-Palm flux and its divergence. It is broadly used in diagnostic studies, (e.g. Palmer, 1982; O'Neill and Youngblut, 1982; Clough et al., 1985), modeling studies, comparison of the two hemispheres (Al-Ajmi et al., 1985) and statistical study (Randel, 1987) and plays an important role in our understanding of wave-mean interactions.

The appearance of the Eliassen-Palm flux in the literature can be traced to as early as 1960 though its full significance was not immediately recognized. In their classical study of vertical wave propagation in the zonal mean flow, Eliassen and Palm (1960) investigated the possibility that the energy of planetary-scale disturbances of the troposphere can propagate into the stratosphere. They showed that poleward eddy momentum flux implies equatorward flux of geopotential and that poleward eddy heat flux implies upward geopotential flux when the zonal mean wind is westerly. Based on this work, Andrews and McIntyre (1976, 1978) have extended the non-acceleration theorem proposed by Charney and Drazin (1961) and introduced the generalized Eliassen-Palm relation for acceleration of zonal mean flows. These theoretical advances are reviewed by Holton (1980). Upon these theoretical advances, it is the work of Edmon et al. (1980) that summarized the presently well-accepted concept of Eliassen-Palm flux (E-P flux hereafter) and emphasized it as a diagnostic tool for the interaction of wave with the mean flow.

The meteorological significance and physical advantages of E-P flux concept and its divergence as diagnostic tools relate to the concept of residual mean meridional circulation and to the transformed Eulerian-mean equations introduced by Andrews and McIntyre (1976a,1978a) and Boyd (1976). Alternative definitions of the residual circulation are given by Andrews and McIntyre (1978b) and Holton (1981). The whole set of these diagnostics arise out of the zonal momentum and thermodynamic equations. For $\beta - plane$ geometry, under quasi-geostrophic approximation, a residual circulation, $(\bar{v}^*, \bar{w}^*)$, can be defined by (see e.g. Andrews, 1986),

$$\bar{v}^* \equiv \bar{v}_a - \rho_o^{-1} \left(\frac{\rho_o \overline{v'\theta'}}{\theta_{oz}} \right)_z \tag{2.3}$$

$$\bar{w}^* \equiv \bar{w}_a + \left(\frac{\overline{v'\theta'}}{\theta_{oz}} \right)_y \tag{2.4}$$

where the $\bar{v}_a, \bar{w}_a$ are ageostrophic parts of the ordinary Eulerian mean components of meridional circulation; θ_o is a reference potential temperature; $\rho_o = \rho_o(z) \equiv \rho_s e^{-z/H}$ is the basic density with $\rho_s \equiv P_s/RT_s$ the density in the surface, $H = RT_s/g_0$ the scale height.

The quasi-geostrophic transformed Eulerian momentum and thermodynamic equations can be written:

$$\bar{u}_t - f_o \bar{v}^* = \rho_o^{-1} \nabla \cdot F \tag{2.5}$$

$$\bar{\theta}_t + \bar{w}^* \theta_{oz} = Q \tag{2.6}$$

The E-P flux vector in this quasi-geostrophic $\beta - plane$ case is,

$$F \equiv (0, -\rho_o \overline{v'u'}, \rho_o f_o \overline{v'\theta'}/\theta_{oz}) \tag{2.7}$$

and its divergence is given as:

$$\nabla \cdot F \equiv -(\rho_o \overline{v'u'})_y + (\rho_o f_o \overline{v'\theta'}/\theta_{oz})_z \tag{2.8}$$

where the subscript y and z denote differentiation with respect to that variable and overbar and prime denote the zonal mean and the departure therefrom. The above equations 2.3–2.8 are in log-pressure coordinates.

The key point of transformed Eulerian mean formalism over ordinary Eulerian mean equations is that the rectified eddy-forcing term on the right of equation 2.5 i.e. E-P flux divergence, depends explicitly on certain basic physical properties of wave or eddy disturbances; whereas in the Eulerian-mean the relation between the physical properties of eddies and the the zonal mean is not so obvious (as will be discussed below). For example, from the generalized E-P flux theorem (to be discussed in next section), it can be shown that if the disturbances are steady, linear, frictionless and adiabatic and if the mean flow is conservative to $0(\alpha^2)$, where α is the amplitude of the waves, the total eddy force of mean flow will disappear. By contrast the Eulerian-mean momentum and thermodynamic equations on $\beta - plane$, for quasi-geostrophic, frictionless flow can be written:

$$\bar{u}_t - f_o \bar{v}_a = -(\overline{v'u'})_y \tag{2.9}$$

$$\bar{\theta}_t + \bar{w}_a \theta_{oz} = \bar{Q} - (\overline{v'\theta'})_y \tag{2.10}$$

Under the same conditions discussed above, the forcing on the zonal mean flow by eddies does not vanish when $\nabla \cdot F = 0$, but induces a non-zero Eulerian-mean circulation $(\bar{v}_a, \bar{w}_a)$ that precisely cancels their effect. Generally one can not easily anticipate how the zonal-mean flow will respond to a specified "eddy momentum flux" $\overline{u'v'}$ or "eddy heat flux" $\overline{v'\theta'}$ in the ordinary Eulerian mean equations 2.9 and 2.10 or turn to anticipate what physical properties of the waves control these eddy fluxes.

The advantages of using the E-P flux and its divergence together with transformed Eulerian mean equations in dynamical studies have been discussed in detail by Edmon et al.(1980), Dunkerton et al. (1981), Palmer (1982) and O'Neill and Youngblut (1982), They can be summarized as follows:

(1) The effects of the eddy momentum and heat flux on the mean flow become a single term, the E-P flux divergence, in the transformed Eulerian mean formalism. This gives succinct theoretical description of wave-mean flow interaction.

(2) E-P flux is non-divergent under steady, linear, conservative conditions. The non-divergence of E-P flux is a necessary condition of non-acceleration of mean flow (Andrews, 1986; Tung, 1986) and E-P flux can be considered a measure of net wave propagation from one height and latitude to another (Edmon et al.,1980;

McIntyre, 1982).

(3) The E-P vector indicates the relative importance of the magnitude and direction of eddy heat and momentum flux.

(4) The divergence of E-P flux for all eddies reflects the magnitude of a combination of transience and irreversible eddy processes at each height and latitude and is found to be proportional to the northward eddy flux of quasi-geostrophic potential vorticity (Edmon et al., 1980, Palmer, 1982; Andrews,1987).

(5) E-P flux vector is parallel to the projection of the group velocity onto the vertical-meridional plane when the planetary wave on a slowly varying zonal flow satisfies the WKBJ approximations and when only a single wave packet is present.

As mentioned before, the transformed Eulerian mean equations and E-P flux definition showed above are all in log-pressure coordinates for conciseness and on a $\beta - plane$ for simplicity. Our later study, however, will concentrate on the isentropic coordinate. All the above statements about E-P flux and related concepts are correct in isentropic coordinates on a sphere except that the residual circulation will be replaced by diabatic circulation and all the equations and definitions have to be written accordingly. The whole set in spherical isentropic coordinates will be given in chapter 5 and 6.

Some difficulties of applying E-P flux diagnostics have also been noticed by several workers. These mainly include errors caused by either the adoption of quasi-geostrophic rather than primitive form of E-P flux definition or errors caused by wind estimation (Robinson, 1986; Boville 1986) which lead to a probably spurious 'dipole structure' of high latitude divergence and mid-latitude convergence and to the appearance of positive region of E-P flux divergence in long term averaged analysis (monthly or seasonal mean) (see Hartmann,1984; Andrews,1987). The latter is partly explained theoretically by Andrews (1987) and detailed dynamics of it remains to be investigated.

2.4 Generalized E-P flux theorem

As mentioned in the previous section the divergence of the E-P flux depends explicitly on certain basic physical properties of the flow. This is one of the conclusions of the generalized E-P flux theorem, as ~~to be~~ seen below, and is the main reason for adoption of transformed Eulerian mean formalism in preference to the ordinary Eulerian mean equations.

The foundations for the generalized E-P flux theorem is also laid on the pioneering work of Eliassen and Palm (1960) and Charney and Drazin (1961). The former authors considered steady, linear waves on a basic flow $\bar{u}(\phi, z)$, with no frictional

and diabatic effects. Using a set of linear disturbance equations, they were able to show that the E-P flux divergence (later so called) vanishes for linear, steady, conservative waves on a purely zonal basic flow. This result was generalized to include the nonconservative effect and wave transience and to allow for spherical geometry by the later authors (Andrews and McIntyre 1976a, 1978a; Boyd 1976) giving the now well recognized generalized E-P flux theorem, which takes the form:

$$\frac{\partial A}{\partial t} + \nabla \cdot F = D + 0(\alpha^3) \tag{2.11}$$

where A and D, like F are mean quadratic functions of disturbance quantities. "A" appearing in 2.11 is called the "wave-activity density". The quantity D contains the frictional (X', Y') and diabatic (Q') effects. The $0(\alpha^3)$ term, where α is wave amplitude, represents nonlinear wave effects and negligible for sufficiently small wave amplitude. As the term $\partial A/\partial t$ vanishes for steady waves and Q vanishes for conservative waves, Eliassen and Palm's result that F is non-divergent for such waves emerges as a special case of equation 2.11.

The importance of generalized E-P flux theorem is that it shows explicitly the dependence of E-P flux divergence on the physical properties of the wave, viz. transience and nonconservative effects. The demonstration of the so-called non-acceleration theorem by generalized E-P flux theorem is straight forward, i.e. when A_t and D are zero so is $\nabla \cdot F$ and it follows from equation 2.5 and 2.6 that $\bar{u}_t = 0$. It shows how the waves induce no mean-flow changes under the stated condition. Such

a result is not obvious at all from the ordinary Eulerian-mean formalism, where such eddies will induce non-zero Eulerian-mean circulation $(\bar{v}_a, \bar{w}_a)$ that precisely cancel their effects.

Another significance of generalized E-P flux theorem is that it fundamentally, when the terms in the right hand side of equation 2.11 are zero, takes the form of a conservative law for wave properties. Such a law is of considerable interest in many branches of physics (McIntyre, 1980, 1981). The advantages of generalized E-P flux theorem over wave-energy equation is discussed by Andrews (1986). Briefly, in the presence of mean flow shear, terms that represent the exchange of energy between mean-flow and disturbance in the wave-energy equation are generally nonzero. Those terms do not appear in the generalized E-P flux theorem and thus the theorem can be used as a diagnostic of wave propagation in complicated mean flow.

One of the main difficulties which prevent the early application of generalized E-P flux theorem to observational studies is that the original fundamental definitions of quantities A and D involve parcel displacements from some undisturbed basic or 'initial' state. For example the quasi-geostrophic wave activity can be defined as $A = \frac{1}{2}\rho_o \overline{\eta'^2} \bar{q}_y$, where q is quasi-geostrophic potential vorticity, η' is northward parcel displacement. Finding the parcel displacement by constructing trajectories is practically extremely difficult. Moreover, for the real atmosphere, as opposed

to that of an idealised mathematical model, it is far from clear what should be regarded as the basic state. In the hope to overcome these difficulties, Andrews (1987) gave the Eulerian definitions of these quantities in various coordinates and for different approximations. These definitions, which do not need the knowledge of parcel displacements, paved the way for possibility of direct applying the E-P flux theorem in the diagnostic studies of stratospheric wave activities using satellite derived data. The main part of the present study is to apply the version of the E-P flux theorem given by Andrews (1987) for a diagnostic study of wave activities using LIMS data in a period in which the small amplitude condition required by the theorem is satisfied in the real stratospheric atmosphere. Also based on this version of the theorem, the wave-mean flow interactions, the limitation of linear theory and the relations between chemical and dynamical tracers as well as possibility of parameterizing tracer transport effects in the view of the theorem are also investigated. The detailed expression of the Eulerian definitions used in the present study will be quoted in detail in chapter 5.

2.5 Understanding of stratospheric sudden warming

In the normal northern hemisphere winter, the zonal-mean zonal winds in the stratosphere are generally westerly and increase with height, peaking in the polar night

jet, and the temperature decreases toward the winter pole. The stratospheric sudden warming is a phenomenon in which the normal zonal-mean feature is dramatically disrupted, with polar temperatures increasing rapidly with time, leading to a poleward increase of zonal mean temperature and sometimes a reversal of zonal-mean winds to be easterly. It is such a spectacular phenomenon in the stratosphere that it has been studied by various means and throughout the decades since it was discovered by Scherhag (1952). We only briefly present the theoretical understanding of it in this section from the view of models and wave-mean flow interaction.

Matsuno (1971) was the first to suggest the dynamical mechanism underlying the sudden warming in terms of the upward propagation from the troposphere of planetary (Rossby) waves and their interaction with the zonal mean stratospheric flow. In a test of this idea using a numerical model, he supposed that the mean flow changes observed during sudden warnings, including zonal flow deceleration and mean temperature rise near the pole, are due to the rectified nonlinear effects of vertically propagating planetary waves forced in the troposphere by large scale disturbances there. Bearing in mind that steady, conservative, linear planetary waves do not induce zonal mean change by Charney-Drazin non-acceleration theorem, he sought conditions violating the theorem in his model. One of these conditions occurs when a transient packet of planetary waves first starts to propagate up-

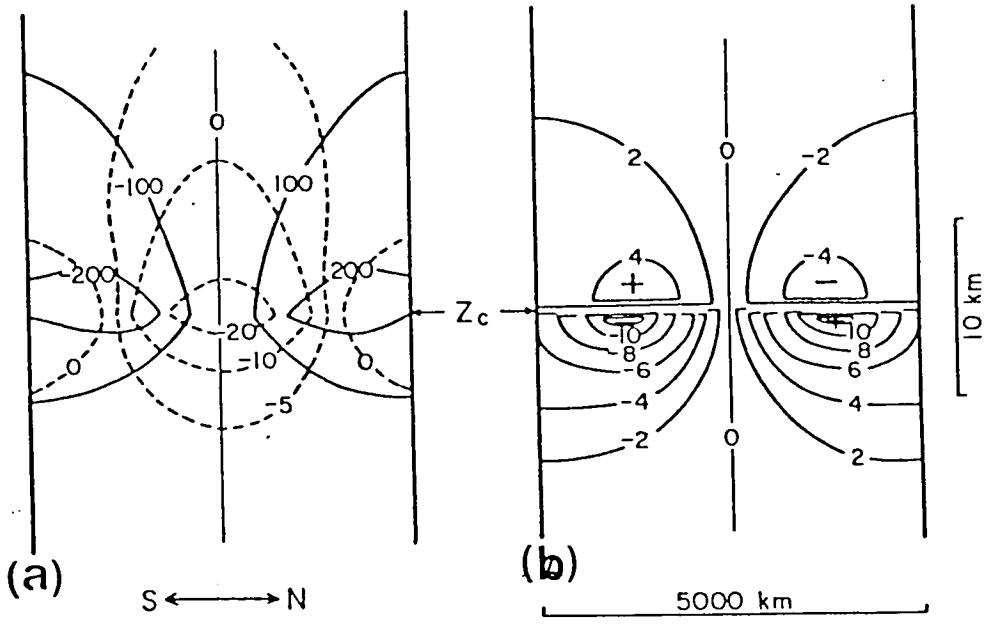


Figure 2.3: The changes of zonal mean fields caused by planetary waves incident on a critical level (Z_c) as shown in the meridional plane. (a) Time change of isobaric heights (solid line, m/day), and that of zonal winds (dashed lines, m/s.day). (b) time change of temperature ($°C/day$). [From Matsuno, 1971]

---

ward, wave amplitude are growing with time at the leading edge of such a packet. Matsuno was able to show that these amplifying waves can cause a deceleration of the zonal mean wind and an increase of zonal mean temperature poleward of the latitude where the height wave amplitude is maximum and a decrease on the equatorward side as shown in figure 2.3. Matsuno indicated that the magnitude of deceleration should increase with height because of increasing of wave

amplitude with the decrease of air density. As a result the mean zonal flow may become easterly at high levels of the stratosphere, forming a critical level which prohibits further vertical propagation. Matsuno's mechanism and model results were confirmed by later authors (Geisler,1974; Holton, 1976; Schoeberl,1980). Actually most subsequent 'mechanistic models' of sudden warmings have been the generalization of Matsuno's.

From the view of the generalized E-P flux theorem, Matsuno's mechanism can be schematically described as follows. As shown in previous section the quasi-geostrophic wave activity can be defined as $A = \frac{1}{2}\rho_o \overline{\eta'^2} \bar{q}_y$, if $\bar{q}_y > 0$, for growing wave at the leading edge of packet $\partial A/\partial t > 0$ as the particle dispersion grows. From generalized E-P flux theorem, E-P flux divergence $\nabla \cdot F < 0$ for quasi-conservative waves. This induces a negative force per unit mass on the zonal-mean flow by equation 2.5 and causes a deceleration in the vicinity of the patch of negative $\nabla \cdot F$. A mean temperature increase will occur on the lower poleward flank of this patch in consistent with thermal wind balance. The whole picture is showed schematically in figure 2.4 and the mean-flow changes with the leading edge of transient planetary wave packet are qualitatively similar to those observed during a sudden warming (Andrews et al. 1987).

The above view from the generalized E-P flux theorem is confirmed by the model study of Dunkerton et al. (1981) using transformed Eulerian-mean equations.

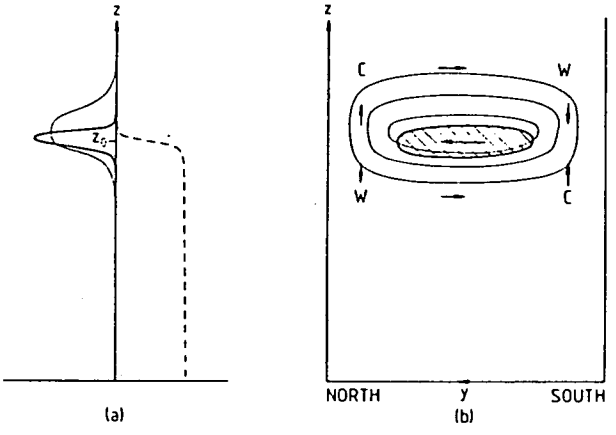


Figure 2.4: Schematic description of the interaction with the zonal mean flow of a transient vertically propagating planetary-wave packet that is maintained by continual forcing at z=0 in a zonal channel, to illustrate aspects of Matsuno's model of the sudden warming. (a) Mid-channel height profiles of wave activity A (dashed), E-P flux divergence $\nabla \cdot F = -\frac{\partial A}{\partial t}$ (heavy line), and acceleration $\bar{u}_t$ (thin line); z_0 is the height reached by the leading edge of the packet, and is of order (vertical group velocity) $\times$ (time since packet started at z=0). (b) Latitude-height cross section showing region where $\nabla \cdot F < 0$ (hatched), contours of induced acceleration $\bar{u}_t$ (thin lines), and induced residual circulation $(\bar{v}^*, \bar{w}^*)$ (arrows). Regions of warming (W) and calling (C) are also indicated. [From Andrews 1986]

They showed that the vertically forced planetary waves exhibited significant negative values of $\nabla \cdot F$ due primarily to wave transience, the associated zonal force per unit mass $D_F \equiv (\rho_o \cos\phi)^{-1} \nabla \cdot F$ was particularly large in magnitude in the high-latitude polar cap where both ρ_o and $\cos\phi$ were small. The effect of this force was redistributed to some extent by the residual circulation, but nevertheless produced deceleration of zonal mean winds and associated increasing of polar temperature. The diabatic effects were small in the short time scale here and the net deceleration of zonal mean wind is mainly affected by D_F as figure 2.5 shows.

In this section, we mainly concentrated on the role that wave-mean flow interaction plays in the sudden warmings, though other mechanisms, such as wave-wave interaction and instabilities may be also involved (e.g. Hsu, 1981; Palmer, 1981; Palmer and Hsu, 1983; Matsuno, 1984), the reason being that this mechanism and the generalized E-P flux theorem are closely related to our later study in chapter 5. However there are difficulties in extending our study to these large amplitude events, as discussed in that chapter.

2.6 Tracer transport theory

As the chapter 6 of this thesis will involve the investigation of tracer transport in meridional plane from an extended view of generalized E-P flux theorem, the

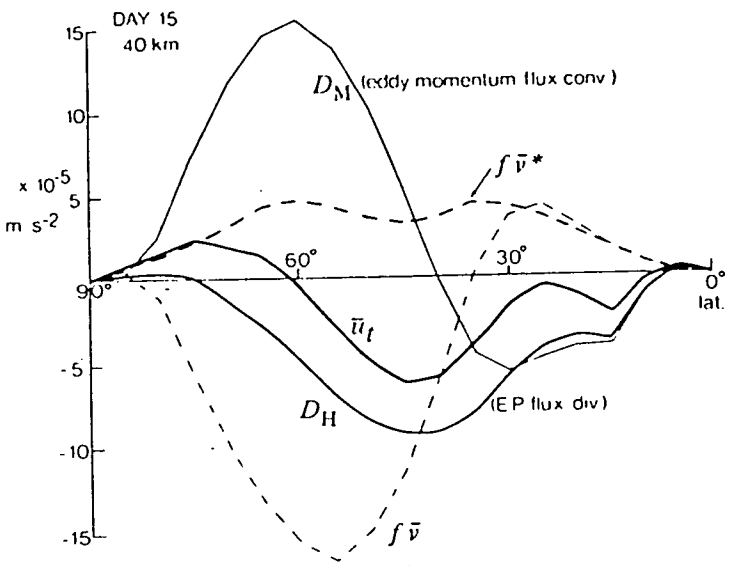


Figure 2.5: Heavy curves: typical balance of terms in the zonal momentum equation of the transformed Eulerian-mean equations, from the model warming described by Hsu (1980). Light curves: principle terms in the conventional momentum equation, where $D_m = -(a\cos^2\phi)^{-1}(\overline{v'u'}\cos^2\phi)_\phi$. [From Dunkerton et al (1981)]

tracer transport theory will be briefly presented in this section.

An atmospheric tracer is defined as any quantity that "labels" fluid parcels. It may be dynamical or chemical, conservative or quasi-conservative, passive or active. The transport of a tracer means the movement of the tracer from one place to another by the atmospheric flow field, thus the transport problem is in principle a matter of tracking parcel movement. Since a trace-gas field will be mainly determined by chemical sources and sinks if the chemical time scale is much less than the dynamical time scale, transport effects are usually considered for conservative and quasi-conservative tracers or for tracers whose chemical time scale is greater or comparable with the dynamical time scale.

As for the parameterization of transport effects in models, since in three dimensional cases, the wind components are calculated explicitly and the flux term in the tracer continuity equation can be directly found and the distribution of the tracer straight forwardly determined, the need for parameterization of the transport of tracers in meridional plane only arises in two dimensional zonal-mean models and in one-dimensional models. The present study will deal only with the parameterization of effects of transport in meridional plane results (in chapter 6).

The flux form of tracer continuity equation in isentropic coordinates (for tracer

with mixing ratio χ) can be written as:

$$(\sigma\chi)_t + (\sigma u\chi)_x + (\sigma v\chi)_y + (\sigma Q\chi)_\theta = (\sigma S) \tag{2.12}$$

where σ is density.

Averaging zonally on a θ-surface and defining a mass-weighted mean for any field B:

$$\bar{B}^* \equiv (\overline{\sigma B})/\bar{\sigma} \tag{2.13}$$

and taking into account the mean continuity equation,

$$\bar{\sigma}_t + (\bar{\sigma}\bar{v}^*)_y + (\bar{\sigma}\bar{Q}^*)_\theta = 0 \tag{2.14}$$

we obtain the mean transport equation:

$$\bar{\chi}_t + \bar{v}^*\bar{\chi}_y + \bar{Q}^*\bar{\chi}_\theta = \bar{S}^* - \bar{\sigma}^{-1}(\overline{\sigma'\chi'})_t - \bar{\sigma}^{-1}\nabla\cdot\left(\overline{(\sigma V)'\chi'}\right) \tag{2.15}$$

where $\nabla = \left(\frac{\partial}{\partial y}, \frac{\partial}{\partial\theta}\right)$, $V = (v, Q)$ and $Q = \frac{D\theta}{Dt}$.

From equation 2.15 we can see that the transport consist of two processes: that by mean meridional circulation $(\bar{v}^*, \bar{Q}^*)$ (i.e. the mean advection) and that by effect of eddies. However these two processes are not independent. Holton (1986) gave a schematic picture, reproduced as figure 2.6, of the relationship between the two processes. He indicated that the equilibrium tracer slope results from the competition between the slope steepening effects of advection by the diabatic circulation and the slope flattening effects of quasi-isentropic eddy transport and photochemical loss. On the other hand, the diabatic circulation itself is primarily

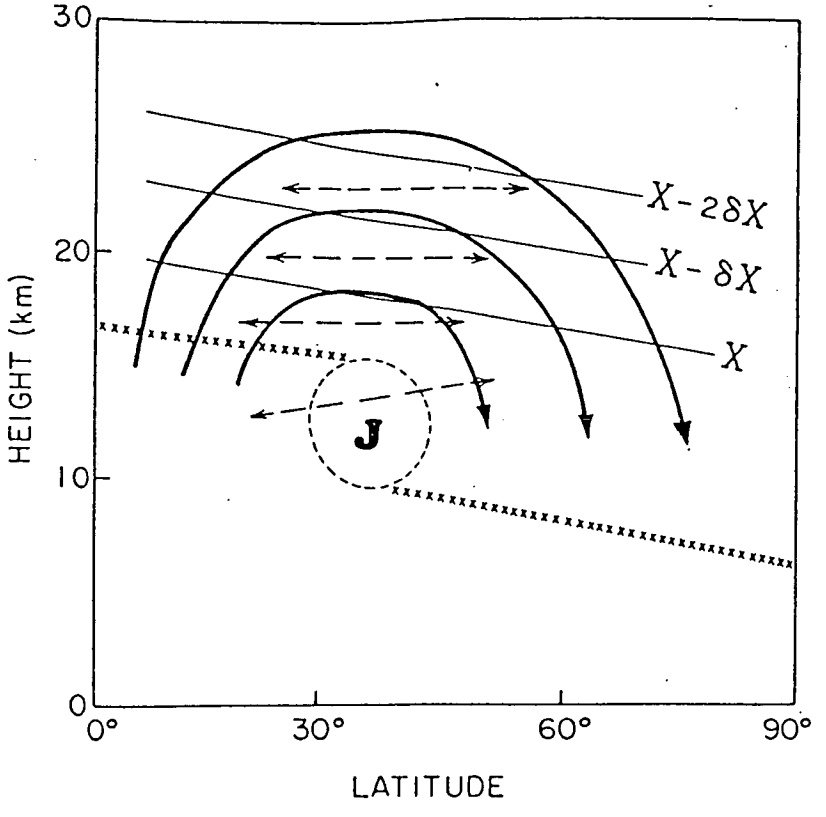


Figure 2.6: A schematic meridional cross section illustrating transport in the stratosphere. Heavy lines with arrows show the streamlines of the diabatic circulation. Dashed arrows indicate quasi-isentropic eddy transport. The mean tropopause is indicated by crosses and the J indicates the mean location of the subtropical tropospheric jet stream. Light lines labeled with mixing ratio values (X) show the mean slope of a typical vertically stratified tracer. [From Holton, 1986]

driven by eddy transports, which maintains the departure of temperature from the state otherwise close to radiative equilibrium with very weak meridional motions and tracer transport. Thus these two processes often cancel each other to some extent. This point can also be understood from equation 2.15. For example when averaged over a few weeks, the tendency term $\bar{\chi}_t$ is small if the source $\bar{S}^*$ is small, in which case the 'mean advection' and 'eddy' terms in equation 2.15 must almost balance. Generally for weak eddies the departure from radiative equilibrium will be small, the diabatic circulation will be weak and the tracer mixing ratio slopes should be small. As the magnitude of the eddy transport increases, the diabatic circulation will increase and the tracer slopes will tend to steepen. However, if the eddy transport becomes strong enough to completely destroy the pole to equator temperature gradients, further increase in eddy strength can not increase the diabatic circulation but will tend to flatten the tracer mixing ratio slopes through quasi-isentropic mixing. The determination of net meridional transport due to eddies and the calculation of $\bar{\chi}_t$, therefore, needs careful considerations. It is generally agreed that this net transport by eddies depends on the departure from the conditions required by the non-transport theorem, which states that steady, linear and conservative waves produce no net transport.

Knowing the underlying mechanism of the transport processes, the next step is to parameterize the eddy flux terms on the right of the equation 2.15 to allow it to

be used for 2-dimensional models. Assuming the wave amplitude is small and the flux-gradient relation:

$$\nabla \cdot \left(\overline{(\sigma V)'\chi'}\right) \equiv -(K^{(c)} + K^{(s)})\nabla\bar{\chi}, \tag{2.16}$$

we can then obtain the parameterized transport equation (See Andrews et al. 1986):

$$\bar{\chi}_t + \bar{v}^*\bar{\chi}_y + \bar{Q}^*\bar{\chi}_\theta = \bar{S}^* - \bar{\sigma}^{-1}(\overline{\bar{\sigma}\eta'\gamma'})_{ty} + \bar{\sigma}^{-1}\nabla \cdot (\bar{\sigma}(K^{(c)} + K^{(s)})\nabla\bar{\chi}) \tag{2.17}$$

where $K^{(s)}$ is the symmetric 'diffusion' tensor with definition to be given in chapter 6 and $K^{(c)}$ is the 'chemical diffusion' tensor:

$$K^{(c)} = \tilde{A}\begin{pmatrix} \overline{\eta'^2} & \overline{\eta'q'} \\ \overline{\eta'q'} & \overline{q'^2} \end{pmatrix}. \tag{2.18}$$

where $\tilde{A}$ is a linear relaxation coefficient so that $S' = \tilde{A}\chi'$.

Under the assumption of small amplitude eddies, the principal component of $K^{(s)}$ namely $K_{yy} = \frac{1}{2}(\overline{\eta'^2})_t$ represents the reversible dispersion of fluid parcels and can be positive or negative depending on whether the parcels disperse or return to their equilibrium latitude. For large wave amplitude, a parameterized equation of the form of 2.17 is more difficult to justify rigorously. Especially in those events like planetary wave breaking which involve rapid, irreversible deformation of otherwise wavy material contours, nonlinear processes normally prevail, so that the parcels will not return to their original latitude and parameterization based on small amplitude theory would be invalid. Nevertheless, the numerical model experiments

by Plumb and Mahlman (1987) suggested that two dimensional models could still reproduce well the zonally averaged evolution of tracers in such irreversible mixing events with the above scheme of parameterization, though the components of $K^{(s)}$ need to be revised. They were able to show that the errors in the 2-dimensional reproduction results partly from the finite model resolution rather than from errors in formulation. This justifies the practical use of parameterization in the study of meridional transport in 2-dimensional models.

This section lays the theoretical understanding of transport mechanism and its formulation. The methods of estimating the diffusion tensors will be reviewed in chapter 6 along with our own investigation from the new point of view described there.

2.7 Summary of the chapter

The above discussions have shown that planetary waves, their propagation and breaking, play a very important role in the stratospheric dynamics and tracer transports. The use of isentropic coordinates and isentropic maps of Ertel potential vorticity can provide us with a powerful tool to understand more clearly many physical mechanisms involved in the dynamical and transport processes. The divergence of the E-P flux, through the transformed Eulerian-mean equations, can

represent the mean effects of the total force exerted on the zonal mean flow by the waves. The generalized E-P flux theorem gives explicitly the dependence of divergence of the E-P flux on the physical properties of the waves, wave transience and non-conservative effects. Moreover, the version of the theorem derived by Andrews (1987) with each term expressed in Eulerian quantities makes it possible to apply in diagnostic studies using satellite observed data, and forms the basis of this study. In later chapters the theoretical background and ideas put forward in this chapter will be adopted for investigation of planetary wave activities, their transport effects and relations between dynamical and chemical tracers as well as wind divergence effects on the area enclosed by contours of potential vorticity on an isentropic surface. Next chapter will start with the investigation of seeking a better approximation to diagnose the stratospheric winds from the satellite derived data.

Chapter 3

Higher Approximation Methods of Estimating Winds from Satellite Observational Data

3.1 Introduction

As the main body of this thesis involves diagnostic quantities in terms of winds and as winds are not readily available, this chapter contains a discussion of attempts I have made to determine the best way of deriving winds from other measurements.

Observational studies of stratospheric dynamics and tracer transport rely heavily on estimates of the wind fields. Measurements of stratospheric wind fields by conventional means such as radio-sonde or rocket-sonde, however, are too sparse for global synoptic analysis to be pursued and the direct measurement of $\vec{v}$

wind fields by satellite is not yet available at present. On the other hand, the geopotential temperature and height are readily derived from satellite detected radiance data. The determination of the wind field in the stratosphere, therefore, has to resort to an indirect estimation from satellite derived temperature or height fields using some assumptions such as geostrophy and hydrostatic balance together with some lower boundary conditions of height or wind.

Geostrophic wind has long been taken for granted in stratospheric dynamical studies as a good approximation to the true winds due to the dominant of large scale motion and small influence by the underlying topography, though there was recognition of an inadequate representation by the geostrophic approximation in the largest scales of stratospheric motion (Dickinson 1968). It was the more recent work by Robinson (1986) that elicited more attention and research into the problems caused by the geostrophic assumption. In his paper Robinson showed that even in a low wave number model, geostrophically derived winds can seriously misrepresent the E-P flux divergence, leading to spurious sources of wave activity in polar latitudes. Boville (1987), after comparison of eddy momentum flux, heat flux, and E-P flux calculated by geostrophic winds and model winds respectively, found that the geostrophic winds , though they produced small errors in eddy heat flux calculations, can induce substantial errors in eddy momentum fluxes and the error in the winter stratosphere was even larger than in the troposphere.

He deduced that higher order approximations would give more accurate results in the stratosphere diagnostic studies. Elson (1986) has inferred ageostrophic wind components from satellite data and was able to show that terms usually neglected in the zonal momentum equation are significant during large amplitude wave events.

For middle and high latitudes, there are several possible methods which may give better approximations than geostrophic, viz. the gradient wind equation, the quasi-geostrophic equation and several balance equations. The aim of present chapter is to compare those different ways of obtaining better estimates of winds and determine a best one for use in the later chapters. The next section will discuss the observational data used and techniques of transforming them into isentropic surfaces. Section 3.3 will list and discuss several schemes of higher approximation.

The fundmental idea of comparing the wind-approximation schemes in this chapter is based on the two sets of tests. For the first test we have LIMS (Limb Infrared Monitor of the Stratosphere) observational data of geopotential height and temperature at p-sufaces and we also have the corresponding FGGE data in certain levels of the analysis wind, which is interpolated to isentropic surfaces from pressure coordinates by Duncan (personal communication). These data allow us first to diagnose winds from LIMS data using every method described in section 3.3 and then compare the derived winds with the FGGE analysis winds, which

are considered as 'true' winds, to determine the errors of each approximation scheme. Since LIMS data and FGGE data are actually two separate data sets and as the FGGE wind analysis may be constrained by information from the height field, a second test using model data is thus necessary and is pursued. In the second test, we have model output fields of both geopotential height and winds, again we diagnose winds from the model output fields of geopotential height and then compare with the 'true' winds, i.e. the model output wind fields, to analyse the errors of each scheme. It is possible to choose an approximation method with the smallest error through the two tests.

Following the above described ideas, section 3.4 will give some computational details of the two sets of the tests. It will show the resultant winds of the two tests and the error analysis. The errors of some higher level diagnostic such as horizontal momentum flux from the different schemes will also be discussed in this section. Finally in section 3.5 we analyse the possible cause of error in each method and give a summary of this chapter with discussion.

It is necessary to note here that Randel (1987) has done a very similar work and gets very similar results to ours though the procedure for solution and data as well as error analysis are completely different. Unfortunately Randel's work did not come to our attention until the summary stage of the present work. Thus all the work described here is done independently and is supplementary to his, rendering

it more robust by demonstrating that his results were not dependent upon the choice of a particular occasion.

3.2 Data and methods of transforming data to isentropic surfaces

The main data sets for the present thesis are derived from satellite observations. These include synoptically mapped LIMS measurements of temperature, geopotential height and constituents such as ozone, water vapour and nitrogen dioxide. The LIMS data are provided to us in Fourier coefficients and ranged from $63°S$ to $84°N$ in latitude and 100mb to 0.05mb in height. The detailed derivation of each field (See Gille and Russell 1984) is beyond our present scope and will not be discussed here. Only the methods of transforming the data sets into the form which is needed by the calculations in this chapter and the following chapters are discussed here.

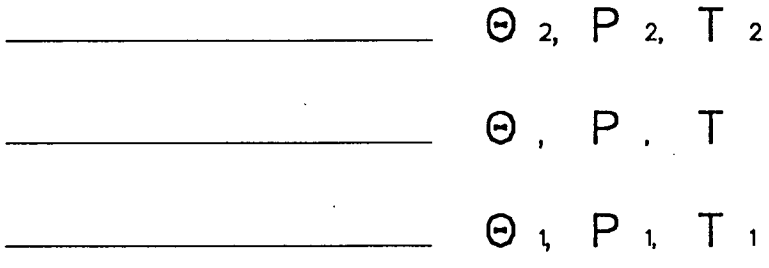
Since the isentropic coordinates have physical and conceptual advantages as discussed in section 2.2, many computations involved in the current thesis are performed on isentropic surfaces. However, all the observational data are on pressure surfaces; a technique is therefore, needed for transforming the data from pres-

sure to isentropic coordinates. A suitable transformation scheme for temperature, Montgomery potential and other constituents from pressure surface to isentropic surface, then, becomes the prior task and will be described in the following.

Practically every field can be numerically interpolated into isentropic surface directly or indirectly from pressure surfaces by various means of interpolation. However, for derived fields such as vorticity and potential vorticity a choice has to be made whether to compute them first on pressure surfaces and then interpolate to isentropic coordinates, or to interpolate the basic fields first and compute the derived fields in the new coordinates. For example in their celebrated paper about planetary wave breaking, McIntyre and Palmer got pressure and vorticity fields on isentropic surface by a cubic spline interpolation of these quantities from pressure coordinates (McIntyre and Palmer 1983). Butchart and Remsberg (1986) also obtained their vorticity fields on isentropic surfaces by cubic spline interpolation of vorticity on pressure surfaces. Though these interpolation may work well for their purpose, we try to calculate our derived fields directly on isentropic surfaces by interpolating the basic fields to isentropic surfaces first. Taking into account the fact that the stratospheric temperature changes approximately linearly with logarithm of pressure, we design an iterative method to get temperature and pressure fields and further, the Montgomery potential field on isentropic surfaces.

Suppose that there are two grid points on pressure surfaces P_1 and P_2 respectively

with the same horizontal location. The original temperature data on these two pressure surfaces are known and the potential temperatures, θ_1 and θ_2, of the two points are readily derived. Now we want to know the values of temperature and pressure of a third point on an isentropic surface θ located vertically between the two grid points and having the same horizontal coordinates as them (see the picture shown below). Assume that temperature changes linearly with logarithm



of pressure within these two pressure levels, i.e.

$$T = a + b\ln(P) \tag{3.19}$$

where a, b are constant. Writing equation (3.19) on the two pressure levels, we can easily have

$$b = \frac{T_2 - T_1}{\ln(P_2/P_1)} = \gamma \tag{3.20}$$

γ is assumed constant between P_1 and P_2 . For an arbitrary pressure surface P between P_1

and P_2, its temperature can be written as

$$T = T_1 + \gamma \ln(P/P_1) \tag{3.21}$$

From the definition of potential temperature, we have

$$P = P_0 \left(\frac{T}{\theta}\right)^{c_p/R} \tag{3.22}$$

where $P_0 = 1000mb$.

For a given value of an isentropic surface θ between (θ_1, P_1) and (θ_2, P_2) equation (3.21) and (3.22) form a closed set of equations and can be solved to obtain pressure P and temperature T on isentropic surface θ respectively.

Replace P in equation (3.21) by (3.22) we have

$$T = T_1 + \gamma \ln(\frac{P_0}{P_1}) - \gamma \ln(\theta)^{c_p/R} + \frac{\gamma c_p}{R} \ln(T) \tag{3.23}$$

or

$$T - \frac{\gamma c_p}{R} \ln(T) - C = 0 \tag{3.24}$$

where $C = T_1 + \gamma \ln(\frac{P_0}{P_1}) - \frac{\gamma c_p}{R} \ln(\theta)$ is known. By Newtonian iteration we can obtain T from equation (3.24) and take it back into equation (3.22) so giving P on a given isentropic surface θ.

Knowing temperature and pressure fields on a given isentropic surface by above discussed procedure, we now turn to discuss how to obtain Montgomery potential

on an isentropic surface. The hydrostatic relationship in isentropic coordinates gives

$$\frac{\partial M}{\partial \theta} = \frac{c_p T}{\theta} = c_p \left(\frac{P}{P_0}\right)^{R/c_p} \tag{3.25}$$

write (3.25) on two isentropic surfaces by central difference scheme and suppose the Montgomery potential fields on θ_1 surface are known, we can calculate the Montgomery potential fields for θ_2 surface by:

$$M_2 = M_1 + c_p T_* \frac{\theta_2 - \theta_1}{\theta_*} = M_1 + c_p T_* \ln\left(\frac{\theta_2}{\theta_1}\right) \tag{3.26}$$

where $\theta_* = (\theta_1 + \theta_2)/2$ is a reference level and T_* is the temperature on θ_*. T_* can be obtained by the similar method to that iteration of equation (3.24).

Obtaining Montgomery potential on the bottom isentropic surface by its definition $M = c_p T + gz$ with T and z directly interpolated from pressure coordinates and integrate up to any given level, we can get Montgomery potential fields on that isentropic surface.

Since this method is based on the definition of potential temperature and the fact that the temperature is linearly proportional to logarithm of pressure, it should produce reliable data on isentropic surfaces. The comparison with results of those by Butchart and Remsberg (1986) in chapter 4 and other calculations later in this chapter all confirm our confidence.

The related chemical elements such as ozone mixing ratio, NO_2 etc. on isentropic

surfaces which are needed in later chapters are all obtained by linear interpolation.

3.3 Several schemes of better approximation

We turn now to schemes for obtaining wind from Montgomery potential or geopotential height fields.

For the ease of comparison we firstly list below the geostrophic balance,

$$V = \frac{1}{f}\vec{k} \wedge \nabla\Phi \tag{3.27}$$

the balance maintained between pressure-gradient force and Coriolis force (produced by the rotation of the earth). In equation 3.27, Φ can be geopotential if the gradient operator is taken on an isobaric surface, or Montgomery potential if it is taken on an isentropic surface; $V = (u, v)$ is the horizontal wind vector; f and k are the Coriolis parameter and the vertical unit vector respectively.

3.3.1 The linear and nonlinear balance winds

The linear and nonlinear balance winds are approximate winds derived, respectively, from linear and nonlinear balance equations which imply a relationship between nondivergent winds, V_ψ and geopotential height. The balance equations can be derived

ROYAL UNIV. EDINBURGH

from the divergence equation, which can be written, in pressure vertical coordinate, as (see Holton 1979):

$$\frac{\partial}{\partial t}(\nabla \cdot V) + \nabla^2 \left(\Phi + \frac{V \cdot V}{2} \right) + \nabla \cdot [\vec{k} \wedge V(\zeta + f)] + \omega \frac{\partial}{\partial p}(\nabla \cdot V) - \frac{\partial V}{\partial p} \cdot \nabla \omega \simeq 0 \qquad (3.28)$$

where ζ is the vertical component of vorticity, ω the vertical velocity component in pressure coordinates and other symbols are conventional.

Divide velocity field into a nondivergent part V_ψ plus an irrotational part V_d such that

$$V = V_\psi + V_d \qquad (3.29)$$

where $\nabla \cdot V_\psi = 0$ and $\nabla \wedge V_d = 0$. For two dimensional flow, the nondivergent part can be expressed in terms of the streamfunction ψ defined by

$$V_\psi = \vec{k} \wedge \nabla \psi \qquad (3.30)$$

and this easily leads to the expression of vorticity as $\zeta = \vec{k} \cdot \nabla \wedge V = \nabla^2 \psi$.

Scale analysis of the vorticity equation shows that (Holton 1979; Haltiner and Williams 1980) the ratio of horizontal divergence to the relative vorticity is not larger than the order of the Rossby number, i.e. $\left| \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) / \zeta \right| \leq R_0$. Equivalently, this requires that the divergent part of the winds must be smaller than the rotational part, i.e. $V_\psi \geq V_d$. If the Rossby number is small, such as in midlatitude synoptic scale case, the motion V is quasi-nondivergent, i.e. $V_\psi \gg V_d$, the terms involving ω and V_d in equation (3.28) may all be neglected. The remaining terms give the

nonlinear balance equation relating V_ψ and Φ. Using definition (3.30)the nonlinear balance equation can then be written:

$$\nabla \cdot (f \nabla \psi) = \nabla^2 \Phi + \nabla \cdot ((V \cdot \nabla) \cdot V) \tag{3.31}$$

This is a higher order approximation than geostrophic since for synoptic or larger scale motions, the largest terms neglected are of order R_0^2 (see Haltiner 1980 p69). The geostrophic approximation, however, is an order R_0 approximation.

If the nonlinear terms in equation (3.31) are further neglected for consideration of large scale flow (e.g. midlatitude synoptic scale or larger where the nonlinear terms are smaller than the linear terms), we then obtain linear balance equation

$$\nabla \cdot (f \nabla \psi) = \nabla^2 \Phi \tag{3.32}$$

If the variation of the Coriolis parameter can be neglected, this reduces to the geostrophic equation. It gives nondivergent wind and also includes the β effect.

The method of solving the equation (3.31) and (3.32) can be found e.g. in the work of Houghton (1968) and others. The code used here was developed and supplied by Harwood (personal communication). To solve equation (3.31) to obtain ψ, the equation has to be elliptic. But in some cases when large amplitude planetary waves are present, the equation becomes hyperbolic and insolvable and some adjustments has to be made to remove the hyperbolic points. After solving the

equation (3.31) and (3.32) for streamfunction, the wind fields can be obtain from equation (3.30).

3.3.2 Quasi-gradient winds

In order to obtain winds that are more dynamically consistent than those given by geostrophic assumptions, Marks (1987) designed a scheme based on the equations:

$$\frac{uu_\lambda}{a\cos\varphi} + \frac{v\,(u\cos\varphi)_\varphi}{a\cos\varphi} - fv + \frac{\Phi_\lambda}{a\cos\varphi} = 0 \tag{3.33}$$

$$\frac{uv_\lambda}{a\cos\varphi} + \frac{vv_\varphi}{a} + fu + \frac{\Phi_\varphi}{a} + \frac{u^2}{a}\tan\varphi = 0 \tag{3.34}$$

i.e. from the primitive equations for steady horizontal flow without external force or friction.

Three steps are taken in calculating the winds. Firstly the geostrophic wind (u_g, v_g) is found. Then v_g is substituted into equation 3.34 leaving a quadratic for u, which can be solved for u, using the principle of gradient wind to choose the root with smaller absolute magnitude, in order to ensure that for cyclonic or anticyclonic flow a 'regular' low or high is produced. In the last stage, u is substituted back into 3.33, leading to a linear equation for v.

According to the designer this scheme is simpler than stream-function/velocity-potential approach of Boville and more systematic than the ad hoc wave, mean-

flow decomposition of Elson (1986). It is intended to give much better balance in the equations of 3.33 and 3.34 than geostrophic in middle latitude. In further work Marks (1989) showed that when the drag on the zonal flow is considered and the continuity and thermodynamic equations are included together with the use of the diabatic heating rate to give the three dimensional circulation field, the scheme based on the revision of equation (3.33) and (3.34) can produce much better meridional wind component and the horizontal wind field compares well with existing mesospheric climatologies derived from radar and rocket data. Below we use the model output to test this scheme for diagnosing daily winds.

To summarize the approximations implied by equation (3.33) and (3.34), they are (1) the neglect of change with time of the wind components due to climatological consideration for use with monthly mean satellite data; (2) the implication of small Rossby number since the first approximation is geostrophic winds and the principle of gradient winds balance, which is related to geostrophic winds by the Rossby number, is involved. The approximation level seems between $O(\bar{R}_0)$ to $O(\bar{R}_0^2)$.

The term quasi-gradient is used here because the principle of gradient balance for circular flow on an f-plane is implied and similar technique is used for the choice from solutions for u as described above.

3.3.3 Quasi-geostrophic wind

From the primitive equation we can derive straightforwardly the following relations with only the vertical advection term neglected.

$$u = -\frac{1}{f}\left(\frac{\partial \Phi}{\partial y}\right) - \frac{1}{f}\left(\frac{\partial v}{\partial t} + V \cdot \nabla v\right) - \frac{u^2}{fa}\tan\varphi \tag{3.35}$$

$$v = \frac{1}{f}\left(\frac{\partial \Phi}{\partial x}\right) + \frac{1}{f}\left(\frac{\partial u}{\partial t} + V \cdot \nabla u\right) - \frac{uv}{fa}\tan\varphi \tag{3.36}$$

where $V \cdot \nabla = u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}$ with $dx = a\cos\varphi$ and $dy = ad\varphi$. The quasi-geostrophic wind components (u_q, v_q) are calculated by two steps. Firstly we use geostrophic balance and get geostrophic wind components (u_g, v_g) . Then we replace all the u and v on the right hand sides of equation 3.35 and 3.36 by (u_g, v_g) . The resultant u,v of the left hand sides is then called quasi-geostrophic wind and denoted by (u_q, v_q).

The equations (3.35) and (3.36) are the same as the quasi-gradient approximation of equations (3.33) and (3.34) if the steady condition is assumed, though the solving procedures are more straight forward. The level of approximation for quasi-geostrophic is $O(R_0^2)$. We will calculate and discuss these methods in the following sections of this chapter.

3.4 Computational details

As indicated in the first section of this chapter, all the approximate schemes described in the above section are applied in two sets of test: the one using LIMS derived height fields and the one using model output height fields to diagnose the wind fields. The resultant wind fields are compared with the corresponding FGGE analysis winds and model output winds.

3.4.1 Test with LIMS data

For those tests with satellite observational data, the calculations are carried out on the isentropic surface with the Montgomery potential calculated by the method described in the previous section. The level chosen is $ln(\theta) = 6.26$ (near 50mb) which is one of the interpolated surface from the FGGE analysis data. Only the winds in the lower stratosphere are discussed because the interpolated FGGE analysis winds above level of 20mb are extrapolated from below and are believed to have large error (Duncan personal communication). The case chosen is on 21 February 1979 before a major warming.

Because the northmost latitude of the LIMS data is $84°N$, the northern boundary is set accordingly at $85°N$ in diagnosing the winds.

3.4.2 Test with model output data

The model output data of wind, temperature and height fields, kindly supplied to us by A. O'Neill from the U.K. Meterological Office, were from a run from model day of 22 December 1984 to 18 January 1985. The model is the Meteorological Office stratosphere-mesosphere multi-level model (Fisher 1987) based on the primitive equations. These equations are solved to fourth-order accuracy in the horizontal, and to second-order accuracy both in the vertical and in time, using energy-conserving 'box' type finite differences and leapfrog integration. The model utilizes a regular grid in spherical coordinates with gridpoints at intervals of 5° in latitude and longitude and 32 levels between 100 and 0.001mb, which are equally spaced in log pressure and are approximately 2km apart. To avoid having to represent the troposphere, a lower boundary condition is imposed near the tropopause, namely the geopotential height of the 100mb surface specified from analysed observations. A major warming event was assimilated during the run and the simulated winds are believed to be realistic.

For the ease of comparison with the model output wind the calculations using the model data are chosen on the 10mb and 1mb isobaric surfaces respectively. Attention is drawn to two cases, one is on model day 24 December 1984, another on 10 January 1985 for two different patterns of flow.

Near the equatorial area, because the Coriolis parameter, f becomes very small, both the quasi-geostrophic and balance wind are no better than the geostrophic wind due to the appearance of f as denominator in the additional terms of the right hand side. We, therefore, set the south boundary at $15°N$ for all tests, with both the LIMS data and model output data.

In all calculations, the time-dependent terms in the equations 3.35 and 3.36 are found to be at least one order smaller than the others and hardly affect the results. Thus they have been dropped in all calculations leading to the following discussions and this actually shows that the quasi-geostrophic winds practically has the same approximation level as the quasi-gradient winds.

3.5 Analysis of wind errors from the different approximations

In this section the results from the two tests described in the introductory section of this chapter will be analysed and discussed. In brief, in test one the winds diagnosed from LIMS derived height fields by each approximate scheme will be compared with according FGGE analysis winds, considered as the 'true' winds; and in test two winds diagnosed from model output height fields will be compared

with model output winds. By these comparison the wind errors caused by different approximations will be shown and compared. For the analysis of wind errors, the figures of wind fields itself is not persuasive, so attention will mainly be paid to the error fields. Some statistics of related quantities of wind errors will also be discussed in this section.

3.5.1 Wind error caused by different approximation

In the following discussions we will not show the results from quasi-gradient approximation because they are very similar to quasi-geostrophic.

Test with LIMS derived data

Figure 3.1a shows the analysis wind field on $\ln(\theta) = 6.26$ isentropic surface (near 50mb) in FGGE data on 21 February 1979 before a major stratospheric warming. This is a wave-2 pattern flow with cyclonic centres located over north America and north Asia respectively. Figure 3.1 b,c,d and e show the wind errors to the analysis wind fields by using approximations of geostrophic, quasi-geostrophic, linear balance and nonlinear balance assumption respectively. From figure 3.1b we can see that geostrophic winds have noticeable error in the cyclonically curved area, for example around $120°E, 50°N$, where it over estimated the wind speed.

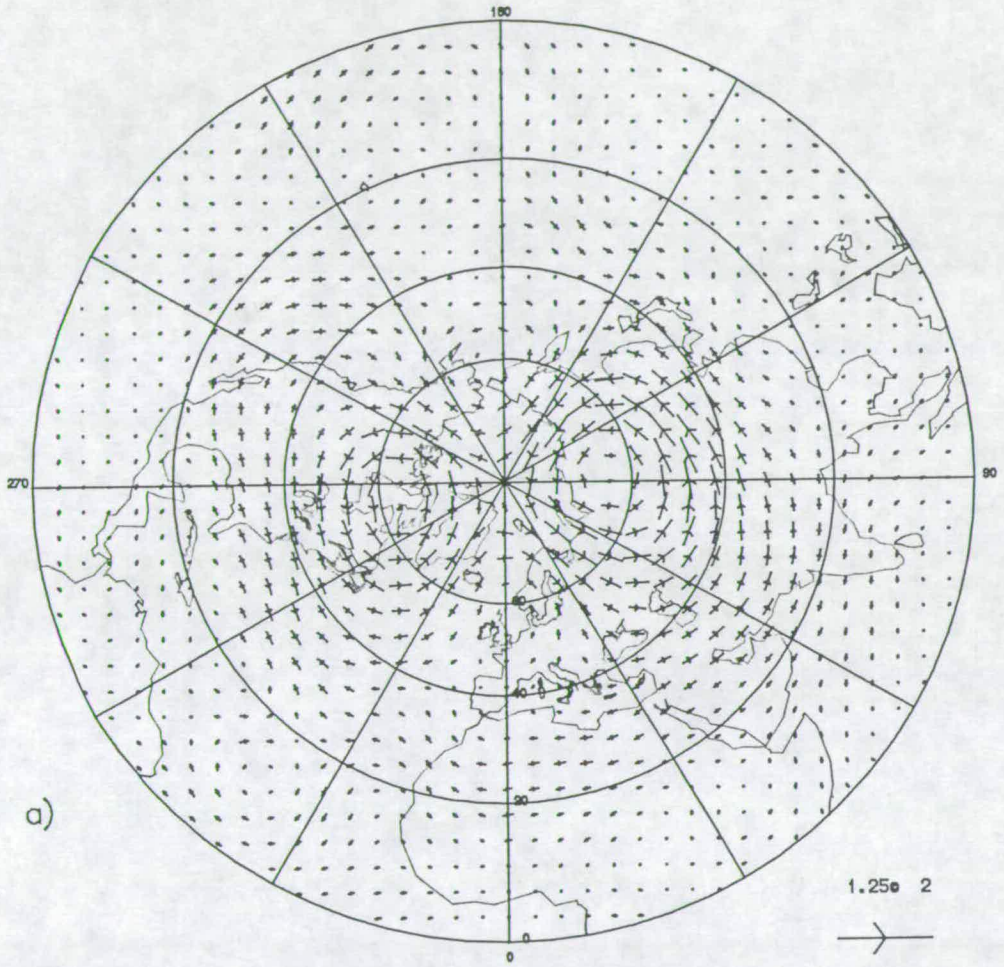
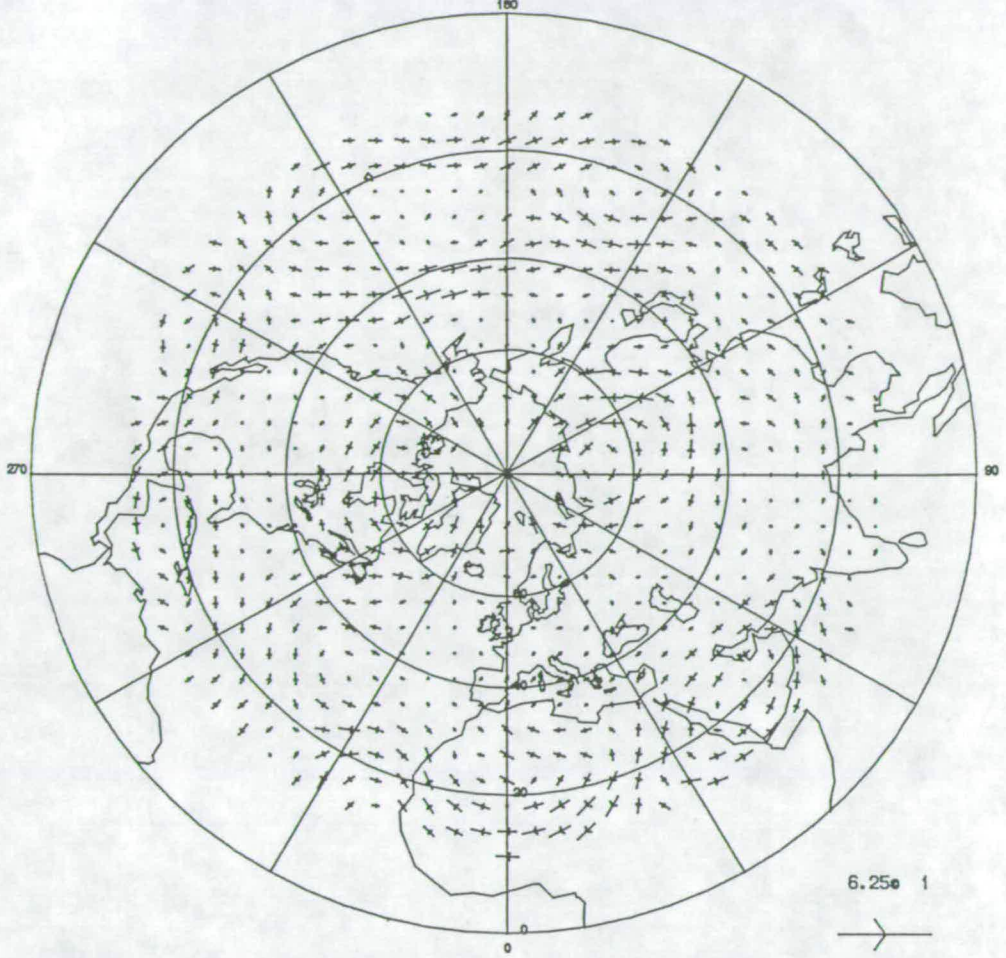


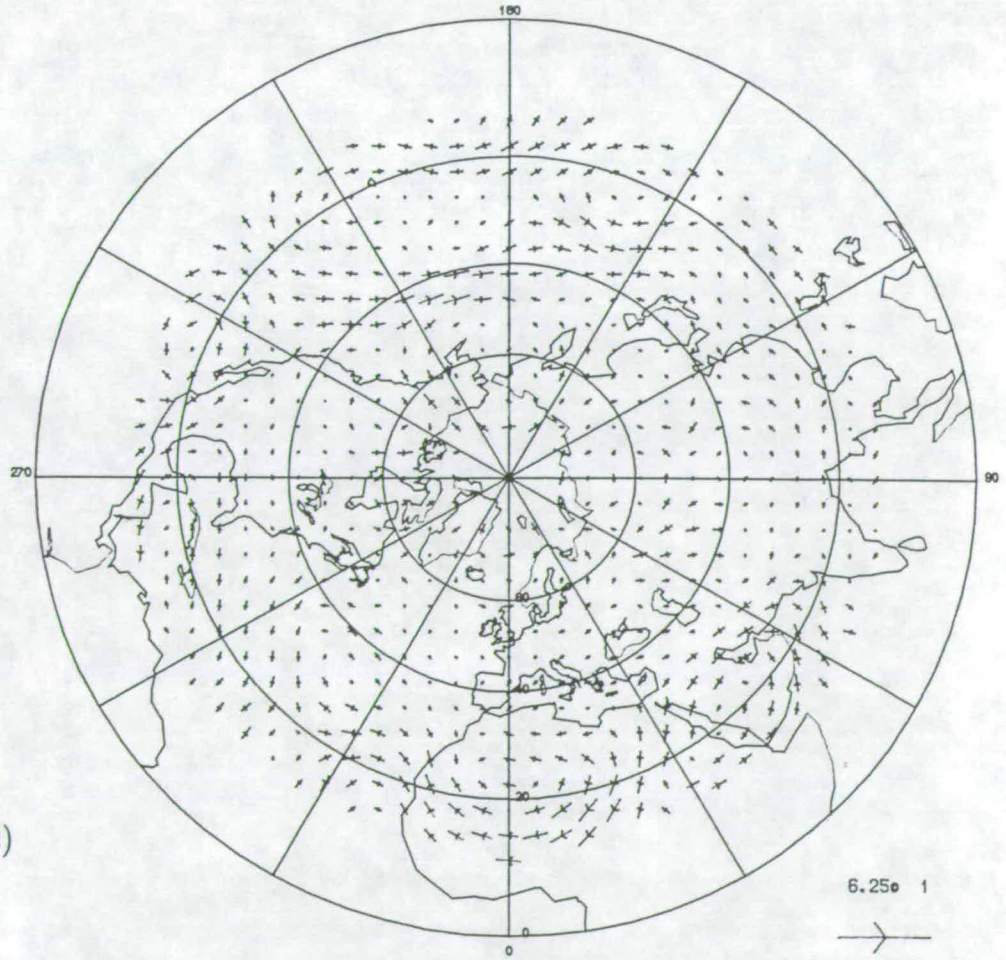
Figure 3.1: Wind fields and errors estimated from LIMS data. (a) Corresponding FGGE analysis wind field near 50mb on 21 February 1979. (b) Error field caused by geostrophic approximation. (c) As (b) by quasi-geostrophic. (d) as (b) by linear balance. (e) as (b) by nonlinear balance. Unit: m/s

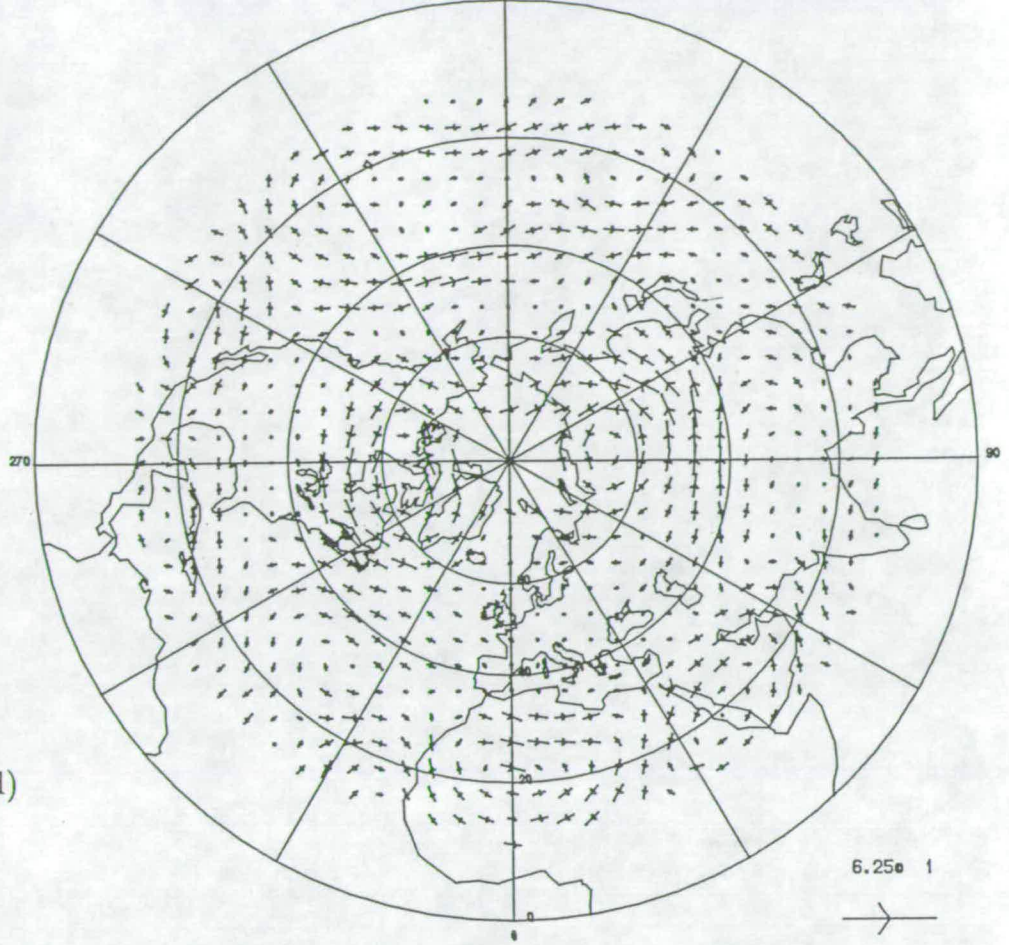
* Note: the scale in (b), (c), (d) and (e) is different from (a). The same lenth in error field represents only half the wind speed in (a).

(b)

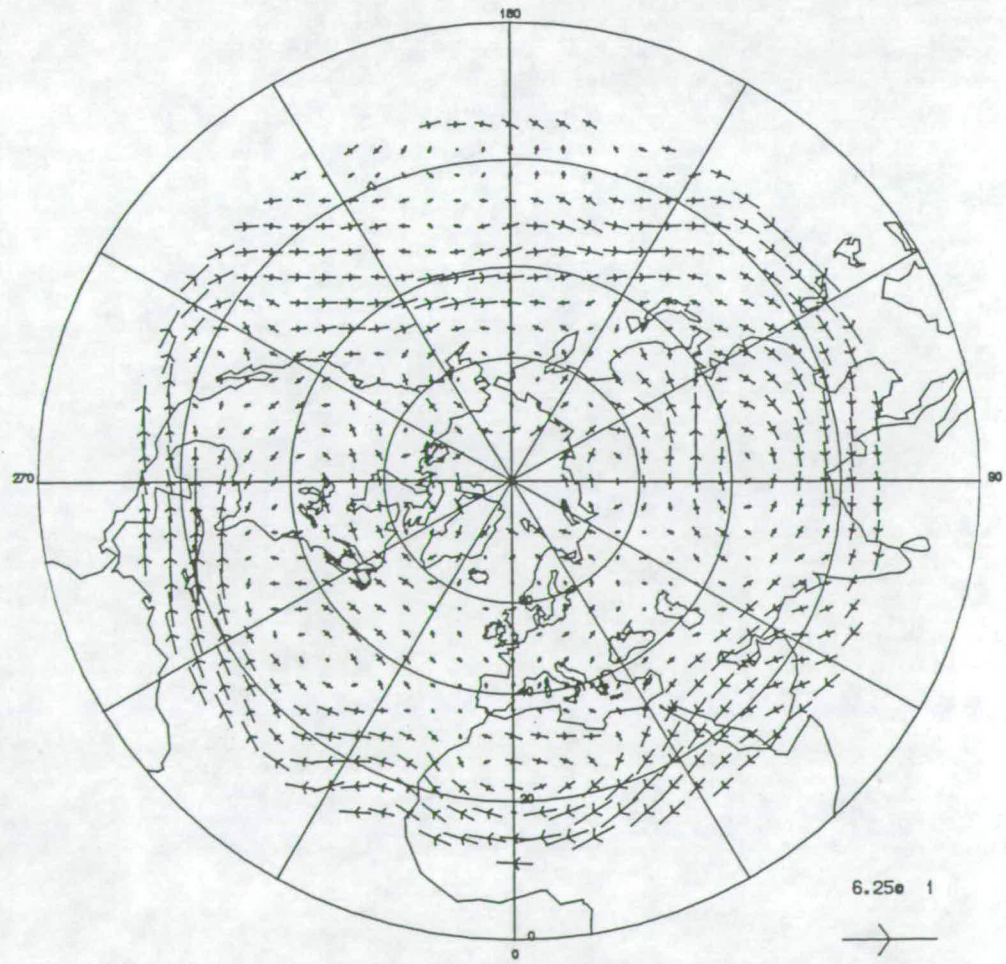


(c)





(d)



(e)

The quasi-geostrophic wind error in figure 3.1c, however, improved the errors of geostrophic wind in the cyclonically curved region significantly with the errors in the other extratropical areas also reduced. The linear and nonlinear balance wind error fields of figure 3.1d,e show no improvement over geostrophic winds with nonlinear wind error even larger in most areasthan geostrophic, which is not in accord with our expectation nor with Boville's (1987) results. For example in the region between $20 - 60°N$ latitude and $90 - 120°$ longitude and also in the area around 20° latitude circle, the errors caused by nonlinear approximation is particular larger than the errors caused by geostrophic. The reason of the failure of nonlinear balance wind may lie in the restriction on the flow by the ellipticity criterion demanded for solving the nonlinear equation. In the process of solving the nonlinear equation, we often have to remove the points where the ellipticity criterion is violated and the equation becomes hyperbolic, especially in the first iteration. This may further suggest that the physical balance in the real complex atmosphere may not be satisfied with pure rotational flow underlying the derivation of nonlinear balance equation.

The r.m.s error, defined as $\sqrt{\sum_{i=1}^{n}(x_{i_d} - x_{i_t})/n}$, where x_{i_d} and x_{i_t} denote wind components from diagnostic and 'true' winds respectively, is also calculated and the resultant numbers are shown in table 3.1.

We can see from the table that the r.m.s. errors of both u and v components from

Scheme	r.m.s. error	
	LIMS data	
	21 Feb. 79	
	u	v
Geo.	5.967	4.960
Quasi-geo.	5.128	4.405
Linear	8.307	5.338
Nonlinear	9.508	4.785

Table 3.1: r.m.s. errors of wind components from approximate schemes to the 'true' wind from FGGE analysis

quasi-geostrophic approximation have the smallest values than other schemes in this test of using the LIMS satellite observational data provided that we believe the FGGE analysis are 'true' winds. This confirms the above analysis that the quasi-geostrophic method achieves the best approximation in this case.

Test with model output data

Figure 3.2a shows the original model output wind fields for model day 10 January 1985 on 1mb. It can be seen that the flow is a pattern of a weak wave-2 with small wave amplitude.

Assuming the model output wind as 'true' winds and taking the model output height fields as data, the previously described approximations have been applied to deduce the approximate wind and wind error fields. Figure 3.2b,c,d show the error fields of winds by geostrophic, quasi-geostrophic and nonlinear approximation on 10 January model day respectively, with linear balance winds not shown because the result is similar to geostrophic. Quite similar results to the LIMS case can be observed here with the quasi-geostrophic approximation having the smallest error and the errors from the other approaches are also small. This can be further confirmed from the r.m.s. error of u, v-component shown in table 3.2.

The tests on model day 24 December, however, are rather different. Figure 3.3a

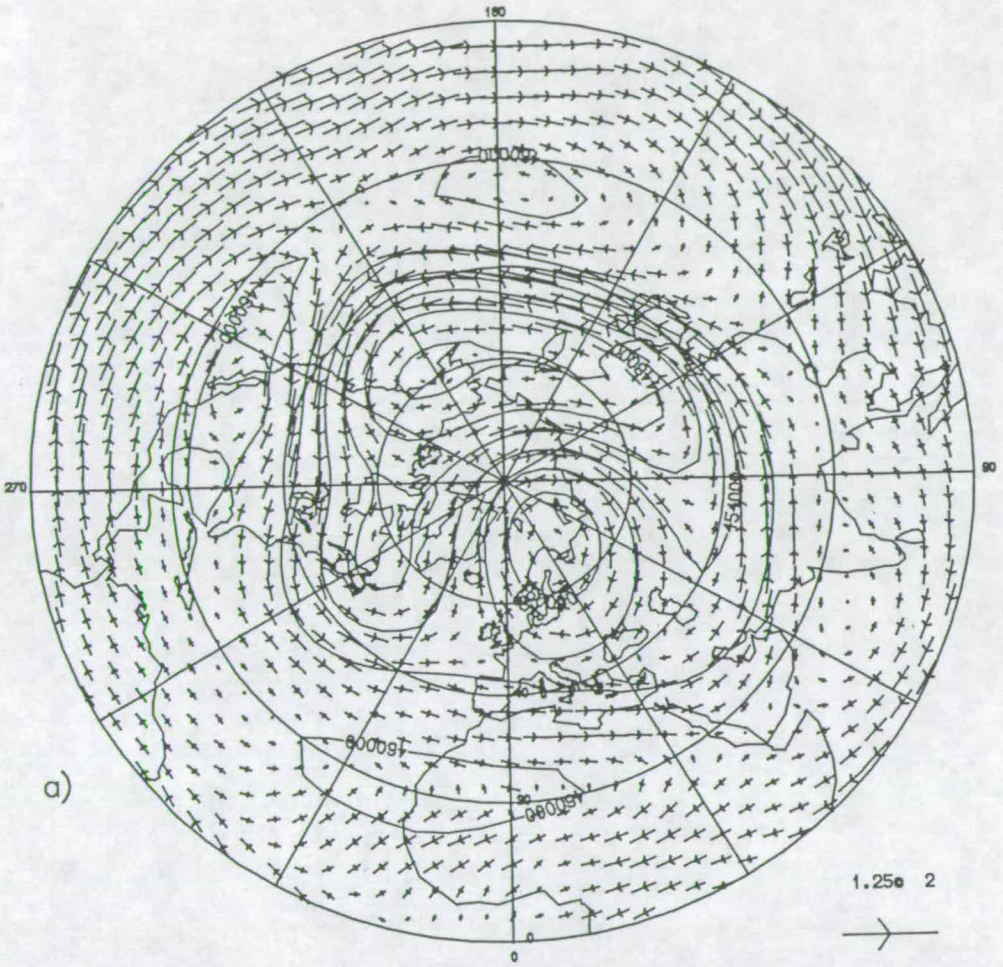
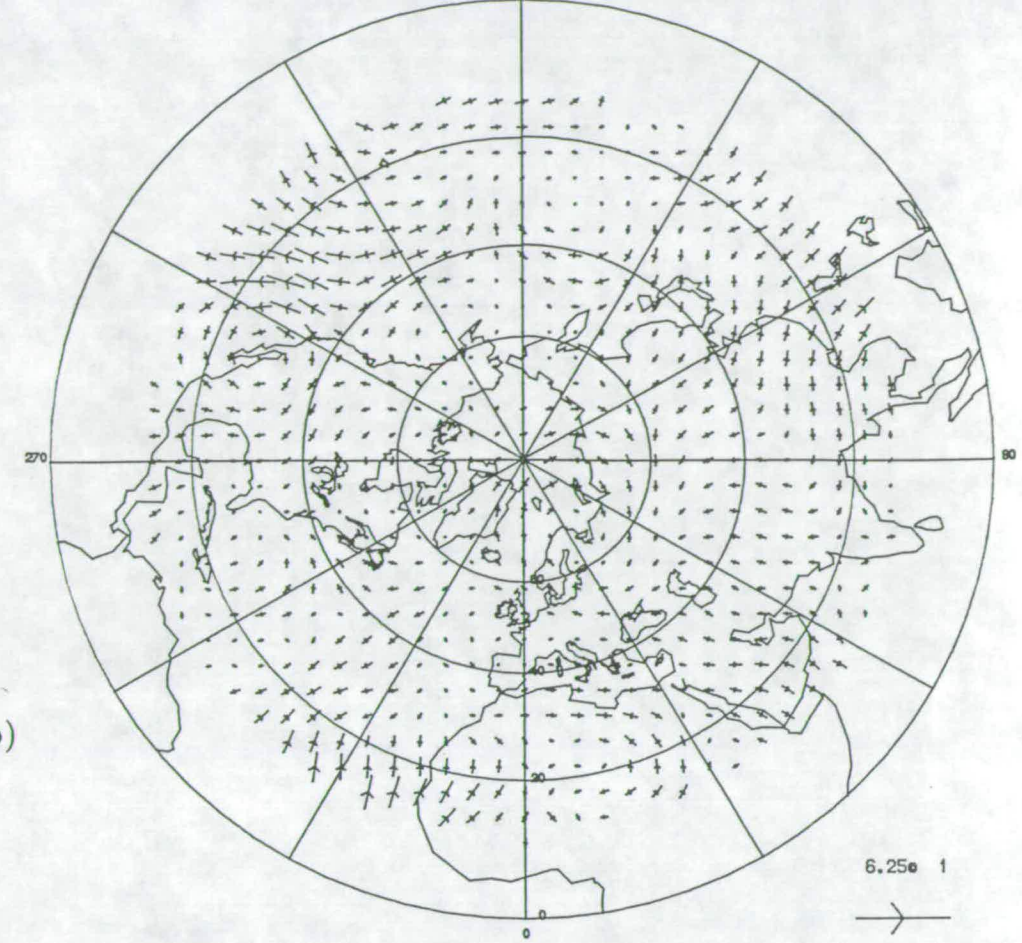


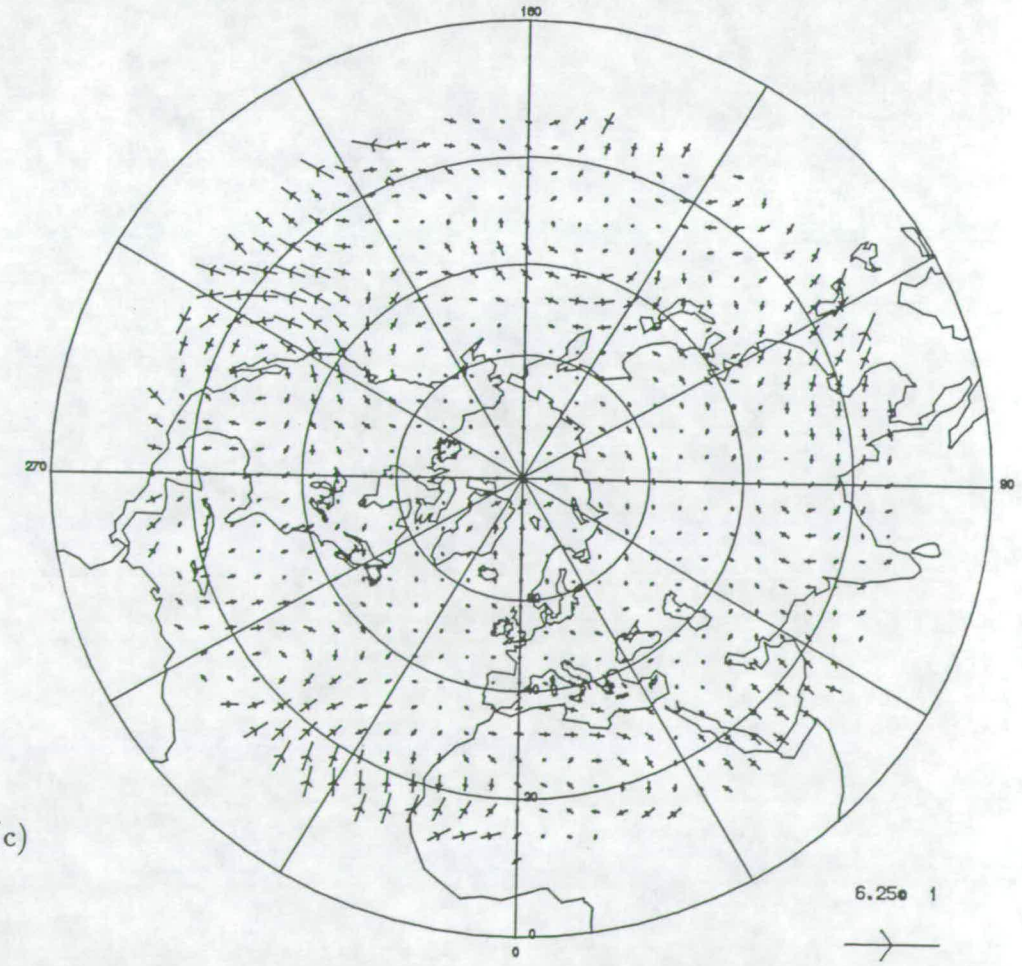
Figure 3.2: Wind fields and errors estimated from model output data on 1mb model day 10 January 1985. (a) Model output wind field. (b) Error field caused by geostrophic approximation. (c) As (b) by quasi-geostrophic. (d) By nonlinear balance. Unit: m/s

* Note: the scale in (b), (c) and (d) is different from (a). The same lenth in error field represents only half the wind speed in (a).

(b)



(c)



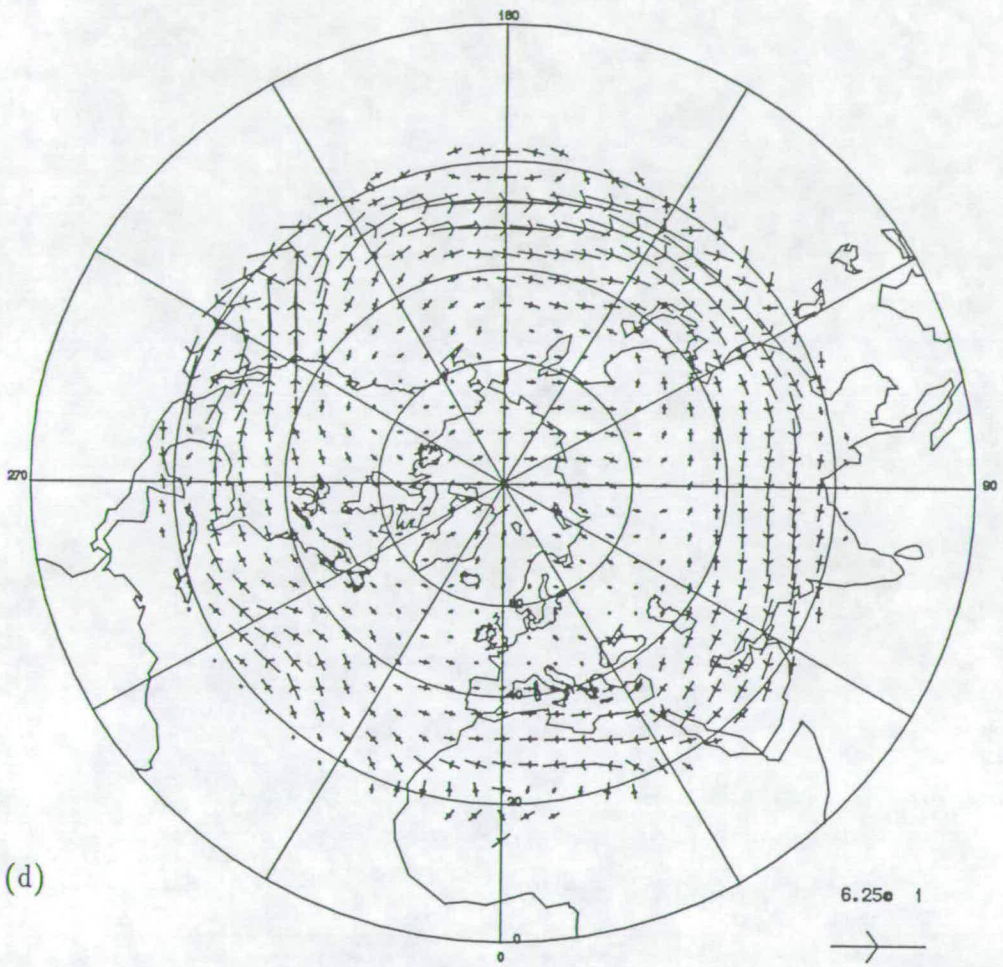


Figure 3.2: (figure continued)

Scheme	r.m.s. error			
	Model output data			
	10 Jan. 85		24 Dec. 84	
	u	v	u	v
Geo.	4.111	4.950	13.077	9.513
Quasi-geo.	3.362	4.370	12.767	11.958
Linear				
Nonlinear	11.313	4.393	16.062	10.100

Table 3.2: r.m.s. error of wind components from approximate schemes to the 'true' winds from model output.

shows the model output wind field on the model day 24 December 1984 on 1mb. In comparison with figure 3.2a, different pattern of flow appears in this figure, with wave-1 prevails and the wave amplitude and the wind speed are very large. (figure 3.3a,b,c,d the original wind field from model output and the error fields of winds on 1mb), where almost every scheme has significant errors to the 'true' winds with on the one hand geostrophic winds over-estimating the wind speed and on the other hand quasi-geostrophic approximation under-estimating in the large area of cyclonically curved regions. Nonlinear balance winds, however, have smaller error in the curved regions (This may reflect the strong nonlinearity in those areas and the relative fitting of the nonlinear balance) but with much larger errors in the other areas. The r.m.s. error in Table 3.2 also suggests that it is hard to say which approximation scheme is better than others with quasi-geostrophic winds having smallest value in deviation of u-component and geostrophic in v-component. The reason why almost all the approximation schemes collapse in this occasion will be discussed in the last section of this chapter.

3.5.2 Average momentum flux error

As discussed in the introduction of this chapter, it is the error in the E-P flux divergence caused by geostrophic assumption in obtaining the winds that provides our principal motive for seeking a better approximation scheme to diagnose the

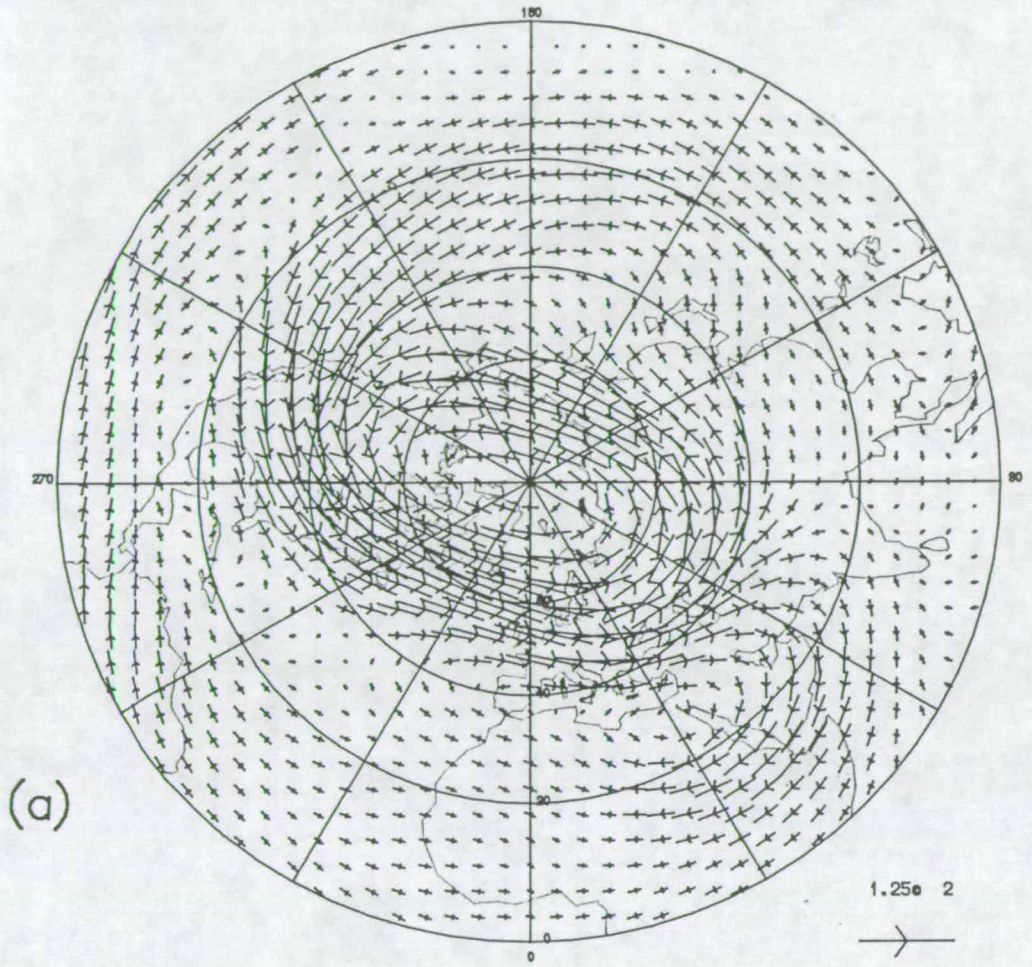
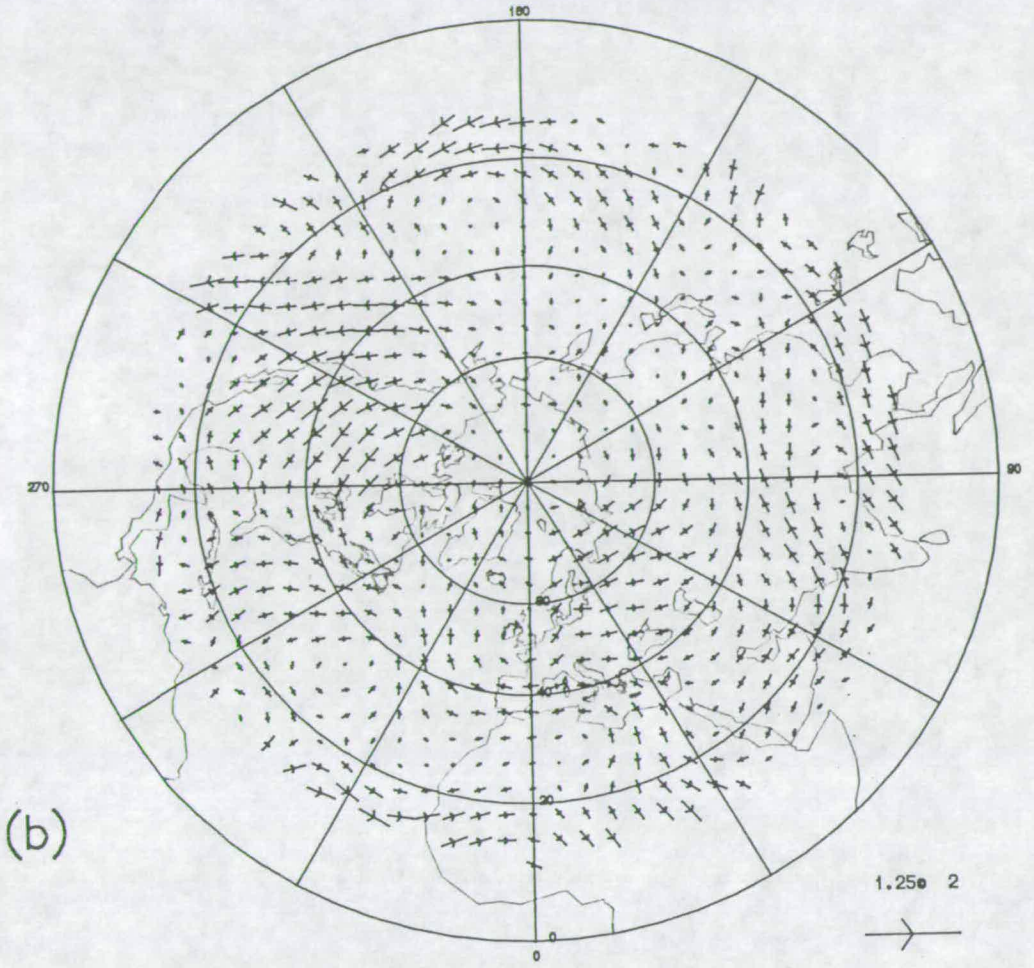
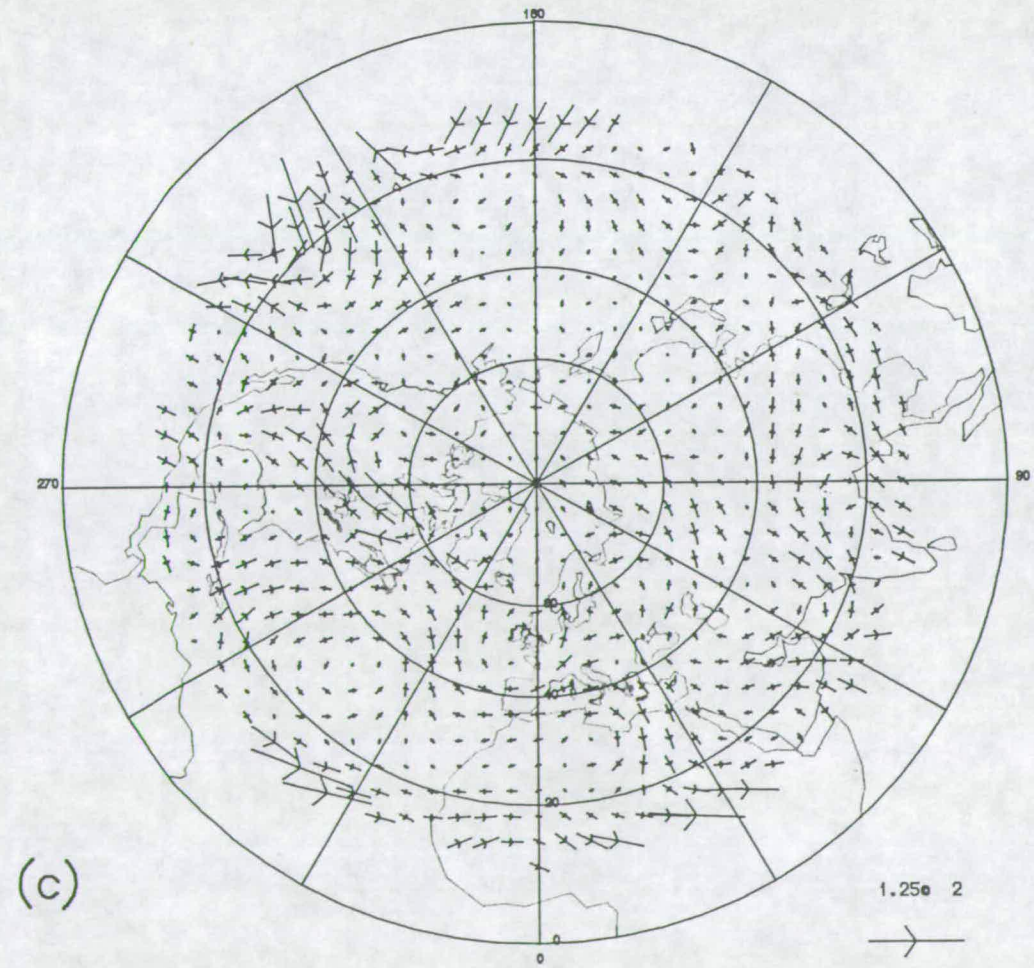


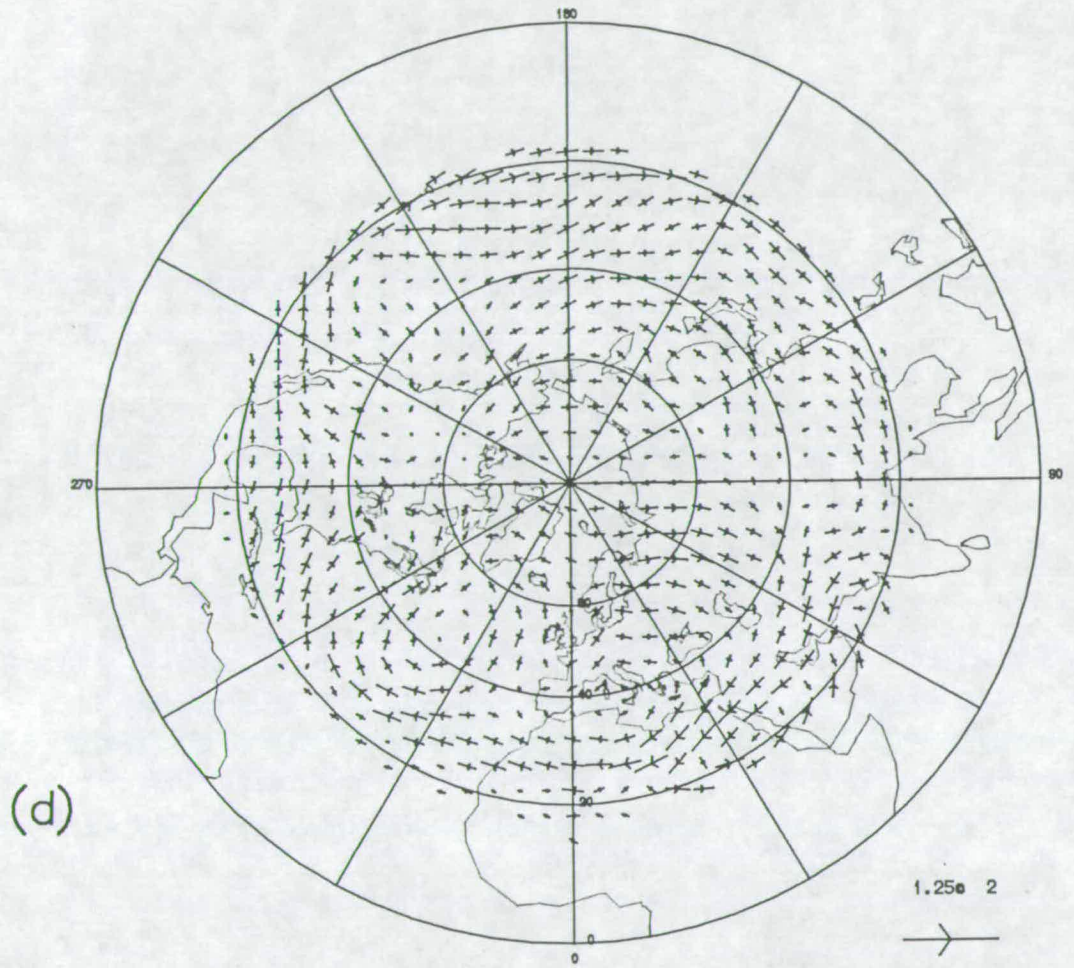
Figure 3.3: As figure 3.2 but from model output data on 1mb model day 10 January 1985. (a) Model output wind field. (b) Error field caused by geostrophic approximation. (c) As (b) by quasi-geostrophic. (d) As (b) by nonlinear balance. Unit: m/s



(b)



(c)



(d)

Figure 3.3: (figure continued)

stratospheric winds. Boville's (1987) investigation found that the error in the E-P flux divergence can mainly be attributed to the error in the geostrophic momentum wave flux while the error in the geostrophic heat wave flux contributes relatively little. Accordingly the errors in momentum wave flux caused by each method are contrasted here.

Test with LIMS derived data

Figure 3.4a shows the averaged momentum wave flux on 21 February derived from the different methods and the 'real' winds on the level, where the linear balance result are again not included because of their similarity to the geostrophic. It is obvious that except near the north pole boundary where every scheme drifts away from the 'real' flux, the quasi-geostrophic momentum wave flux curve is the nearest one to the real flux with geostrophic and nonlinear having much large error (about 40% at mid-latitude $56°N$). This is in agreement with the results discussed previously.

Test with model output data

Figure 3.4b is the same as Figure 3.4a but for 1mb on 10 January 1985 with model data. We can see that the quasi-geostrophic momentum wave flux has the same trend as the real flux and is proportional to it with smaller error; whereas the

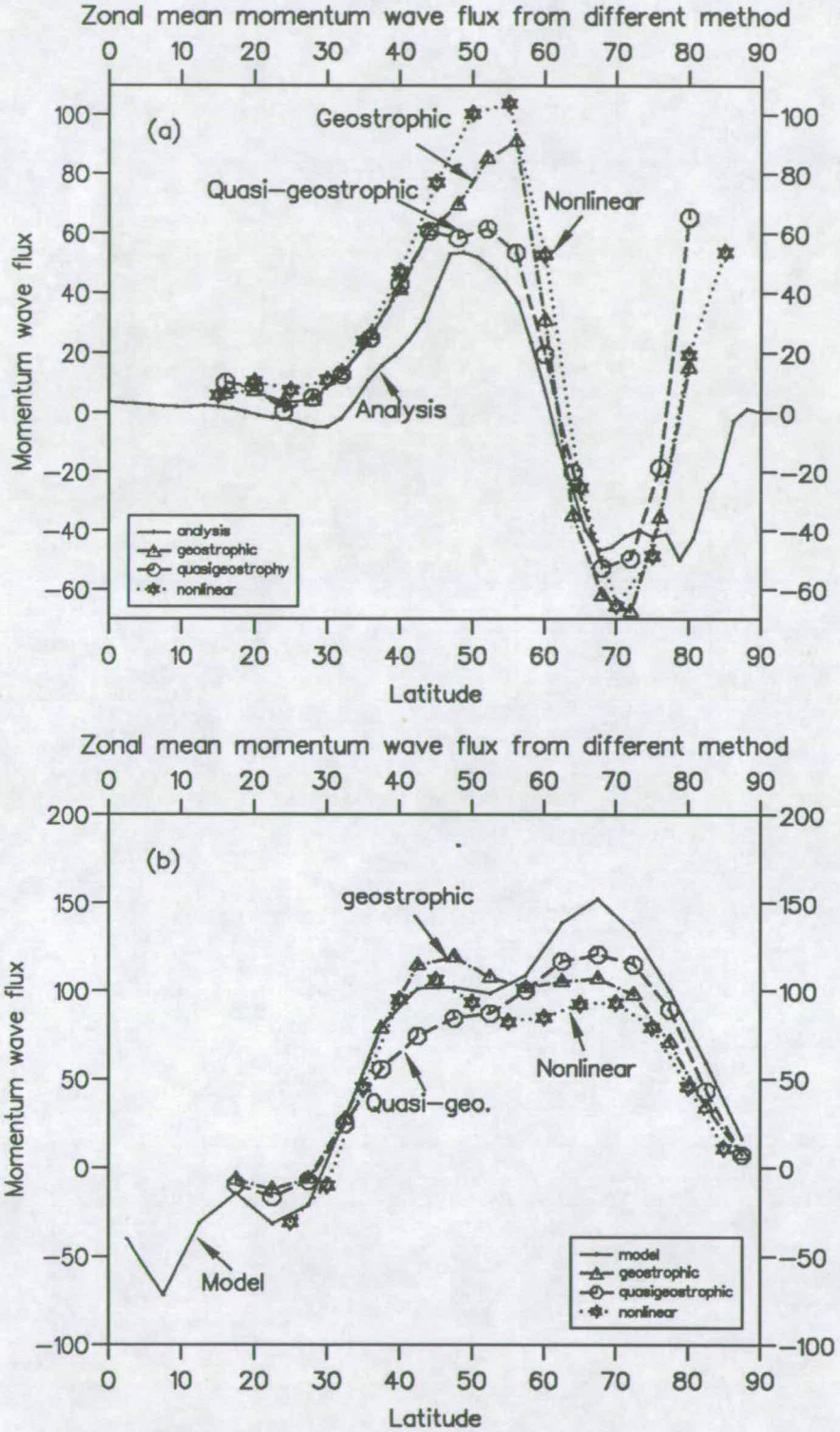


Figure 3.4 Mean momentum wave flux derived from analysis,model output and approximated winds resp.. With a) on 21 Feb. 1979, b) on model day 10 Jan. 1985 and c) model day 24 Dec. 1984. Unit: m**2/s**2
Note: a) From LIMS data. b) and c) From model data.

geostrophic approximation over estimates the flux in mid-latitude $40 - 50°N$ and under estimates it in higher latitude $56 - 75°N$ leading to spurious positive E-P flux divergence around $60°N$ (not shown). The momentum wave flux derived from nonlinear approximation in the region of $25 - 45°N$ approximates the 'true' flux quite well but is under estimated northward.

For the case of 24 December model day (figure 3.4c), the result is rather different.

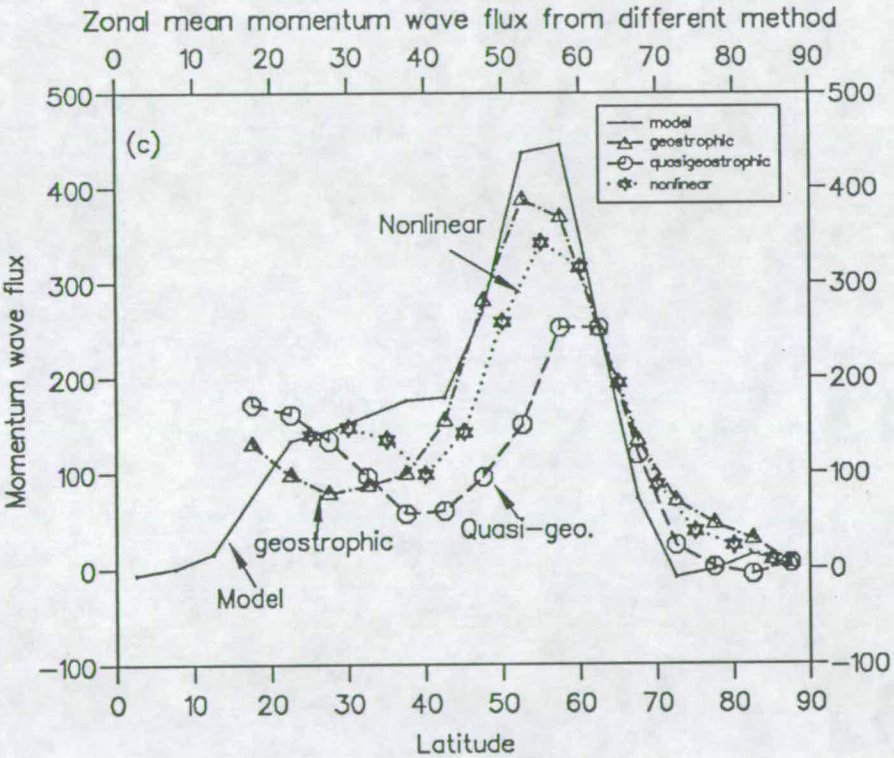


Figure 3.4.c (figure continued)

All the approximations produce large errors in the momentum wave flux with error of momentum wave flux from geostrophic being the smallest and from nonlinear the second and from quasi-geostrophic the largest. This is also in accordance with the wind error and the reason will also be discussed below.

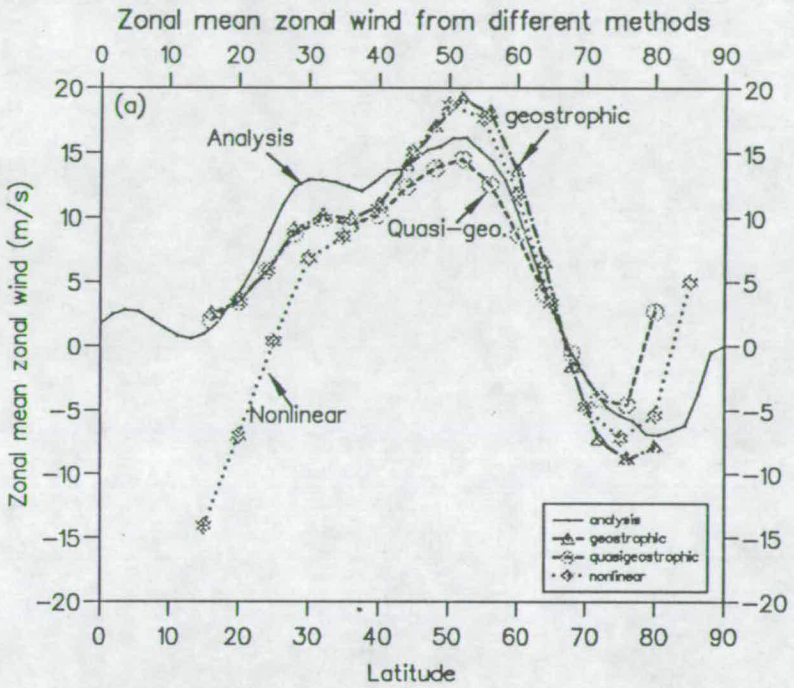
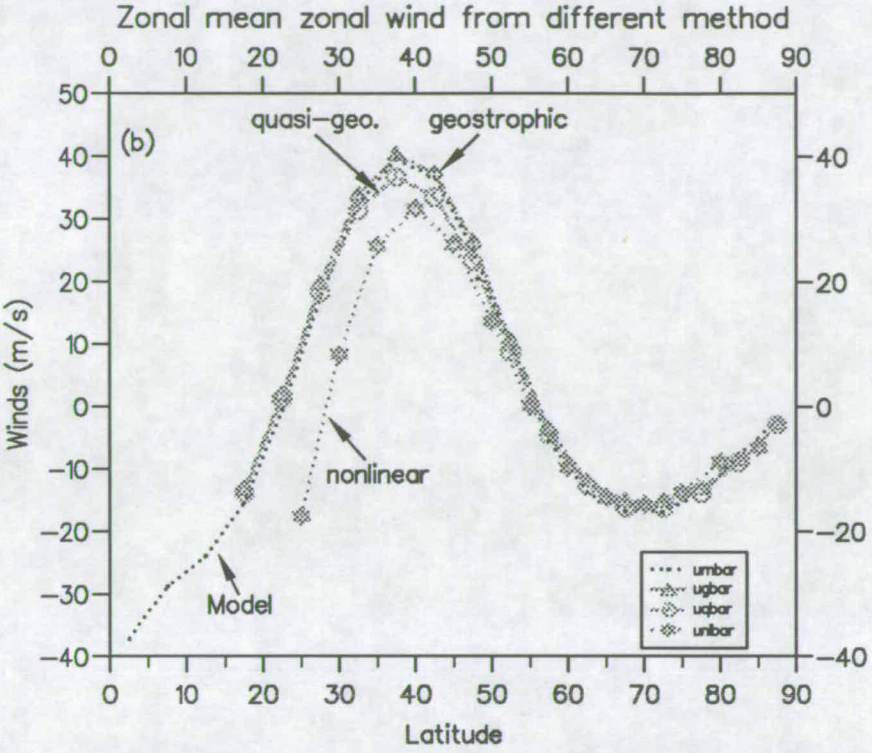


Figure 3.5 Zonal mean zonal wind derived from analysis and model output and approximations. With a) on 21 Feb. 1979. b) model day 10 Jan.1985 and C) model day 24 Dec. 1984. Unit:m/s. Note: a) from LIMS data, b) and c) from model data. Height: a) near 50mb, b) and c) on 1mb.

3.5.3 The errors in the zonal averaged zonal mean wind

It is well-recognized that the geostrophic zonal mean wind often over-estimates the zonal mean wind especially in the stratosphere where it can be 20% too large (see Boville 1987). The zonal mean zonal wind is also checked here. Figure 3.5a,b,c give the zonal mean wind curves from different approximations and analysis and model output results for the three occasions tested, with figure 3.5a from the test 1 with LIMS data and the others from the test two with model output data. In the small wave case (fig3.5a,b), every method gives a good approximation with



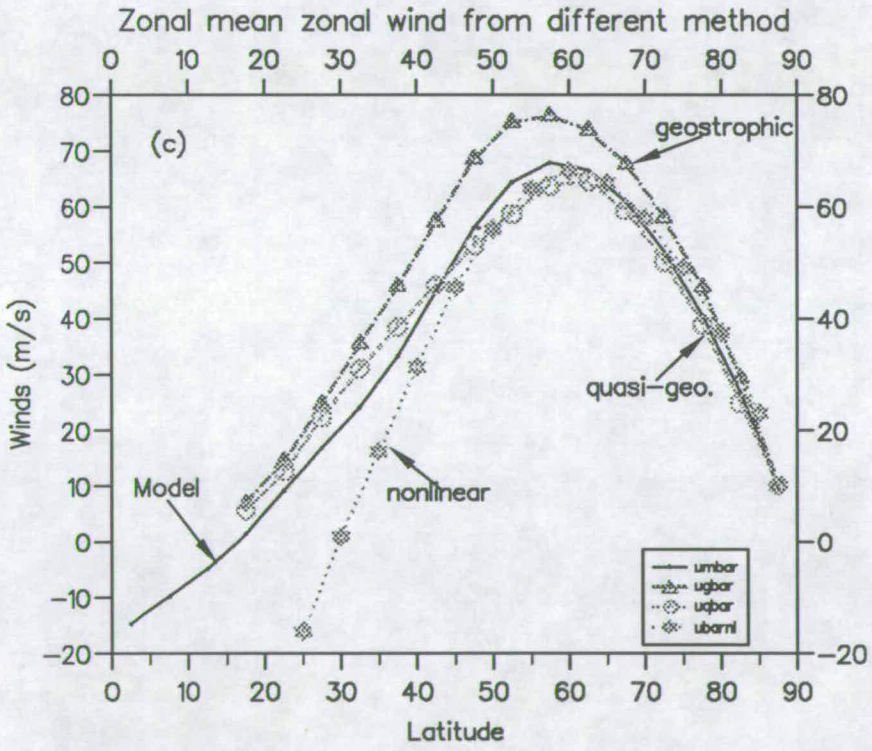


Figure 3.5b,c Zonal mean zonal wind derived from model output winds and different approximations. With b)on 1mb model day 10 Jan. 1985, and c) on 1mb model day 24 Dec. 1984. Unit: m/s.

quasi-geostrophic having the smallest error in using observational data (fig. 3.5a) and nonlinear under estimated so much in low latitude. But in the large wave amplitude case (Figure 3.5c) the quasi-geostrophic is better, with geostrophic over-estimating the zonal mean wind as discussed in the references cited above. It is interesting to see that the zonal mean zonal wind from nonlinear approximation is very close to the model output northward of $45°N$ but is again under estimating in lower latitudes.

3.6 Discussion

We showed above that in most cases quasi-geostrophic winds provide us a good approximation, but when the planetary wave amplitude is very strong almost all the methods have considerable error, for instance the case with model data on model day 24 December 1984. To look into the cause of these phenomenon, we plot out here the Rossby number distribution for each case. Because Rossby number is a criterion of validity of geostrophic balance, we know from theoretical analysis that when the Rossby number is much smaller than 1, geostrophic balance, to the first order, will be valid. Otherwise if the Rossby number is comparable to 1, the acceleration would be about the same order as Coriolis force, in other words geostrophic balance would break down. From the process of calculating

the wind one can see that the quasi-geostrophic wind implicitly relies on the validity of the geostrophic wind, thus if the geostrophic approximation broke down one cannot expect the good behaviour of quasi-geostrophic approach. In this case the nonlinear balance equation often becomes hyperbolic and can not be solved to get a true nonlinear solution and thus no good result either.

Figure 3.6a,b give the plots of Rossby number in the model cases of our calculation. In the LIMS data case, except in the equatorial area (outside our calculation area) the Rossby number is smaller than 0.2 in most midlatitude areas and than 0.4 in a small polar patch (not shown). Therefore in both the LIMS case and the model case of 10 January 1985, the Rossby number is generally small with maximum value not exceeding 0.4. All the methods give relatively good winds with the quasi-geostrophic having the smallest error to the 'true' wind. In the case of model day 24 December 1984 we see that Rossby number is quite large with most areas within the vortex larger than 0.4 and in the cyclonically curved region exceeding 0.6. All of the approximations produce large errors in this case . Therefore the local breakdown of geostrophic balance may explain the wind errors in this case.

In summary, we have compared several schemes to obtain approximate wind fields from the height fields either derived from satellite observational data or produced by a stratospheric numerical model and compared them with the corre-

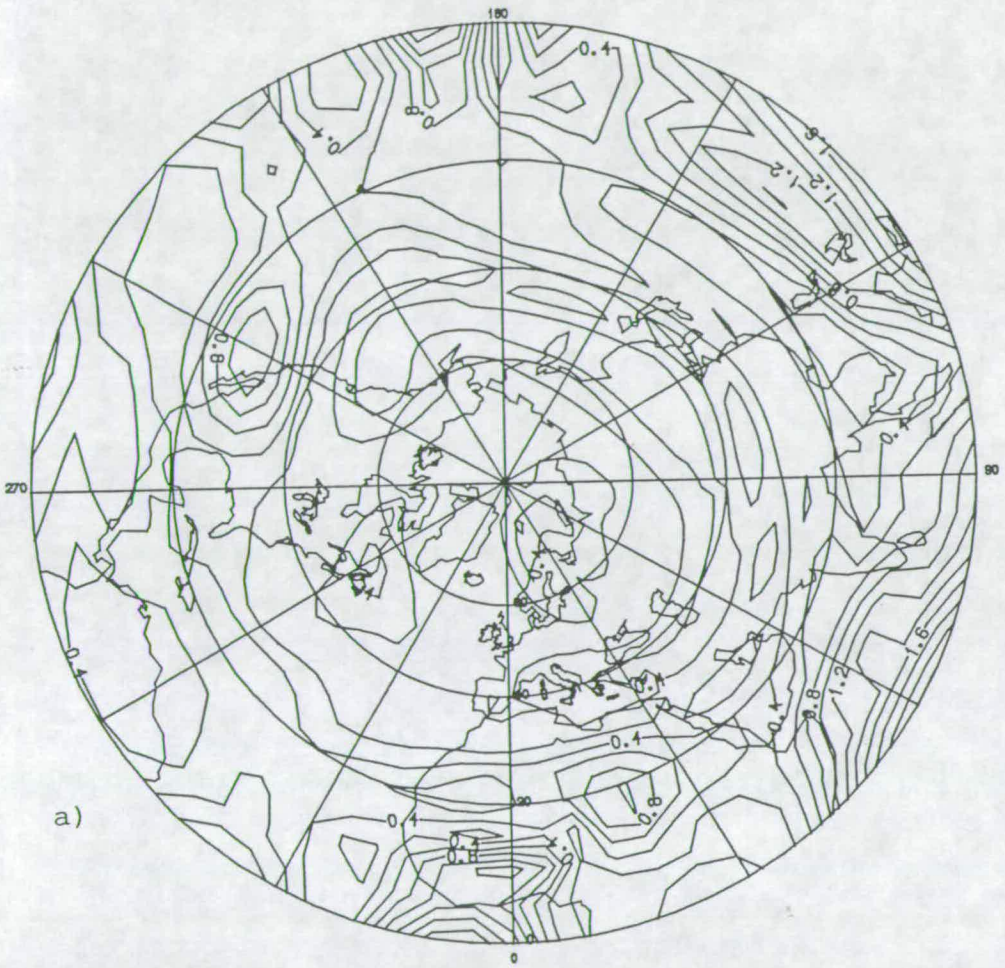


Figure 3.6: Rossby number derived from model output wind field. (a) On 1mb model day 10 January 1985. (b) On 1mb model day 24 December 1984. The contour interval is 0.2.

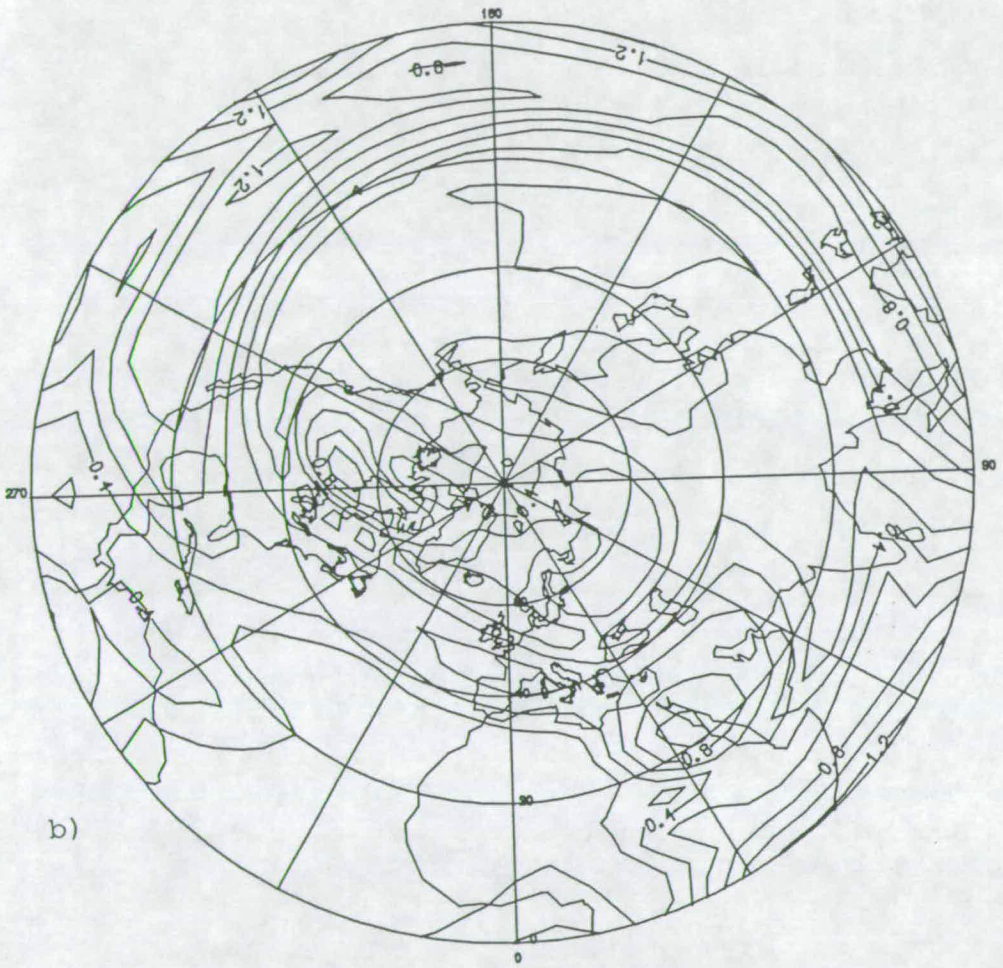


Figure 3.6: (figure continued)

sponding FGGE analysis wind or model output wind respectively. Considerable errors involved in using the geostrophic approximation are found, especially in the strong cyclonical curvature area where either the nonlinear effects may be large or the acceleration term may be important. The error in the eddy momentum flux caused by the geostrophic wind is enormous reaching 40% in the maximum eddy flux region. In accordance with many previous studies (e.g. Boville,1987) this may partly explain the spurious dipole structure of E-P flux divergence in some studies of the stratosphere.

The linear and nonlinear balance winds, overall, show no improvement over the geostrophic wind, though there is reduced error in the curvature region. The nonlinear balance equation, though often converged to a reasonable solution in the tests with the satellite observational data, cannot converge without frequent removal of hyperbolic points in the region near the strongest winds (not shown) in the model data. The ellipticity criterion demanded for solving the nonlinear balance equation is often violated in both observational and model data, with higher frequency in the latter case. This together with the unsatisfactory nature of winds obtained by both balance equations suggests that the physical balance in the real atmosphere may not be solely satisfied by pure rotational winds. Either the local vertical advective terms may be important or motion at unresolvable scale may play a role, or the role of divergence may not be neglected in such a

situation.

Comparison of all the approximations give a clear advantage to the quasi-geostrophic winds in both the observational data and model data. It also produces relatively small errors in the wave momentum flux when Rossby number is small and in the zonal mean winds in all the cases involved.

The Rossby number is found to be crucial for all the approximations. When the Rossby number is small all approximations give better winds than they give in the large Rossby number case. The quasi-geostrophic wind is suggested to be a good alternative for geostrophic winds in the small Rossby number case. While in the large planetary wave case with large Rossby number all the approximations together with geostrophic winds have large errors. It seems that no good wind fields can be found diagnostically from the height fields with those methods discussed. Further work may be needed. One of the possibilities to improve the wind approximation in large wave cases is to take into account the drag effects which as indicated by Marks (1989) are important in the middle to high levels of the stratosphere.

As a result of these tests we have adopted the quasi-geostrophic approximation for the remainder of this thesis.

Chapter 4

The Effect of Wind Divergence on The Polar Vortex

4.1 Introduction

The further recognition and highlighting of the role played by potential vorticity on isentropic surfaces in large scale dynamics is a recurrent theme of the last decade in the advance of atmosphere research, though the introduction and use of potential vorticity in ~~the~~ atmospheric dynamics can be traced back to as early as Rossby and Ertel (Hoskins et al., 1985). In a pair of papers, McIntyre and Palmer (1983, 1984) have emphasized the advantage of using maps of Ertel potential vorticity on isentropic surfaces (IPV hereafter) as a tool to visualize large scale dynamical processes in the Northern Hemisphere wintertime middle stratosphere

and also stressed the new insight these IPV maps may give into many aspects of stratospheric behaviour, such as tracer transport, wave propagation, instability, critical layers and some other tropospheric phenomena. This idea, lately called 'IPV thinking' by Hoskins et. al. (1985), has shed new light on our understanding of stratospheric dynamics and gained support from further use (Clough et al., 1985; Dunkerton et al 1986).

The first breakthrough made by IPV maps and IPV thinking is the provision of direct observational evidence for the existence of planetary wave breaking in the stratosphere by McIntyre and Palmer (1983,1984). In their pair of papers, they also defined and suggested the use of a so-called 'circulation index', the area enclosed by isopleths of potential vorticity on isentropic surface, to quantify the mechanics involved in the formation of 'main vortex' and 'surf-zone' structure. Later work by Butchart and Remsberg (1986) confirmed the so-defined 'area' or 'index' as a practical and useful approach to quantify the effects on the area by various processes. However, their study assumed that divergence did not contribute to area change and it remains to be investigated whether this assumption is justified or not.

The main purpose of present chapter is to estimate the possible influence the wind divergence may have on the change of these areas. There are many other aspects of exploiting the idea of IPV thinking in the investigation of phenomena such as

tropospheric dynamical processes (see for example, Hoskins et al.,1985; Haynes and McIntyre,1987) that are, however beyond our present scope and will not be discussed here. Section 2 will discuss the 'main vortex', 'surf-zone' structure, revealed by IPV maps. Section 3 of this chapter will present the importance of the area enclosed by isopleths of IPV in quantifying wave breaking events. Possible influences on the area change by wind divergence will be discussed in section 4. The last section, section 5 will show the results and analysis from the present investigation.

4.2 IPV maps and 'main vortex', 'surf-zone' mechanics

Based on the isentropic maps of potential vorticity, McIntyre and Palmer (MP hereafter) managed to show a striking key fact that the wintertime extratropical middle stratosphere is clearly divided into two sharply defined, zonally asymmetric regions, which they called the polar 'main vortex' and the 'surf-zone'. The polar 'main vortex' is characterized by steep gradients of PV at its edge; whereas the broader 'surf-zone' surrounds the 'main vortex' with comparatively very weak, systematic, large scale gradients of PV. According to MP, this is the principal region where wave breaking or rapid and irreversible deformation of material contours

and irreversible tracer transport take place. MP also consider the 'surf zone' as an analogue of the theoretical nonlinear critical layer. Given a basic knowledge of the stratospheric circulation combined with a theoretical understanding of nonlinear Rossby waves, it appears that the wave breaking and associated irreversible mixing of constituents is almost certainly happening throughout the winter. During the wave breaking processes, air with high potential vorticity is drawn out from the polar vortex as long thin tongues. These tongues tend to become longer and thinner as time goes on and some pieces may be removed by mixing cascaded down to smaller scales and rapidly disappear from view in any approximate representation having limited horizontal resolution. This mixing into the surrounding 'surf-zone' or erosion of the 'main vortex' by wave breaking is irreversible. Also based on an inspection of the evolution of the 'main vortex' and 'surf-zone' structure and the premise of almost continual wave breaking in the stratosphere of the northern hemisphere, MP predicted that as the winter progresses the 'main vortex' is subject to erosion by the action of the breaking planetary waves in the 'surf zone' and that the breaking is also one of the characteristics leading to a major warming. This structure of the two regions and the hypothesis of wave breaking are confirmed by the corroborated work of Leovy et al.(1986).

Figure 4.1 from Butchart and Remsberg (1986) is a schematic representation of how the gross distribution of potential vorticity looks before and after fairly com-

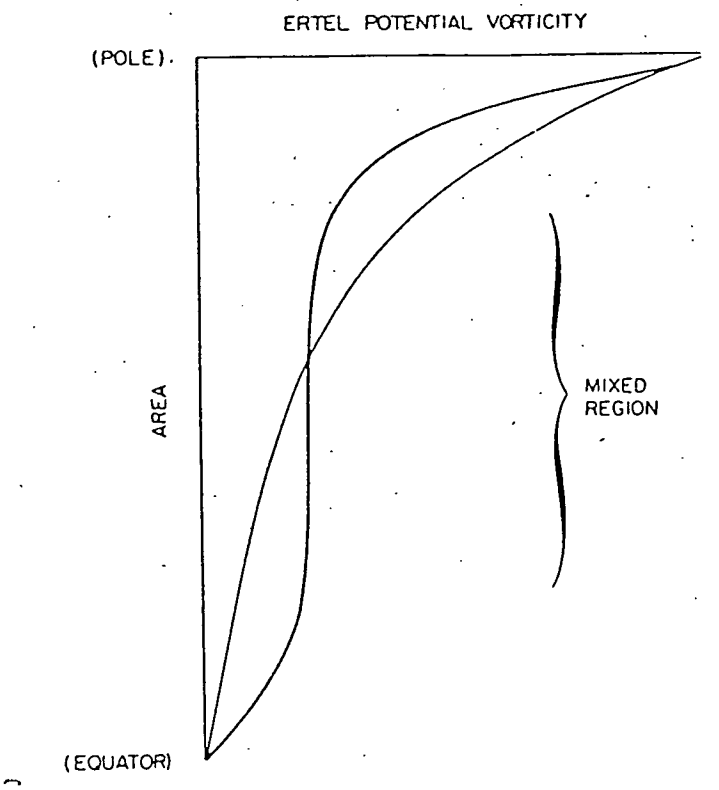


Figure 4.1: Schematic illustration of the isentropic distribution of potential vorticity in the middle stratosphere before (light curve) and after (heavy curve) quasi-horizontal mixing in the middle latitude region. The graph represents the area between the North Pole and the latitude at which a zonally symmetric potential vorticity contour would lie if planetary wave amplitudes diminished to zero. [From Butchart and Remsberg 1986].

---

plete mixing. According MP and Butchart and Remsberg (1986), the systematic redistribution depicted by the figure should also apply to other conservable tracers that have overall gradients in the wave breaking region. The erosion hypothesis

is, therefore, not only a possible description of evolution as seen on IPV maps and an explanation of the two-region structure, but also an important mechanism for determining the transport of pollutants and other tracers.

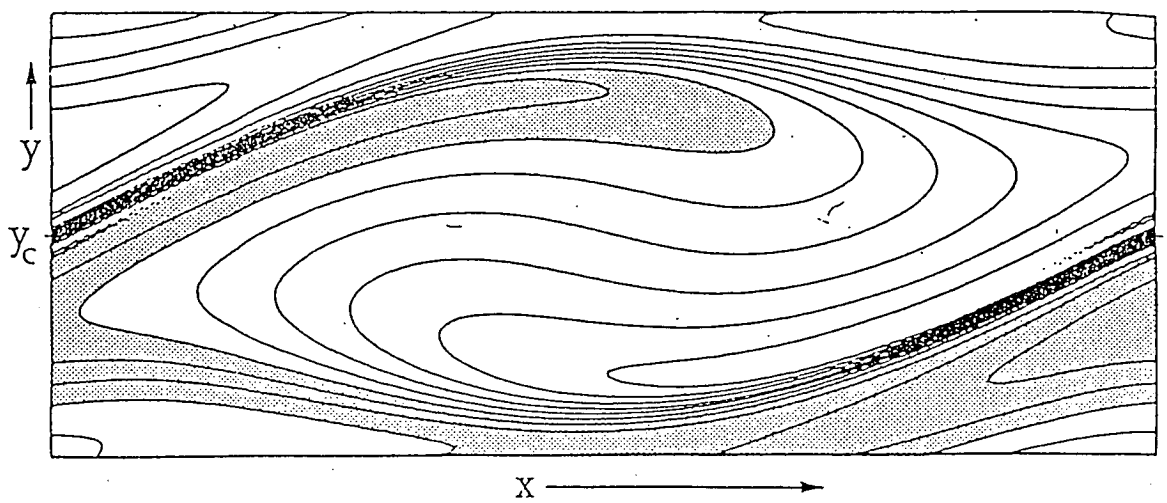
A striking feature of the two-region structure for transport of tracers is the vortex isolation phenomenon or so-called the 'impermeability' to tracers in the region of very strong PV gradients that mark the edge of the 'main vortex'. By analysing the results from high-resolution simulations, Haynes and McIntyre (1989) were able to show that the PV contour embedded in the edge of the 'main vortex' exhibits a peculiar resilience or elasticity that prevents it from being entrained into the turbulent-looking 'surf-zone' to either side. Instead, the contour undulates reversibly and is a material contour to very good approximation. This means that all the material enclosed in the 'main vortex' is chemically isolated from its surroundings and that transport of PV and long-lived chemical tracers is least, not greatest, in the region where PV gradients are greatest. Therefore besides the erosion mechanism, the processes involved in the maintenance and evolution of the 'main vortex', 'surf-zone' structure are complex. This will be further shown by our investigation.

4.3 Area of potential vorticity and its diagnostic formalism

It was suggested by McIntyre and Palmer that the area of main vortex on the isentropic surfaces, defined for instance as the area in which the PV is greater than a specified value, be used as an objective 'circulation index' for the winter stratosphere. They thought such an index can avoid the loss of information incurred by Eulerian space and time averaging in the presence of fluctuations in the position and shape of the main vortex and, therefore, provide a useful means of identifying when the irreversible deformation in the potential vorticity contours is taking place.

The definition of area also gives a way to investigate the erosion hypothesis itself by following the area evolution. For conservative motion, the isopleths of potential vorticity correspond to material contours and should enclose constant area on the isentropic surfaces if the velocity field is non-divergent in isentropic coordinates. The permanent change of area should be caused by either the diabatic heating/cooling or irreversible mixing into unresolvable scales by wave breaking. In reality, however, because of the limited resolution of observation and data, the observed distribution of potential vorticity is, at best, only an approximation to the true distribution. Thus, taking into account the real existence of friction and

Figure 4.2:



diabatic heating, the contours on IPV maps represent only approximate material contours, especially in the 'surf zone' region where overall gradients are weak. The true deformation of material contours like those given by the idealised model shown in figure 4.2 from McIntyre and Palmer (1983) may no longer be resolved by observation. Instead an average or smeared out value of potential vorticity is seen and there exists an apparent mixing of the low resolution distribution of potential vorticity. In the surf zone, even without any dissipation, the mixing of low resolution PV will manifest itself as a decrease of area enclosed by the

isopleths of highest values of potential vorticity. Furthermore for the irreversible contour deformation, the area change will be permanent. The Rossby wave breaking phenomena are then characterized by rapid irreversible contour deformation. The area diagnostics therefore, provide a way of investigating whether the wave breaking happens and the erosion hypothesis stands relevant.

Butchart and Remsberg (1986) were the first to exploit the idea of IPV area to investigate the dynamics involved in the 'main vortex' and 'surf zone' structure and wave breaking processes. They developed the relevant equations for diagnosing the area using real observational data. With LIMS data and neglecting the influence of wind divergence, they estimated the relative importance of the wave breaking and dissipative process and confirmed the prediction made by McIntyre and Palmer that the stratosphere warming should be accompanied by a very rapid decrease of the area of main vortex. They also found the similar change of area enclosed by contours of quasi-conservative chemical tracers, such as ozone and water vapour, and showed the striking similarity between PV and other chemical tracers.

The formula given by Butchart and Remsberg for the rate of change of the area enclosed by isopleths of potential vorticity can be written [see Butchart and Remsberg (1986) equation (13)]:

$$\frac{d}{dt}A(t)_{\chi\geq\chi_*} = \oint_\Gamma \mathbf{F}\frac{ds}{\mid \nabla_\theta \chi \mid} + \int_{A(t)_{\chi\geq\chi_*}} \nabla_\theta \cdot \mathbf{V} dA - \oint_\Gamma \mathbf{V}\cdot\nabla_\theta\chi'\frac{ds}{\mid \nabla_\theta \chi \mid} \qquad (4.37)$$

where $A(t)_{\chi\geq\chi_*}$ denotes the horizontal projection of the region for which $\chi(\mathrm{X},\theta_0,t) \geq$

χ_* on the isentropic surface $\theta = \theta_0$. The quantity **F** represents the effective sources and sinks of χ on the isentropic surface $\theta = \theta_0$. Γ is the curve $\chi(\mathrm{X}, \theta_0, t) = \chi_*$ and *ds* is an element of length along Γ. It should be noted that the prime in the above equation denotes the deviation of the observed values from the true values (e.g. $\chi_{true} = \chi + \chi'$), not the deviation from the mean values as conventionally used. Other symbols are conventional. The region $A(t)_{\chi \geq \chi_*}$ can be multiply connected.

As already mentioned, the basic assumption of Butchart and Remsberg (1986) is that the observed horizontal velocity field is non-divergent on an isentropic surface. This means the second term on the right hand side is neglected in their calculation. In the real atmosphere, however, the wind divergence does exist, though, may be small. That the nonlinear balance wind often becomes hyperbolic suggests the possible importance of divergent wind component (Since this equation neglects divergence in comparison with vorticity). Furthermore, the flux form of the potential vorticity equation on isentropic suface (Haynes and McIntyre 1987; Andrews et al. 1987) can be written:

$$\frac{\partial \sigma P}{\partial t} + \nabla \cdot L = 0 \tag{4.38}$$

where P is potential vorticity and L is a horizontal vector, $L = (\sigma P u - Y + Q v_\theta, \sigma P v + X - Q u_\theta, 0)$ being the flux of potential vorticity generalized to include nonconservative effects, with (X,Y) and Q being nonconservative force and diabatic heating respectively. Equation 4.38 reveals that it is the divergent part

of L which is important for the evolution of potential vorticity distribution. The conservative part of divergence of L is $\nabla \cdot (\sigma \vec{v} P) = P \nabla \cdot U_d + U_d \cdot \nabla P + (\sigma \vec{v})_r \cdot \nabla P$ where subscript r denotes the 'rotational' wind components or flux and d the 'divergent part', $U_d = (\sigma \vec{v})_d$. Taking into account the fact that the rotational flow is predominantly along contours of potential vorticity, so that the last term in the above expansion of dynamical part of divergence of L may be as small as first two, even if the rotational flow is much larger than divergent (Pawson and Harwood 1989). For a similar reason it may be argued that even if the divergence itself is small, the impact of it on the area change may be significant. This study is designed to investigate the role of the wind divergence in the area change, revealing its importance or otherwise confirming the assumption of Butchart and Remsberg (1986) for the reference of any future study.

It is necessary to mention that the seasonal evolution of IPV distribution due to the annual cycle of radiative forcing contributes an important part in the change of area enclosed by potential vorticity contours. Butchart and Remsberg (1986) presented a seasonal change of the area in response solely to the diabatic heating and small frictional damping from a model without rapid irreversible mixing and any eddy induced changes in the effective sources and sinks of PV. They showed that from late October to middle January the areas enclosed by PV values of less

than 2.5 unit are decreasing gradually; for values of larger than 2.5 the enclosed area increase in the same manner. From the middle of January the changes take just the opposite direction. There is no breaking of the vortex in this seasonal variation. They were also able to show that the large scale radiative processes due to planetary waves cannot be responsible for most of the observed reduction in size of the main vortex and expansion of the surf zone and therefore those rapid shrinking and breaking down of the polar vortex during the January and February 1979 should be attributed to other nonconservative processes and irreversible mixing.

4.4 Method of estimation

To estimate the relative importance of wind divergence on the rate of change of IPV area, two estimations are needed, the calculation of the real total rate of change and the rate of change caused by wind divergence. For estimating the real total rate of change on an isentropic surface, we firstly need to know the potential vorticity on the surface. Unlike Butchart and Remsberg (1986), where they calculated the relative vorticity on the isobaric coordinates first and then interpolated into the isentropic surface, our computation is directly performed on the isentropic surface. This is rather simpler and more straight forward.

In isentropic coordinates Ertel potential vorticity is given by:

$$P = \zeta_\theta / \sigma \tag{4.39}$$

where

$$\zeta_\theta = f - \frac{(u\cos(\phi))_\phi}{a\cos(\phi)} + \frac{v_\lambda}{a\cos(\phi)} \tag{4.40}$$

is the vertical component of absolute isentropic vorticity and σ is the density in isentropic coordinates as defined in chapter 3. Using the procedure in chapter 3 we firstly get the Montgomery potentials on the related isentropic surfaces and find u,v by quasi-geostrophic approximation discussed also in chapter 3, leading to the knowledge of the ζ_θ field. Again making use of method given in section 2 of chapter 3, we obtain the density field. Finally by equation (4.39) the potential vorticity field is obtained on a specific isentropic surface.

As the analysis is done on a latitude-longitude grid, each grid point is given an area-weight proportional to the cosine of latitude. Integration of the elementary area for all the grid points within a specific contour of potential vorticity then gives the whole area enclosed by the contour. From time sequences of area of successive days, the real total change rate of the area within each contour is obtained.

For the reason of compatibility and comparison, the same data and level as those used by Butchart and Remsberg (1986) is chosen, That is, the LIMS data is used and calculation is pursued on the 850K level. Figure 4.3a,b shows the area

curves by Butchart and Remsberg (1986) and ours respectively on the same day and same level with the same data. The units of area are surface area expressed as a fraction of one hemisphere and those of potential vorticity, P_*, following Butchart and Remsberg (1986), are $gH_0p_0{}^{-1}10^{-4}Km^{-1}s^{-1}$; g is the gravitational acceleration, p_0 a standard pressure (1000mb) and H_0 a standard pressure scale height taken to be 7000m. The label "equivalent latitude" represents the latitude at which a zonally symmetric contour P_* would lie if it enclosed area $A_{P\geq P_*}$. It can be seen that except near the very north and south boundary both figures are actually exactly the same. Therefore we can have confidence in the estimation of IPV area by both groups. At least our estimation is in the same accuracy as Butchart and Remsberg's (1986). The accuracy of the real change of the area should be also the same order.

To estimate the rate of change of the area caused by wind divergence, i.e. the second term on the right hand side of equation (4.37), a prerequisite is to calculate the wind divergence. The principle of getting divergence wind is given by Pawson and Harwood (1988). Combining the thermodynamic equation and the mass continuity equation on isentropic coordinates (Pawson and Harwood,1988), the wind divergence can be given by:

$$\nabla \cdot \mathbf{V} = -\frac{1}{\sigma_\eta}\left(\frac{\partial \sigma_\eta}{\partial t} + \frac{\partial(\sigma_\eta w)}{\partial \eta} + \mathbf{V}\cdot\nabla\sigma_\eta\right) \tag{4.41}$$

where $\eta = \log(\theta/\theta_0)$ taken as vertical coordinate. θ is potential temperature with

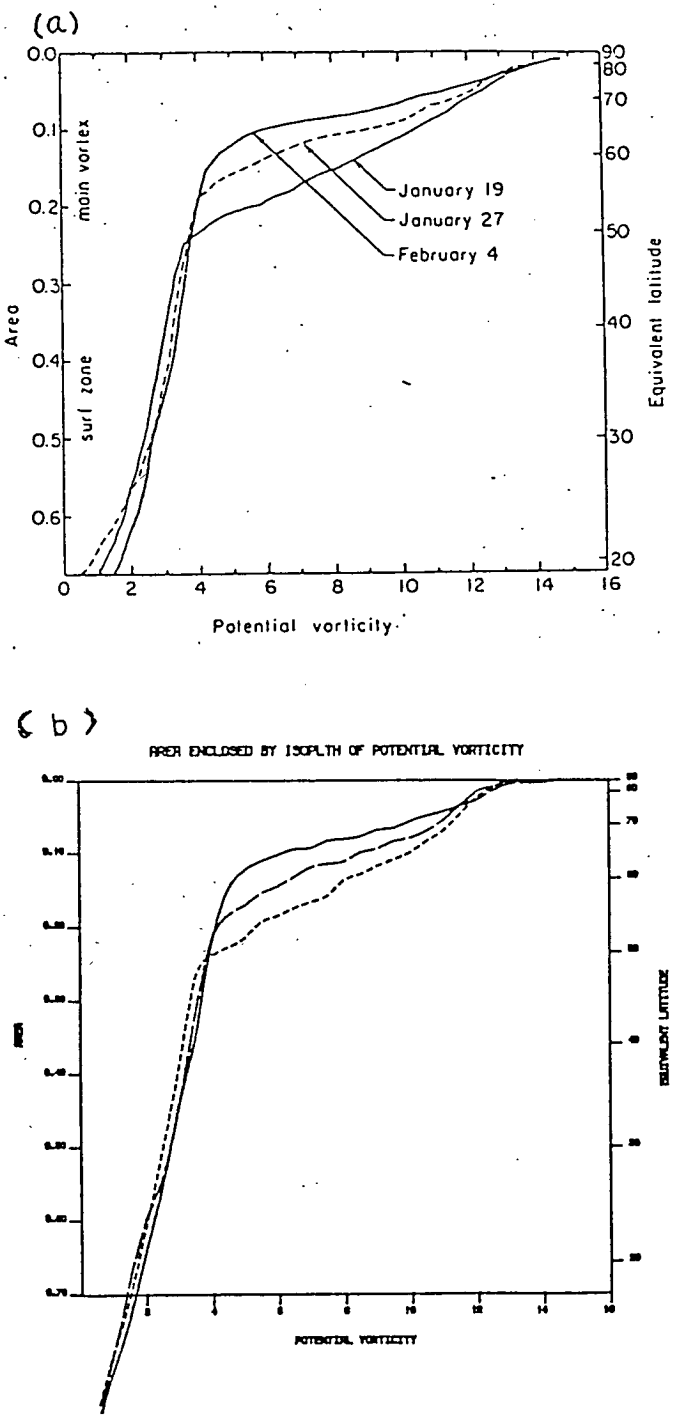


Figure 4.3: Plot of $A_{P \geq P_*}$ against P_* (see text for the definition of 'equivalent latitude') for 19 January, 27 January and 4 February. The unit of area are surface area of one hemisphere and those of potential vorticity,P_* are $gH_0p_0^{-1}10^{-4}Km^{-1}s^{-1}$ (see text for the meaning of symbol). (a) From Butchart and Remsberg (1986) and (b) from our own calculation for comparison. In (b) the solid line is for 4 Feb. long dash line for 27 Jan. and short dash line for 19 Jan.

$\theta_0 = 400$ as a reference level. **V** is the horizontal wind vector. σ is the density in the isentropic coordinates. w is vertical velocity decided solely by the diabatic heating, i.e. from thermodynamic equation:

$$w = D\eta/Dt = J/(c_p T) \tag{4.42}$$

D/Dt represent the material derivative; J/c_p is the net heating rate (expressed as rate of change of temperature) and is obtained from the same code as Pawson (1988) i.e. Haigh (1984).

In their study Pawson and Harwood (1988) were able to neglect the transient $(\partial/\partial t)$ term in equation (4.41) as it is small in the monthly mean. Here however it must be retained as we are calculating daily values. In the present study only the wind divergence is needed not the wind itself, so the computation is even simpler.

Integrating the value $\nabla \cdot V dA$ for all the points within a specific contour P_* gives the change rate of the area enclosed by that contour caused by wind divergence.

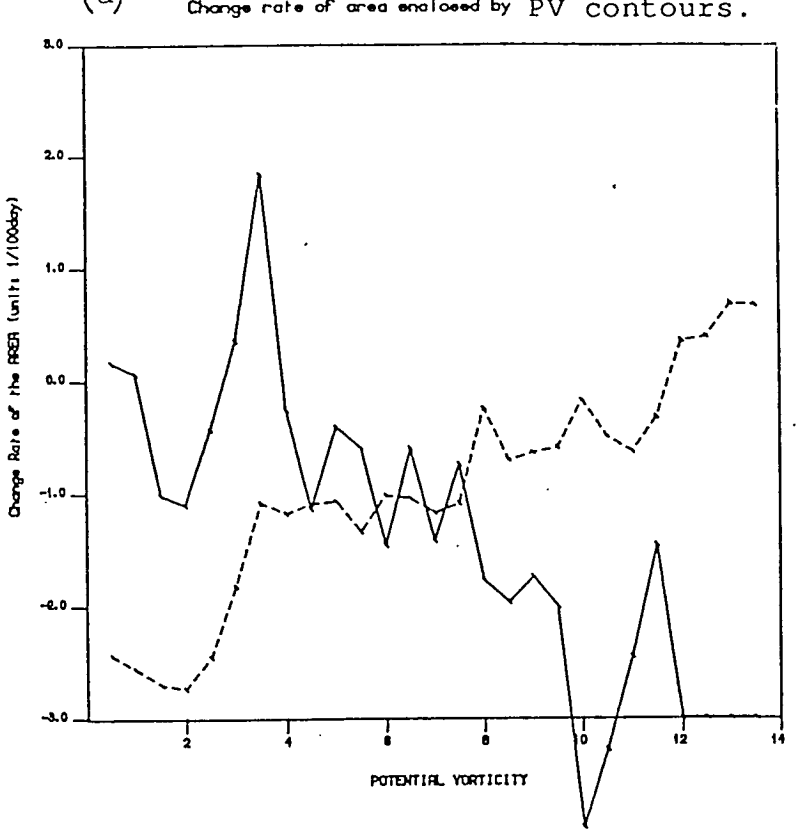
4.5 Results and discussions

Using the procedure described above, we obtain the real rate of change of total area and the rate of change of the area caused by wind divergence for each day of a period from 20 January to 28 February 1979. Since the area is a monotonic

function of the values of contours of potential vorticity where the area value increases with decreasing contour values of potential vorticity, the original rates of change of area appear relatively very large for low PV values and very small for high PV values and this makes it hard to plot and analyse the rate of change, especially to plot the curves of the two kinds of rates of change on a specific day. On the other hand, because our purpose is to compare the relative importance of rate of change caused by wind divergence with the observed rate of total change, a 'weighting' of the two sets by a same function will not destroy any sense of comparison. Therefore weighted values are used in plotting the curves for certain days (e.g. figure 4.4) with both weighted and unweighted values plotted for time evolution analysis (figure 4.6 and 4.5). The weighted values are also plotted in time evolution maps for easier comparison. The weighting function is chosen as $1/(1-\sin\varphi)$, where φ is a set of chosen latitudes,in which the northern hemisphere are equally divided into 32 zones. Taking the whole area of the north hemisphere as 1, this function is equivalent to dividing the change-rate of area enclosed by each IPV contour by the area enclosed by a chosen latitude circle and the pole. It is very similar to $\frac{1}{A}\frac{\partial A}{\partial t}$ but with the equivalent latitude being replaced by a chosen latitude. The direct consequence of this weighting is that the contrast between the smaller rate of change at large IPV values and the large rate of change at small IPV value is reduced, without losing any of the original features of the distribution for comparison (comparison of figure 4.5 with 4.6 is discussed further below).

Figure 4.4a,b,c and d show the curves of weighted rates of area change drawn against their corresponding contours of potential vorticity for 24,27 January and 3, 22 February 1979 respectively with solid line showing the observed rate of change and the dashed line the rate of change of area due to wind divergence. It is worth noticing that figure 4.4a,c and d are before the stratospheric warmings while figure 4.4b is after the warmings. Figure 4.4b shows that on 27 January 1979 there is a decrease of area enclosed by contours of 4–10 unit and an increase of area by those contours of less than 4. Among the change, wind divergence contributes only a small portion, generally less than 30 percent of the total change. The other curves on the days well before and after the warming also have similar signatures with relatively small divergence contributions. Thus these days suggest that in normal stratospheric conditions in winter (not during stratospheric warmings) the effect of divergence on the polar vortex is negligible as Butchart and Remsberg (1986) assumed.

However before the stratospheric warmings as shown by figure 4.4a,c,d, there are signatures in which divergence-caused rates of area change are comparable with the observed rate of change. For example, in figure 4.4a for the area enclosed by contour labels 4–7; in figure 4.4c 8–10 and figure 4.4d 8–10; the rates of change due to divergence are of the same order as the observed rates of change. Furthermore in figure 4.4c and 4.4d both rates of change are of similar trend. These may suggest



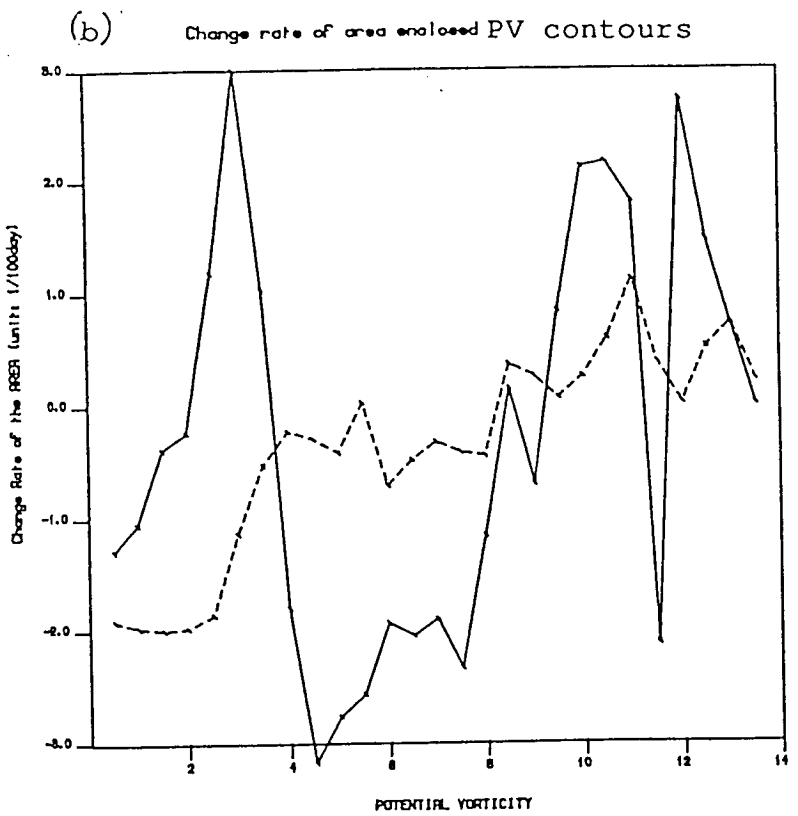
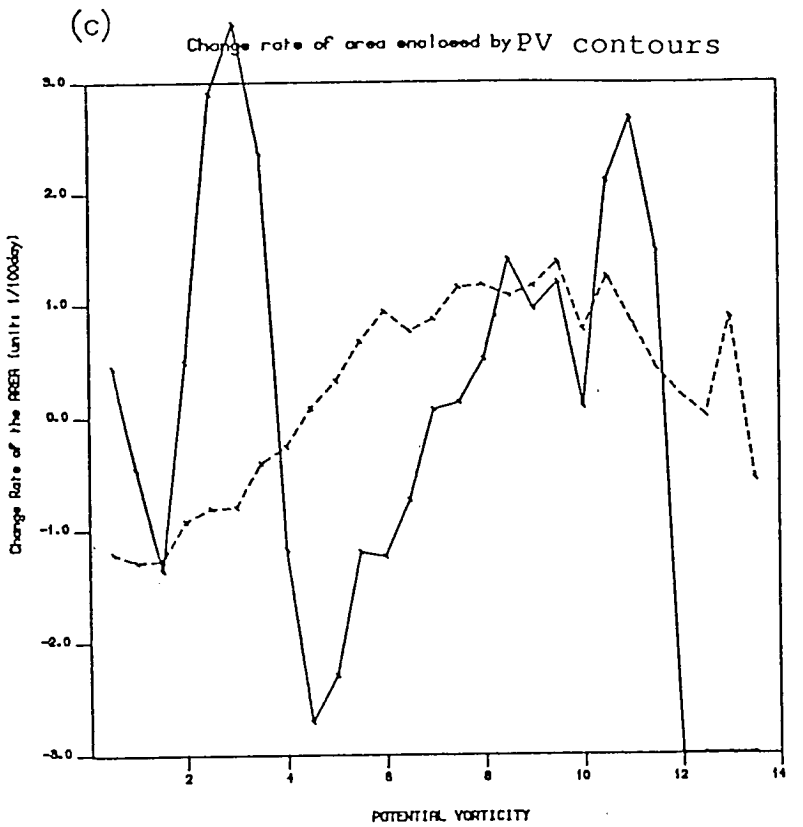


Figure 4.4: Change rate of area enclosed by potential vorticity contours. The change rate curves are drawn against corresponding IPV contour value. Solid line denote the calculated real rate of change; dash line the rate of change caused by wind divergence. (a) On 24 January, (b) on 27 January, (c) on 3 February, (d) on 22 February, 1979.



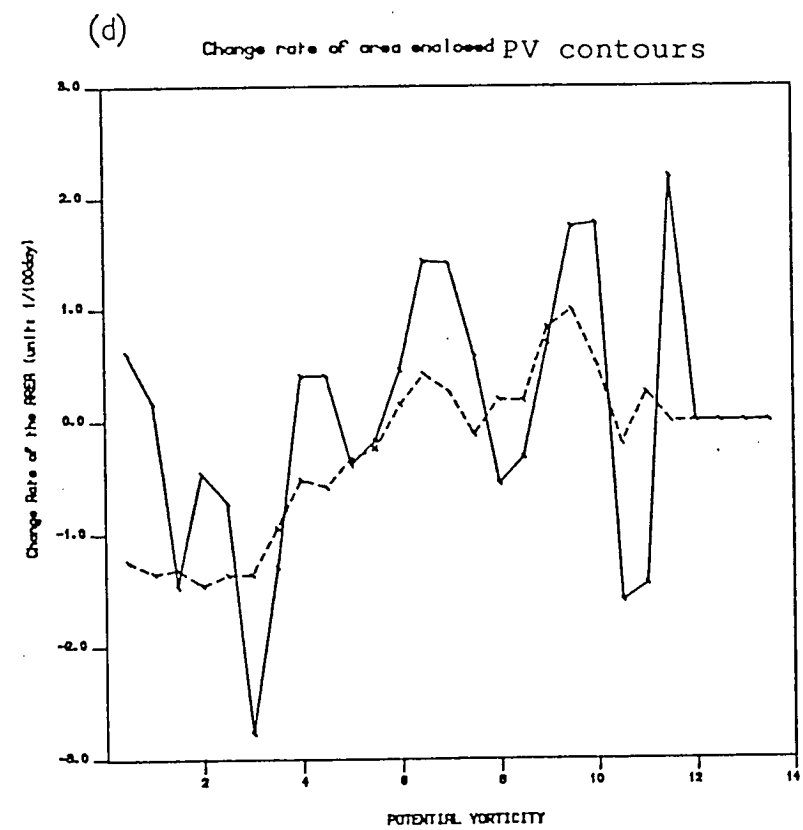


Figure 4.4: (figure continued)

that before special circulations like stratospheric warmings the contribution from divergence to the change of PV area may be significant.

To have an overall picture of the observed change-rate of area and rate of change due to wind divergence, rather than a few isolated occasions, the two terms are computed for the successive days in the period from 20 January to 28 February 1979. Figure 4.5a,b show the time–cross section of observed rate of change and c,d show the rate of change caused by wind divergence in the same period. Both are plotted in their original values (not weighted). From figure 4.5a,b we see that generally the area enclosed by IPV value of larger than 4 decreasing as the winter progresses (negative rate of change); whereas the area inside the lower IPV values increasing in the period (positive values). These features are expected and agree with those presented by Butchart and Remsberg (1986). On inspecting the rate of change of IPV area caused by wind divergence in figure 4.5c,d it seems that in the 'main vortex' region the wind divergence generally causes an increase of the area and in the 'surf zone' region it causes a decrease. The rate of change is generally small, normally less than 30 percent of those rates of observed changes except the three patches which will be discussed later. As indicated previously, the real change is a result of a whole set of processes acting in the atmosphere which include the wind divergence. Despite the contrary contribution from wind divergence, the fact that the observed rate of change still has the features described above

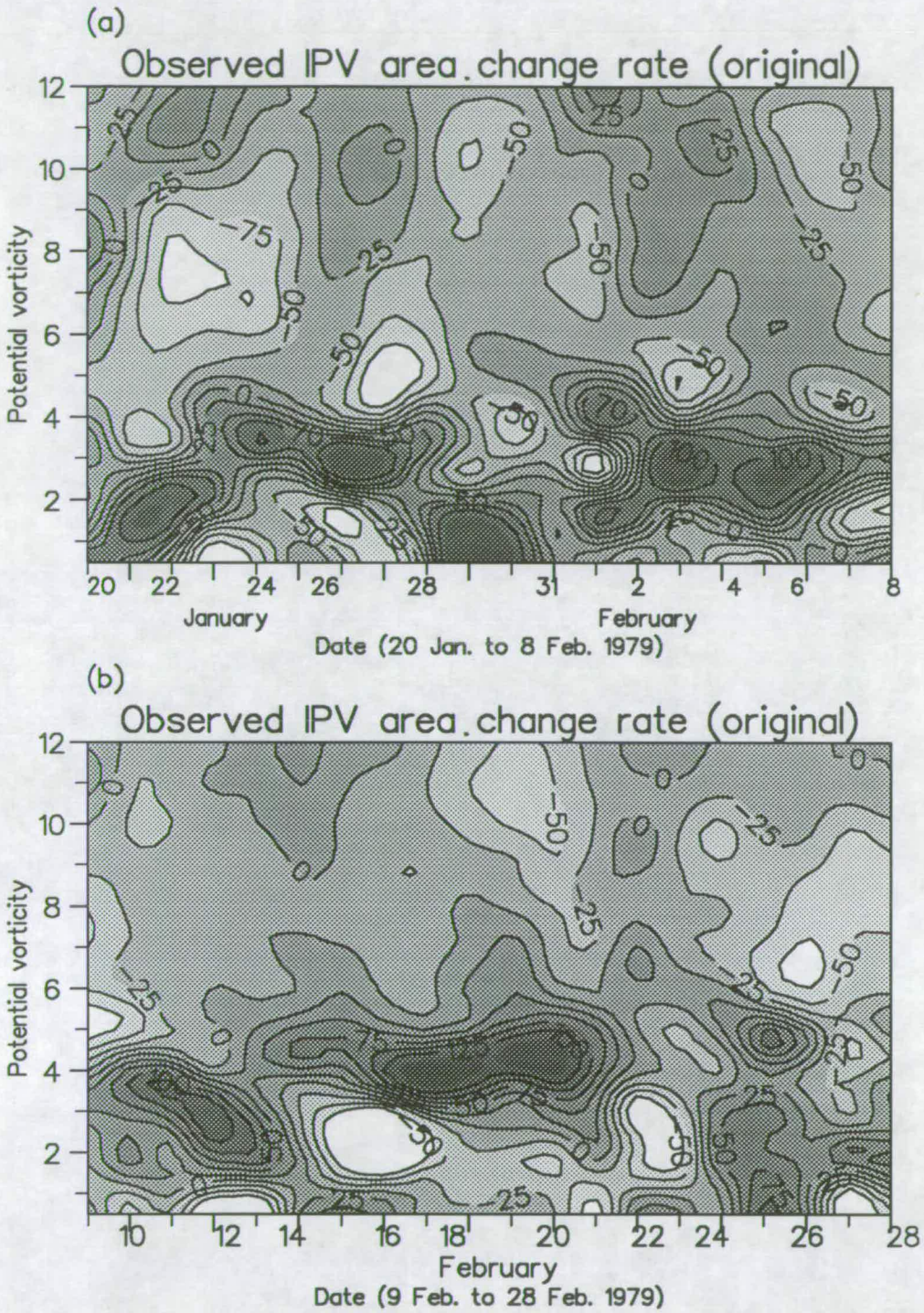


Figure 4.5a,b Time–PV cross section of observed change rate rate of area enclosed by isopleths of PV on 850K isentropic surface. (a) from 20 Jan. to 8 Feb. (b) from 9 Feb. to 18 Feb. 1979. Unit: 1/10000day.

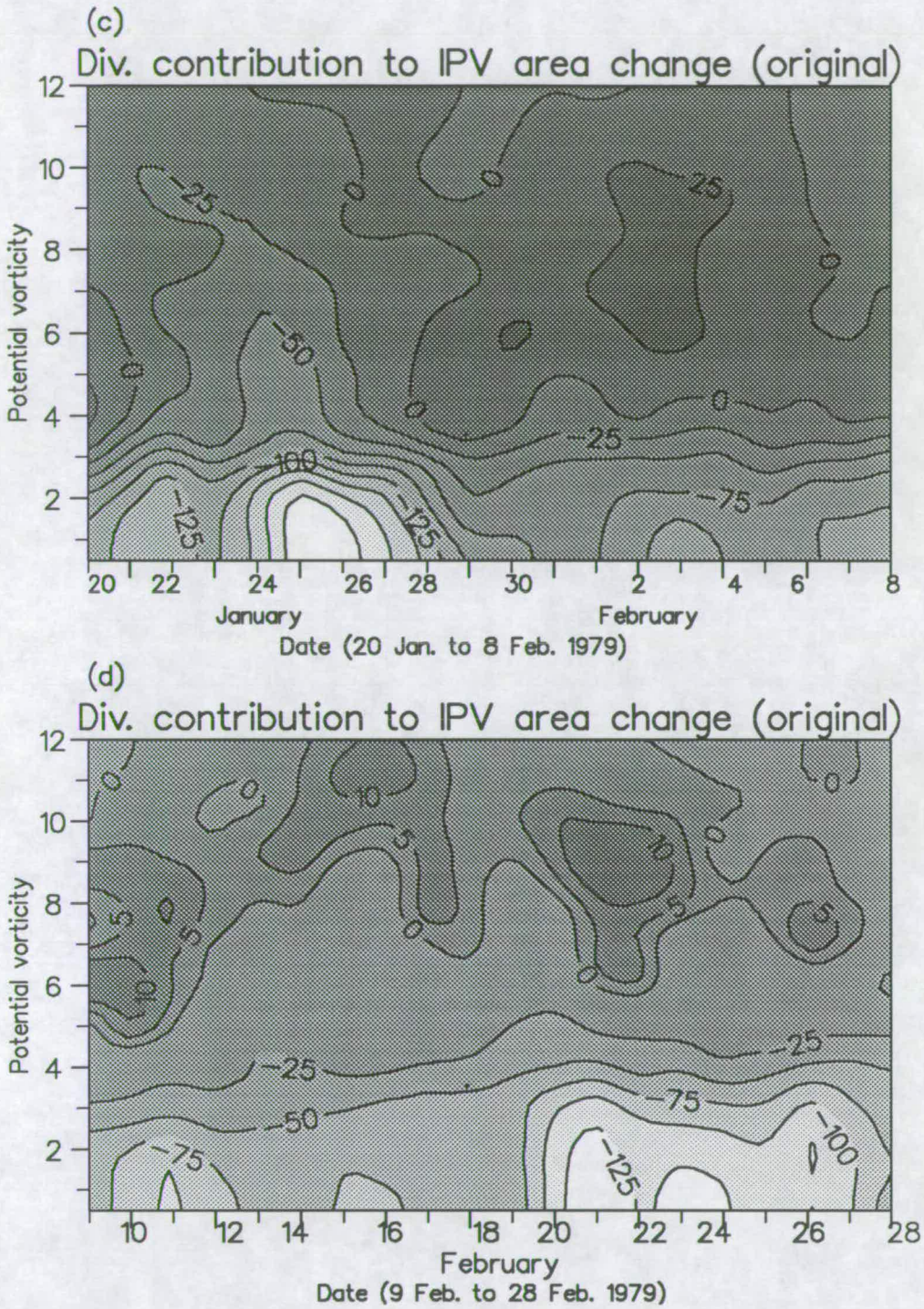


Figure 4.5c,d Time–PV cross section of change rate of IPV area caused by wind divergence. (c) from 20 Jan. to 8 Feb. 1979. (d) from 9 Feb. to 28 Feb. 1979. Unit: 1/10000day.

should indicate the general dominance of erosion mechanism over divergence in the most periods (except immediately before stratospheric warming) of stratosphere in winter. The results are consistent with those found by Butchart and Remsberg (1986).

On the other hand, detailed comparison of figure 4.5a,b with c,d shows that there are three patches where the change caused by wind divergence is comparable to the real change. The first one is around 25 January for IPV values of 4–7 unit, where the observed rate of change is around -25 unit, about the same magnitude as the rate of change caused by wind divergence. This is consistent with figure 4.4a. The second is around 3 February for IPV value of 8–10 unit. Both rates of changes are positive and about the same magnitude, and this feature is again confirmed by figure 4.4c as discussed previously. The third patch is in 22-23 February for IPV values of 8–10 and also shown by figure 4.4d. In these patches the contribution to the total real change from wind divergence is either compatible to the real change or about 50 percent of it. All these features can also be found more easily in figure 4.6, which is the weighted equivalent of figure 4.5. As we mentioned, the three patches occur just before the stratospheric warmings. The disruption during warmings could be expected to give higher heating rates and therefore the larger divergence. Thus, the contribution of the divergence of wind to the change of area enclosed by IPV contours is no longer negligible in the

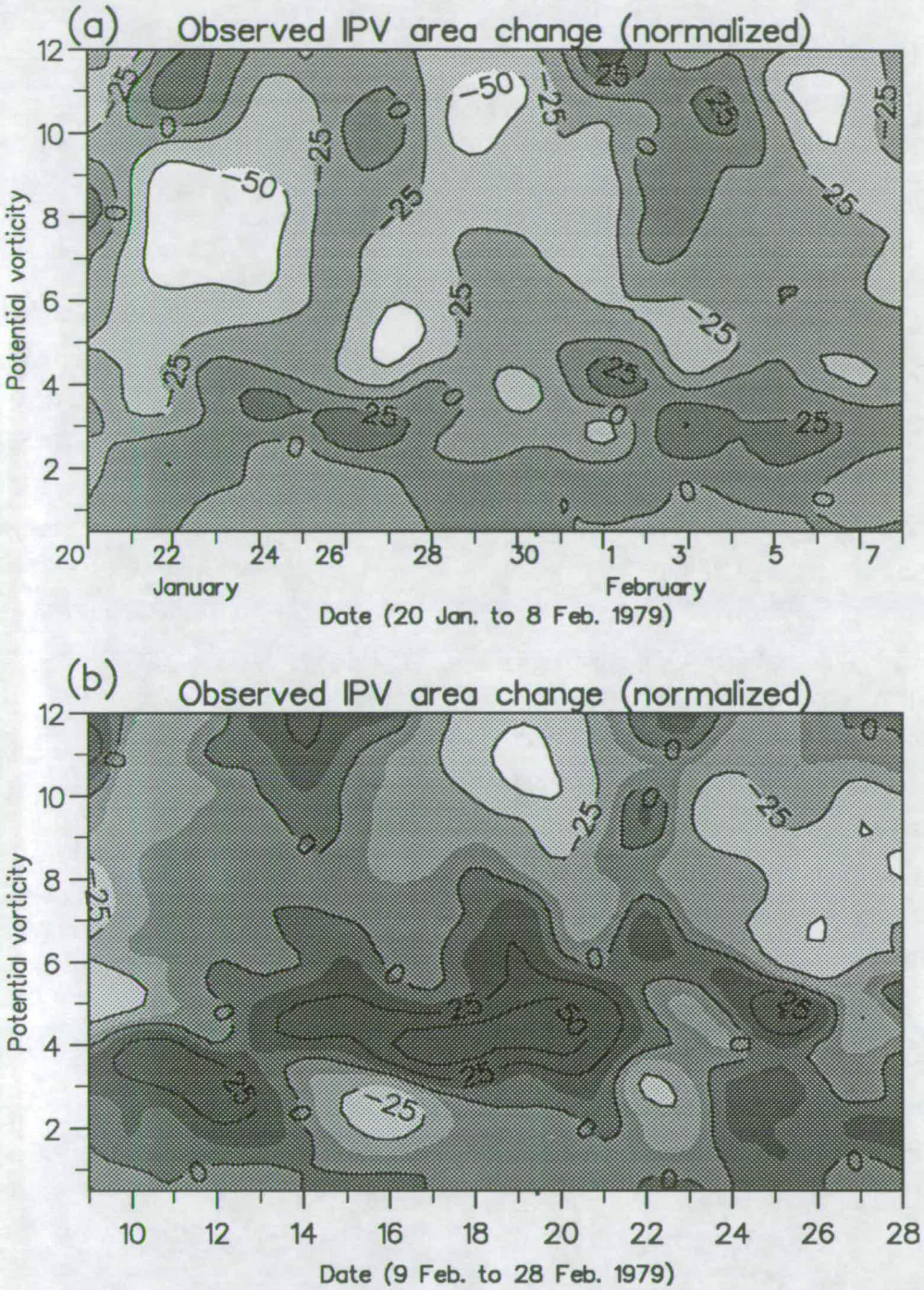


Figure 4.6a,b Time–PV cross section of observed change rate of area enclosed by isopleth of potential vorticity on 850K isentropic surface. (a) from 20 Jan. to Feb.8. (b) from 9 Feb. to 18 Feb. 1979. The number is normalized (see text) with Unit: 1/10000day Unit for potential vorticity is as those of Butchart et al (see text).

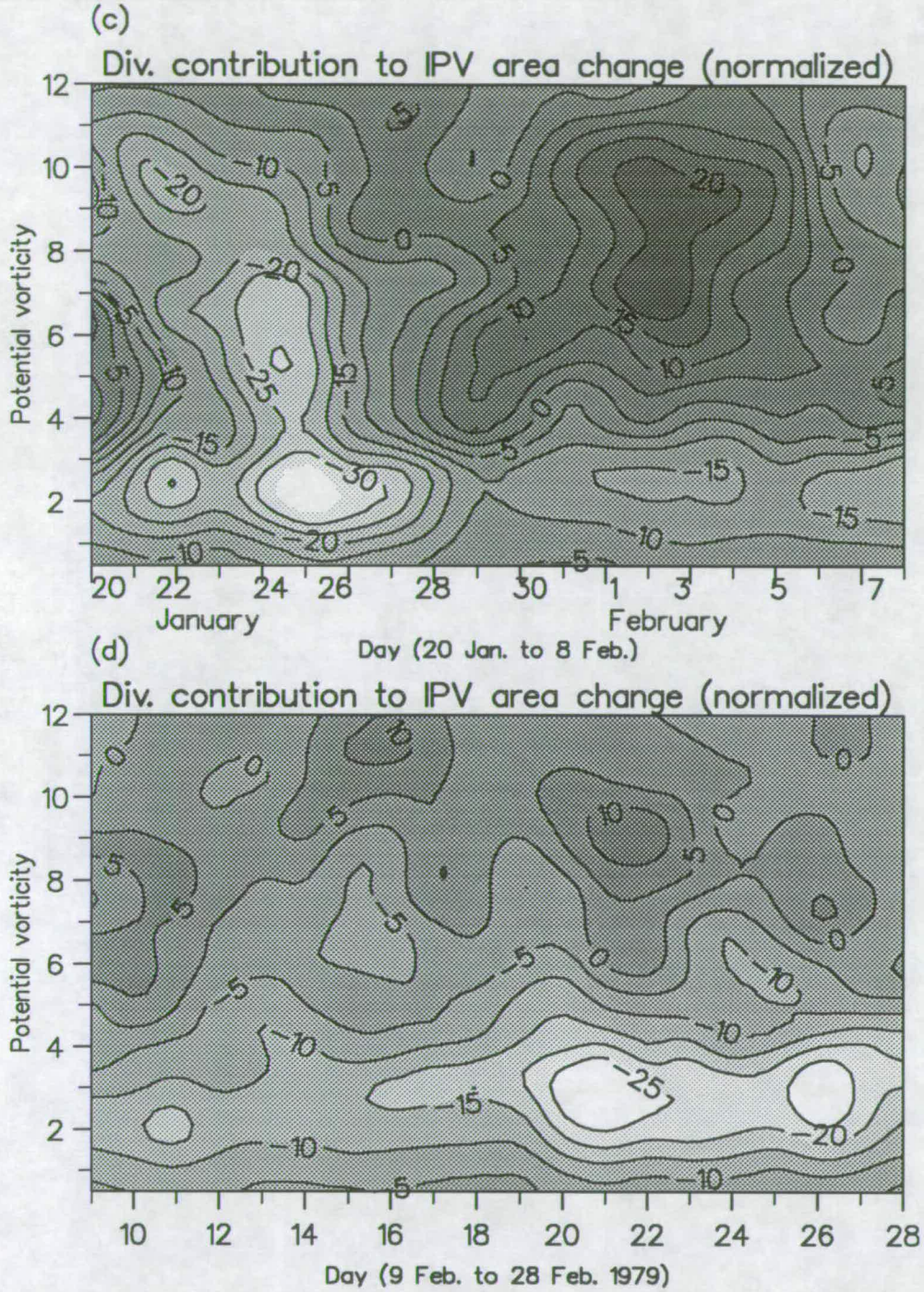


Figure 4.6c,d Time–PV cross section of change rate of IPV area (see text) caused by wind divergence. (c) For period from 20 Jan. to 8 Feb. 1979 (d) from 9 Feb. to 28 Feb. 1979 The number is Normallized (see text) with unit of 1/10000day

occasions discussed above and the attention should be paid to the effects of wind divergence when applying the 'area index' for dynamical analysis around the time of a stratospheric warming.

In summary our present investigation generally support the assumption made in the paper of Butchart and Remsberg (1986) and their results about the area change in the normal situation of winter stratospheric circulation. At least for the LIMS data on 1979 from our calculation, we can say that the contribution from wind divergence to the change of area enclosed by isopleths of potential vorticity on isentropic surface is generally small and opposite in sign to the observed change in the normal situations. Thus the erosion processes proposed by McIntyre and Palmer generally dominate over divergence effects in most of stratospheric circulations and the neglect by Butchart and Remsberg (1986) of the impact of wind divergence on the change of area when using the 'index' to quantify the detailed dynamics involved in the erosion processes is therefore mostly justified by our investigation. However our investigation also indicated that immediately before stratospheric warmings, the wind divergence does have a significant contribution to the change of the area. Attention should be paid to effect of wind divergence when applying the area 'index' to analyse the mechanism involved in the 'main vortex' ,'surf-zone' structure during stratospheric warmings. In the warmings, the area change may be a result of response to nonconservative processes and/or

irreversible mixing to unresolvable scale, and it may also be caused significantly by wind divergence. The mechanism involved in the evolution of 'surf-zone','main vortex' structure during a stratospheric warming may be a mixture of erosion and wind divergence and be very complicated.

Chapter 5

The Observed Daily Variation of Wave Activity in The Northern Hemisphere Stratosphere for Autumn 1978

5.1 Introduction

Since the pioneering work of Charney and Drazin (1961) and Eliassen and Palm (1960) a growing body of theory has sought to elucidate the relationship between changes in the zonal mean state of the atmosphere and the prevalence of waves. The wave activity, A, and the divergence of the Eliassen–Palm (hereafter E–P) flux, F, are key concepts in this theory, $\nabla \cdot F$ being the only term or the principal term (according to the approximating assumptions made) in the equations gov-

erning the rate of change of the mean state (see for example Edmon et. al. 1980), whilst A and $\nabla \cdot F$ are related for small amplitude waves by the generalized E–P flux theorem (Andrews and McIntyre 1976) of equation 2.11.

A major difficulty in applying these ideas in observational studies is that the fundamental definition of the wave activity involves the displacements of air particles from some undisturbed basic or 'initial' state. Finding the particle displacements by constructing trajectories is extremely difficult. Moreover, for the real atmosphere, as opposed to that of an idealised mathematical model, it is far from clear what should be regarded as the basic state. In the hope of alleviating these difficulties, Andrews (1987) has developed some alternative expressions for the wave activity which do not involve particle displacements. The aim of the present study is to attempt to evaluate the usefulness of these expressions by employing them in a diagnostic study of the terms in the generalized E–P flux theorem. As the theorem is a small amplitude result attention is here initially restricted to a period in autumn when the stratosphere is relatively quiescent. An attempt to apply the expressions in more disturbed conditions was less satisfactory because the gradient of zonal mean potential vorticity which appears in the denominator of one of the expressions passed through zero giving singularities. This reversal of the gradient is in any case associated with finite amplitude effects, so the expressions may well be unsatisfactory under such conditions on those grounds also. Accordingly only

the quiescent case is described below.

It should be mentioned that Marks (1988) also carried out a similar study except that the isobaric version of definition of wave activity density and other terms was used. He got a noisy field of wave activity due to the frequent change of sign of gradients of isobaric potential vorticity. Here this problem is prevented by paying more attention to the conditions which the theorem required, namely small wave amplitudes.

The next section sets out the theoretical framework and section 3 deals with the description of data used, computational details and some discussion of errors. In section 4 we discuss the observed wave activity, its spatial distribution and temporal variation. Section 5 analyses the relations among the terms in the generalized E-P flux theorem itself and its relation with changes in the zonal mean flow. An alternative expression for wave activity, defined by substituting ozone mixing ratio for potential vorticity in the theorem, is presented in section 6 and the relation between the both kinds of wave activity discussed. Section 7 describes a diagnosis of how the balance in the theorem, equation 2.11, is maintained. Finally section 8 contains the discussion and conclusions of this chapter.

5.2 Theoretical basis

In this section, the expressions used to calculate the wave activity and the E–P flux are given, together with a brief discussion of the significance of the quantities.

Traditionally when discussing waves on a zonal mean flow, it is common to show diagrams of the amplitude and phase of separate zonal harmonics. It has been realized for some time that such quantities as geopotential amplitude or wave energy density based on Fourier expansion have shortcomings as measures of the strength of waves, for example the geopotential amplitude tends to be large in regions of strong zonal jets, irrespective of the detailed dynamics of the wave propagation (Andrews, 1987). There have been many attempts to investigate the wave-mean flow interaction in terms of the linear theory of forced planetary waves on a zonally symmetric basic state. In order to extract more useful dynamical information on the waves it is important to have a 'conservable' measure of wave amplitude (McIntyre, 1982). Several authors (e.g. Andrews and McIntyre 1976; Edmon et. al. 1980; Palmer (1982); Andrews (1987)) have summarized the main advantages of studying wave–mean flow interactions in terms of the Eliassen–Palm flux as already discussed in section 2.3 of chapter 2.

Many of the relationships between waves and the mean flow become clear when the theory is developed using a Lagrangian coordinate system and the general-

ized Lagrangian mean (Andrews and McIntyre 1976). To the extent that stratospheric flow can be treated as adiabatic, potential temperature can be used as a Lagrangian vertical coordinate with several advantages over pressure or log-pressure. For instance the Charney-Drazin non-acceleration theorem is a finite amplitude result in these coordinates (Andrews 1983). Even when diabatic effects cannot be ignored, the simple relationship between heating and vertical velocity confers advantages by the use of entropy as the vertical coordinate (e.g. Tung 1986). Accordingly η is used as vertical coordinate, defined as $\eta = \ln(\theta/\theta_o)$, where $\theta_o = 400K$ is introduced to avoid any dimensional ambiguity.

The E-P flux theorem was introduced in equation 2.11 but it is reproduced here for convenience.

$$\frac{\partial A}{\partial t} + \nabla \cdot F = D + 0(\alpha^3) \tag{5.43}$$

The definitions of terms in the above theorem in isentropic coordinates are presented by Andrews (1987) as shown in equations 5.44 to 5.47 below.

E–P flux divergence is

$$\nabla \cdot F = \frac{[F^{(\phi)} cos(\phi)]_\phi}{a cos(\phi)} + F_\theta^{(\theta)} \tag{5.44}$$

where $F^{(\phi)}$ and $F^{(\theta)}$ are horizontal and vertical components of E–P flux in isentropic surfaces and are defined as:

$$F^{(\phi)} = -a\overline{\sigma}\overline{u'v'}\cos(\phi);$$

$$F^{(\theta)} = \frac{1}{g}\overline{p'M'_\lambda} \tag{5.45}$$

The wave activity density is defined by

$$A = a\cos(\phi)[\frac{1}{2}a\overline{\sigma}^2\overline{P'^2}/\overline{P}_\phi - \overline{\sigma'u'}] \tag{5.46}$$

and the nonconservative term is defined by

$$D = a\cos(\phi)[a\overline{\sigma}^2\overline{S'P'}/\overline{P}_\phi - \overline{\sigma'X'_1} + \overline{u'(\sigma Q')}_\theta \tag{5.47}$$

The notation is mostly standard with $\sigma \equiv g^{-1}\partial p/\partial\theta$ being the density in (λ, ϕ, θ) space, g the gravity, X'_1 the non-conservative term in the isentropic zonal mean momentum equation; Q is the diabatic heating rate and P the Ertel potential vorticity (PV) in the isentropic coordinates:

$$P = \frac{\zeta}{\sigma} = [f - \frac{(u\cos(\phi))_\phi}{a\cos(\phi)} + \frac{v_\lambda}{a\cos(\phi)}]/\sigma \tag{5.48}$$

P' is the perturbation to PV, satisfying the linearized equation:

$$D_tP' + v'\overline{P}_\phi/a = S' \tag{5.49}$$

S' being the corresponding non-conservative term. $D_t \equiv \partial/\partial t + \overline{u}\partial/a\cos(\phi)\partial\lambda$ is the material derivative. As discussed in detail both theoretically and practically in the next chapter, the Ertel potential vorticity in 5.46 can be replaced with some minor

approximations by ozone mixing ratio as an alternative wave activity density which can be seen both as a check of our present calculation and having some implications for tracer transport.

The following sections will analyse the wave property derived from LIMS data using equation (5.43) to (5.49) as basic formulation.

5.3 Data, computational details and error analysis

5.3.1 Data for wave activity calculations

Synoptically mapped LIMS measurement of temperature, geopotential height, ozone and water vapour (see Gille and Russell 1984) are used in the present work. The geopotential heights are used in the calculation of winds, while the other quantities are needed in the computation of the radiative damping of wave activity in the non-conservative term.

As stated in section 1, we have confined our study to a period of small wave amplitude when the atmosphere is most likely to be near the linear regime required by the theory. The period chosen is from 26 October to 16 November 1978. An attempt to extend the application of the theorem and related quantities in their

original form to an analysis of a period of large wave amplitude such as those during the stratospheric sudden warmings did not produce meaningful results due to the singularity problem caused by the changing of sign of the latitudinal gradient of zonal mean Ertel potential vorticity in some latitudes. This reflects the complex nature of the real atmosphere.

To compute the non-conservative or wave damping term, D, expressed by 5.47, diabatic heating rates are also needed. As the LIMS data ranges only from 100mb to 0.05mb, the most recent CIRA temperature and ozone climatologies (Labitzke and Barnett 1985; Barnett and Corney 1985), supplemented with monthly zonal mean SBUV ozone data, are used to extend the LIMS data to the troposphere and mesosphere. The diabatic heating rates are then obtained after Haigh (1984) and Strobel (1978), by the same code as Pawson and Harwood (1989).

5.3.2 Montgomery potential, winds and Potential vorticity

As stated in section 5.1, $\eta = \ln(\theta/\theta_0)$ is taken as vertical coordinate. η is taken from 0.0 to 1.3, from above 100mb to nearly 1mb with equal interval of 0.13 in the calculation. A regular latitude–longitude grid is used with 4° spacing in latitude and 15° in longitude in accordance with LIMS data.

The Montgomery potential on a given isentropic surface is calculated by the method discussed in chapter 3. Winds are found using the quasi-geostrophic approximation also described in chapter 3.

The density in our η coordinate is related to the vertical coordinate by $\sigma_\eta \equiv \rho(\partial z/\partial\eta) = -g^{-1}(\partial p/\partial\eta)$, where hydrostatic balance is assumed and again calculated by the method given in the section 2 of chapter 3.

Ertel potential vorticity is computed from the definition, 5.48 by the same scheme as described in chapter 4. The accuracy of this method is also discussed there. The zonal average of Ertel potential vorticity is then calculated directly by averaging over the field obtained from 5.48.

The X_1' term in equation 5.47 which represents the frictional effects or drags (e.g. by gravity waves) in the zonal momentum equation on isentropic surface (see B.11 of Andrews,1987) is omitted because its specific expression is unkown and in the lower to middle stratosphere it is believed to be small. S' in the first term of D is calculated from equation 5.49. The third term in equation (5.47) is found to be smaller than the first term. The neglected second term may be important in high altitude.

5.3.3 Error analysis

Following the above procedure and using 5.44-5.47, it is straight forward now to calculate each term in the theorem expressed by equation 5.43. On the other hand, even though the terms can be numerically calculated, for such highly derived quantities, how accurate they are or how large the error may be, still remains a difficult open question. Errors could stem from the limited vertical and horizontal resolution of the LIMS data which may not be satisfactory for calculating such a highly derived quantity; and could also come from the finite-difference routines used to calculate the quantity. Yet again, errors could be caused by the retrieval of the data from satellite observed radiance. It is not possible in our present scope to investigate in detail how much error each error source may cause to each term in the theorem. A practical way to analyse the error is to impose some random error to the data, then see how sensitively the resultant quantities react. If the resultant quantities are very sensitive to the error in the data, any analysis based on such computation will have to be treated suspiciously and cautiously, since any feature therefrom may be the reflection of error rather than of the atmosphere itself. Otherwise if the resultant quantities are insensitive to the errors induced, further analysis may be pursued.

In the present study 5% random error is introduced to the wave coefficients of every field of LIMS data and the wave activity density calculated from this error-

contaminated data is compared with those from the original data. Figure 5.1a gives wave activity density calculated by equation 5.46 from the original LIMS data and 5.1b gives the similar calculation but with the 5% random error imposed. We see that the two figures are generally quite similar in their main features with differences only in strength and some details. Figure 5.1c shows the relative error in wave activity density caused by the 5% random error in the data, i.e. difference of figure 5.1b and 5.1a as a percentage of of figure 5.1a values. We find that errors in wave activity density caused by the induced error are generally small, less than 10%, with very small patches reaching little over 20%. This again confirms that figure 5.1a and 5.1b are qualitatively very similar. The wave activity is, therefore, not unduly sensitive to the errors in the data and we can have some confidence that the calculations based on our scheme are able to reflect the real wave behaviour of the atmosphere.

5.4 Wave activity density

Wave activity density and E–P flux divergence have been calculated as described in the previous sections for the period 26 October to 16 November 1978. The general features of these quantities in that period are discussed in this section.

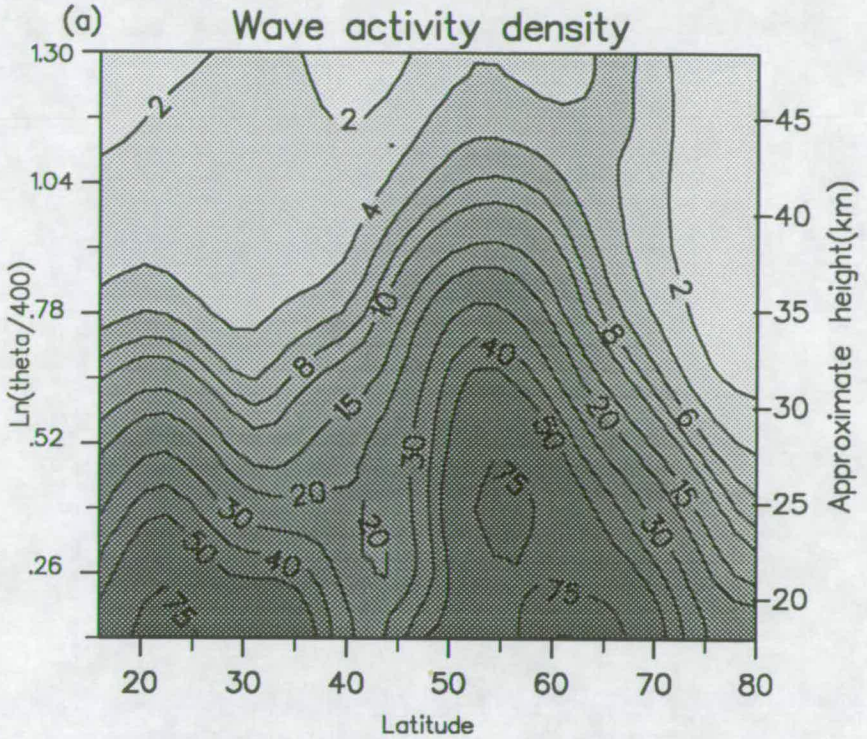


Figure 5.1a Wave activity density on 28 Oct. 1978. Calculated from LIMS data as described in text. Unit: 10000 mb*m*s/K. Note that the interval is not equal as indicated by each label.

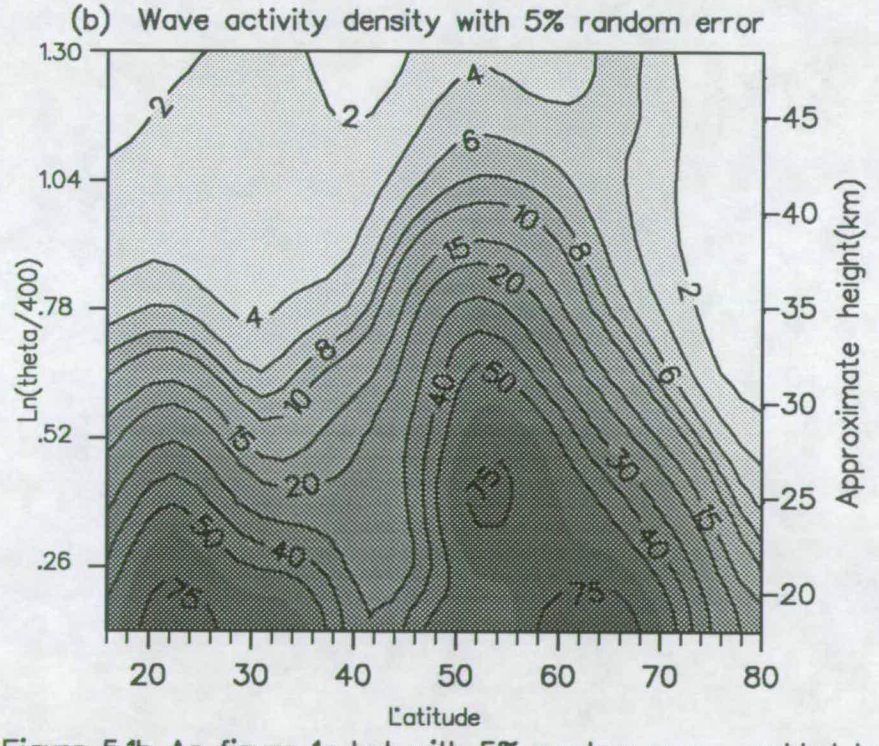


Figure 5.1b As figure 1a but with 5% random error added to the coefficients of waves in the original Fourior coefficients of LIMS data. Unit:as figure 1a.

(c)

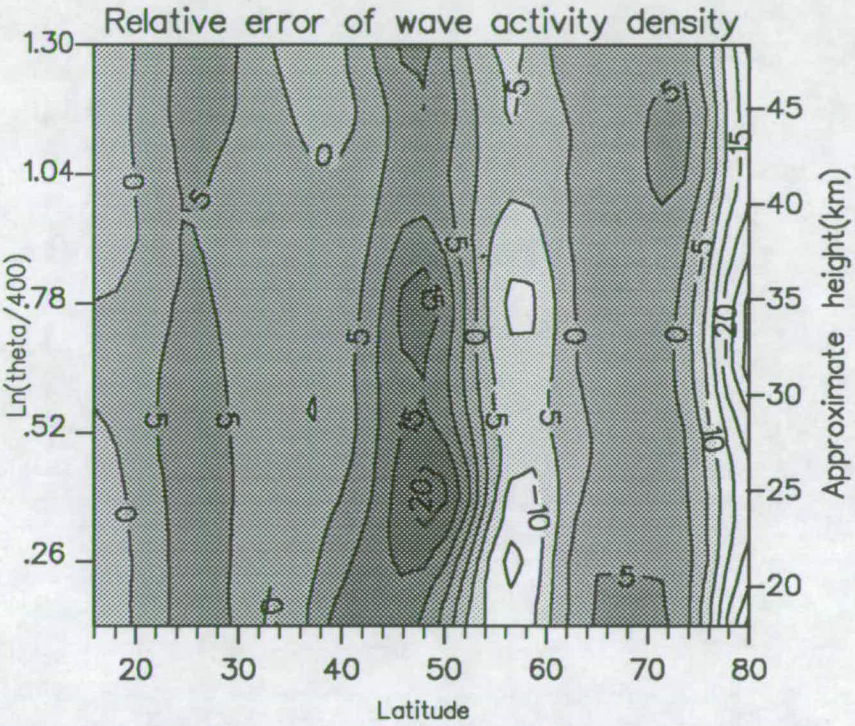


Figure 5.1c Relative error in wave activity density caused by adding 5% random error in LIMS wave component.

A	% of points with R < A	
	26 Oct.to 8 Nov. 78	3 to 4 Feb. 79
0.001	5.9	4.5
0.01	35.5	27.0
0.1	92.6	84.8
0.25	99.5	96.9
0.5	99.9	99.9
1.0	100.0	100.0

Table 5.1: Magnitude of terms in equation 5.46. R is the ratio of absolute value of the second term to the first term in equation 5.46

5.4.1 Relative sizes of terms in definition of wave activity

We note firstly that the second term in the right hand side of equation (5.46) is always considerably smaller than the first. This is true both for the period which forms the subject of this chapter and for a disturbed winter case which has also been investigated. Table 5.1 gives some quantitative information about the relative sizes of the terms.

We have computed the absolute value of the ratio, R in the table, of the second term of equation 5.46 to the first term and found the percentage of points at

which this is smaller than a specific value. It is less than 0.1 at more than 92% of the points in this autumn period, and at more than about 85% of the points in the disturbed winter case. It follow that the second term hardly affects the sign of the wave activity density in our case, so that the sign of A therefore is the same as that of $\overline{P}_y$. This removes one of the minor complications discussed by Andrews(1987) in connection with the interpretation of equation 5.43, namely that the wave activity density itself is of indeterminate sign due to the presence of the second term so that the sign of E-P flux divergence should also be indeterminate when infered from the generalized E-P flux theorem.

The figures presented below were computed with both terms included.

5.4.2 Time–latitude cross sections

The evolution of wave activity is shown in figure 5.2, which shows time-latitude distributions at three isentropic levels with the region of large wave activity shaded. The lowest level, $\eta = 0.39(\theta = 590°K$ around 30mb) is in figure 5.2a. At this level the wave activity seems to be confined to two latitude regions and to pulsate in time. One band of maxima is situated around 20° north, the other is situated in the range $50 - 70°N$. The high latitude maxima have major peaks on 28 October and 13 November with a minimum between them at 6–8 November.

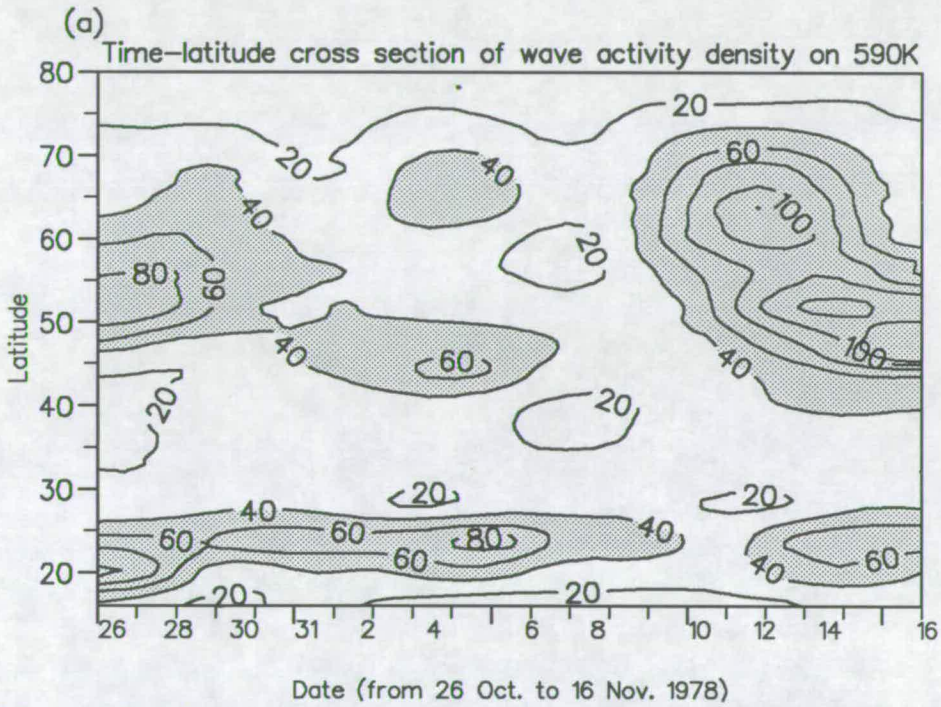
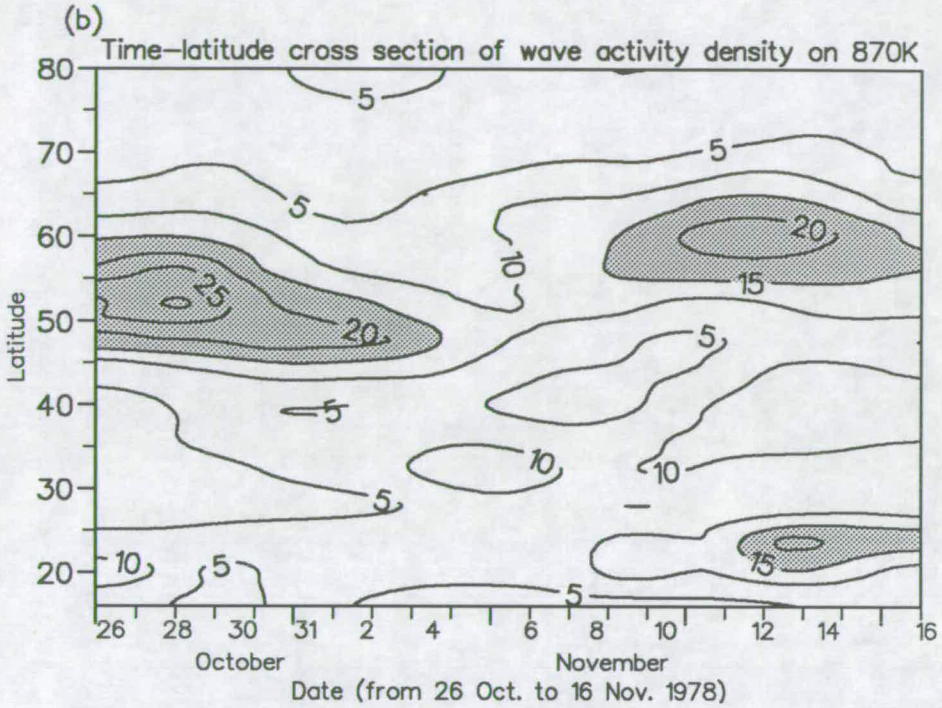


Figure 5.2a Time–Latitude cross section of wave activity density on 590K isentropic level (near 30mb). Unit: 10000 mb*m*S/K.

In the middle stratosphere (Figure 5.2b for $\eta = 0.78, \theta = 870°K$,near 10mb) the pattern is somewhat similar but the high latitude maxima show less temporal variation, the first episode persisting several days longer and reaching its maximum a day or two later than does the lower event. Note that the contour-interval is only a fourth as large in figure 5.2b as in 5.2a and that the peak values in the two figures differ by a factor of about that magnitude. There is some vertical coherence discernible between figure 5.2a and 5.2b in the large values of wave activity in the later part of the period of study.

Figure 5.2c is higher still at $\eta = 1.04$ corresponding to $\theta = 1130°K$ near 40km. The relationship with the previous chart is not totally clear-cut; one straightforward feature is the high latitude maximum around $50°N$ in the early part of the period, which lies directly over its counterpart at the lower level, but the low latitude maximum at greater height is found to lie more or less over a region of minimum in the previous figure, and it is not immediately apparent how the other major features in the middle and later parts of the period are related at the two heights. This is clarified by studying meridional cross sections to which we now turn.



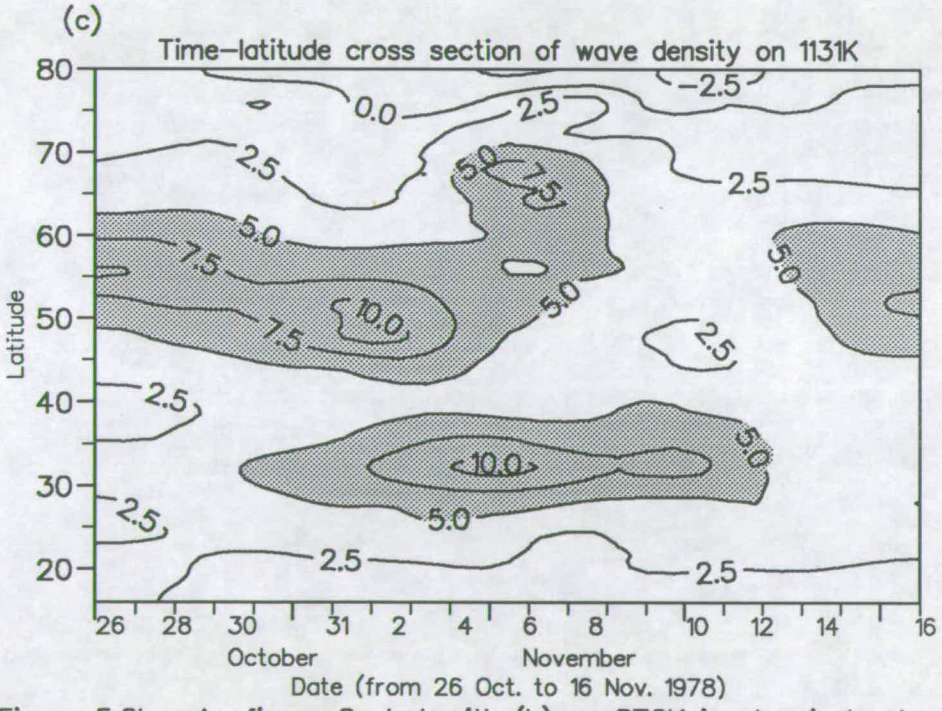


Figure 5.2b,c As figure 2a but with (b) on 870K isentropic level (near 10mb) and(c) on 1131K isentropic level. Figure 2a is taken on ln(theta/400)=.39; 2b on .78 and 2c on 1.04 level.; respectively.

5.4.3 Distribution in the meridional plane

Cross sections of wave activity on three days are shown in figure 5.1a and figure 5.3a,b; with figure 5.3c,d showing an alternative visualisation of figure 5.3a,b. The dates have been chosen to correspond to different stages of the evolution of the values of high latitude activity in the lower stratosphere, figure 5.1a,and 5.3b corresponding to high values and figure 5.3a corresponding to low values.

From figure 5.1a,5.3a and 5.3b we firstly note that wave activity density decreases very rapidly with increasing height, from more than 75 units at 20km to only 2 units above 40km, this is reflected in figure 5.3c,d very obviously. This decrease is caused mainly by the scaling effects of air density in the definition of wave activity density in equation (5.46). It can be verified by scaling wave activity density with $\overline{\sigma} a \cos(\phi)$, then such a feature will disappear. It is worthwhile to point out that after such a scaling the wave activity density will have a sensible unit of the winds: m/s, since by transformed Eulerian-mean formalism the divergence of E-P flux, which has the same unit as dA/dt, appears as the only ~~term of~~ force exerted by eddies on the zonal mean-flow.

Another feature, in agreement with figure 5.2, is that the maxima of wave activity is concentrated in two regions, situated at $20-30°N$ and $50-70°N$ respectively. These regions are the approximate positions of tropospheric and low stratospheric

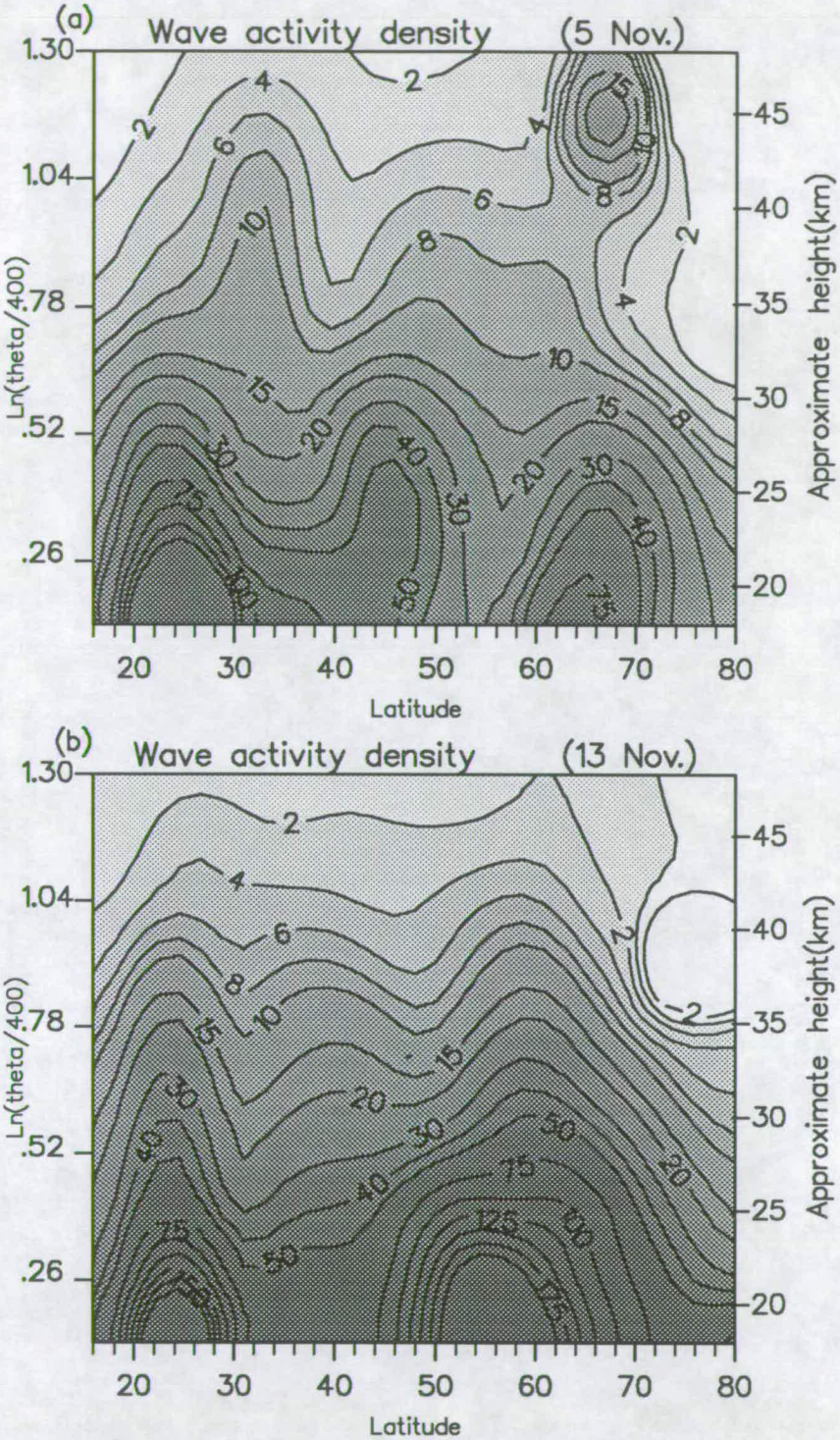
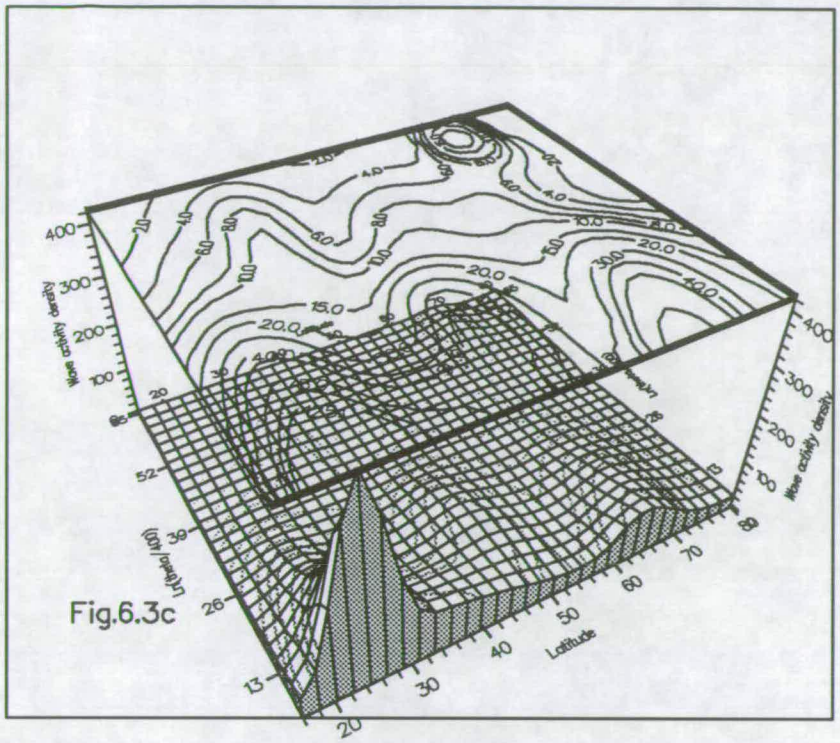


Figure 5.3a,b Typical distribution of wave activity density on a specific day. (a) 5 Nov. (b) 13 Nov. 1978. Unit: 10000 mb*m*s/K as fig.1a. Calculated from LIMS data.



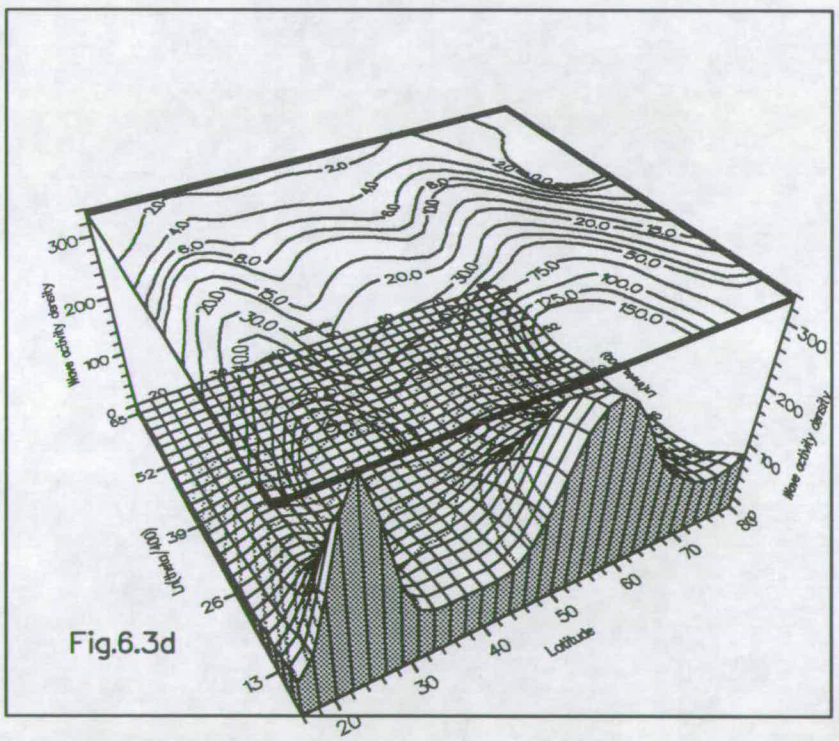


Figure 5.3c,d Equivalent to figure 6.3c,d but for visualised picture.
Note:Ln(theta/400) values is 50 times of original values
for plotting purpose.

jets, where the conventional measure of waves such as wave amplitude or wave energy density are also at their largest. The present measure of wave activity, therefore has also retained some of the features of other common used measurements of waves.

From 26 October to 16 November, ~~though~~ there is low wave activity in the period around 5 November, the wave activity at the low levels is generally strengthened ~~along~~ with the progress of winter. For instance, in figure 5.2a, the maximum at 27 October is 80 units, but it has increased to 100 units from 11 November on. The daily figures of wave activity show that the magnitude of wave activity keeps growing after 16 November so that the small amplitude condition is violated by 18 November (figure 5.3e) and the theorem in the present form is no longer applicable. The further attempt to investigate the later disturbed period, although not totally satisfactory due to the previously mentioned problem of singularity, also confirmed this feature. This agrees with the fact that the waves are generally growing towards winter.

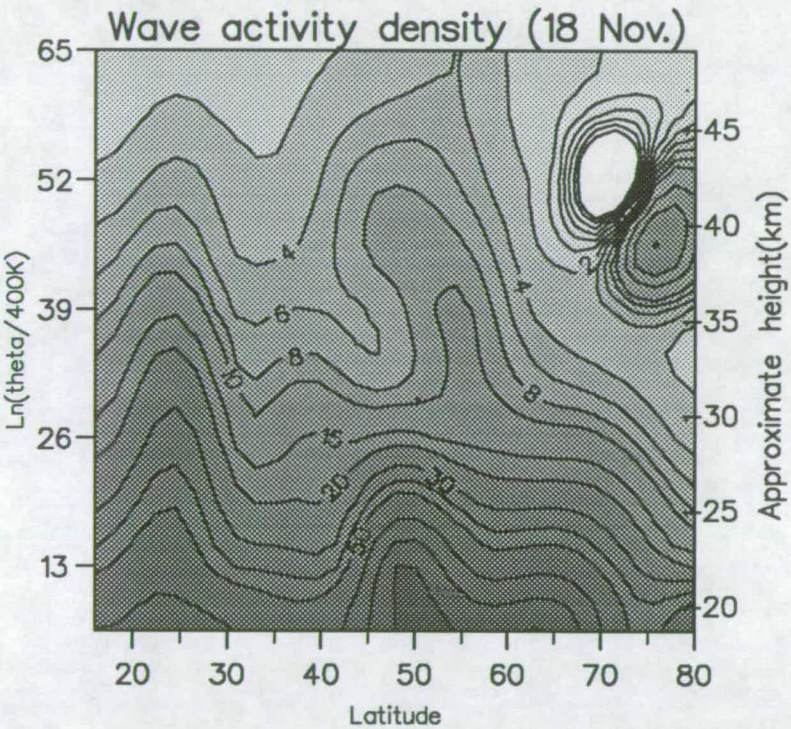


Figure 5.3e As figure 5.3a,b but for 18 November 1978.

5.5 Coherence between the wave forcing, the wave activity density and the zonal mean winds

5.5.1 Relation between wave drive and wave activity density

One advantage which the transformed Eulerian mean formalism has is that the divergence of E-P flux is the principal term of wave driving of the zonal mean flow. The GEP theorem of equation (5.43), however, relates the wave-forcing to wave transience and wave dissipation and thus allows us to investigate how directly these wave quantities contribute to the forcing of the zonal mean-flow by planetary waves.

Neglecting the non-conservative term, the GEP theorem states that if $A > 0$ and the waves are growing ($\partial A/\partial t > 0$) then $\nabla \cdot F < 0$, while for decaying waves ($\partial A/\partial t < 0$) ,$\nabla \cdot F > 0$. On the time evolution maps of A and $\nabla \cdot F$ this relation demands that the large centre of wave activity density should be located where the wave forcing changes sign from negative to positive.

Figure 5.4a,b,c show the time evolution of divergence of E-P flux on the same

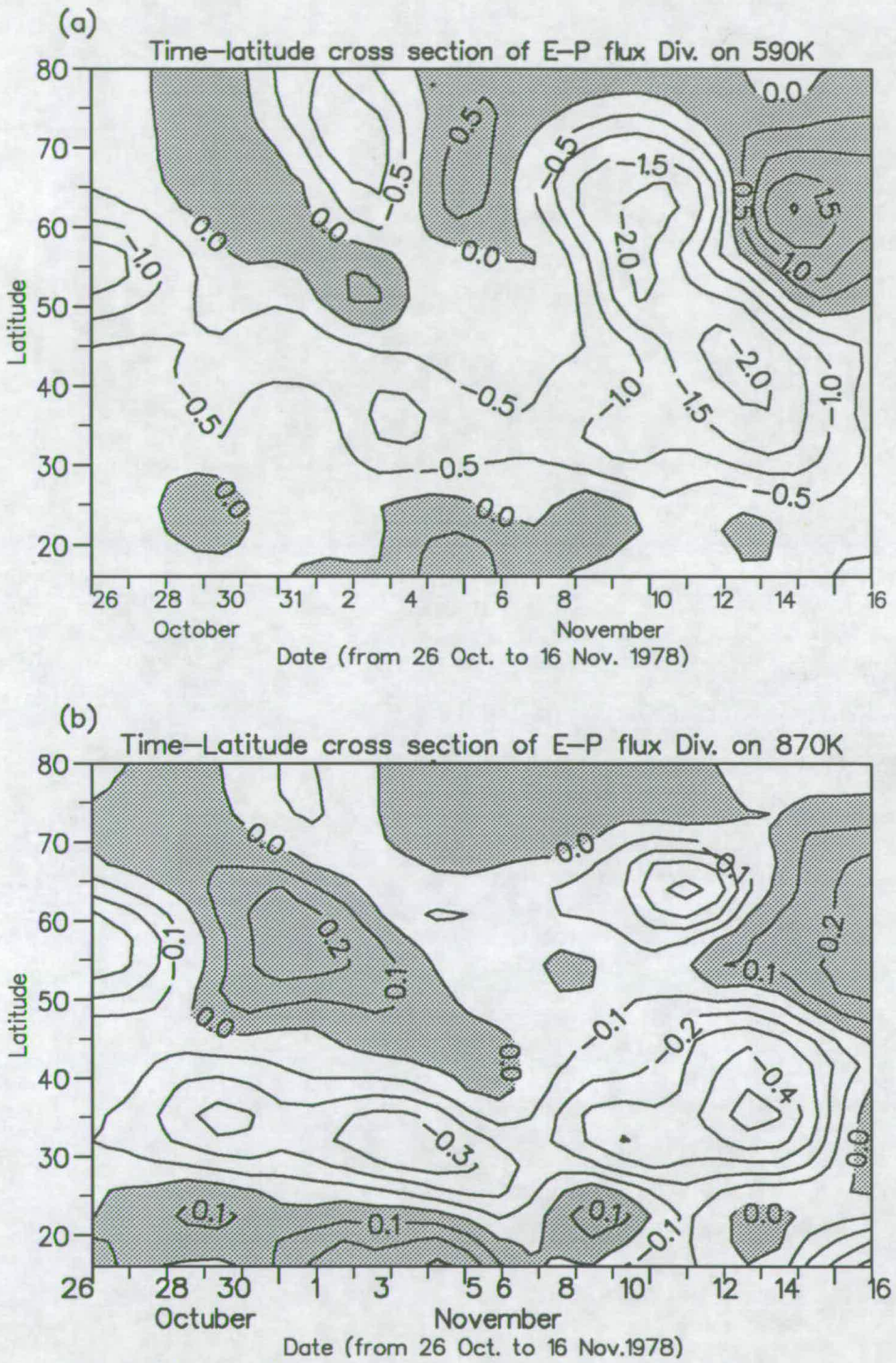


Figure 5.4a,b Time–latitude cross section of E–P flux divergence on isentropic surface. With (a) on 590K level (near 30 mb), and (b) on 870K level (near 10mb). Unit: mb*m/K. The positive area is shaded.

three levels. Comparing figure 5.4a with figure 5.2a we find that in the high latitude region of maximum wave activity the centre of strong wave activity just overlaps the strong gradient of wave drive, i.e. it lies where the divergence of E-P flux changes sign from negative to positive. This is exactly the requirement of the theorem as we already discussed. On the other hand, this also indicates that the theorem reflects the relation of wave transience and wave drive in the real atmosphere when the small amplitude assumption is valid.

The relationship shown in the last paragraph can be seen even more clearly in Figure 5.5 which clearly shows the negative correlation between wave transience ($\partial A/\partial t$,say) and E-P flux divergence. The figure shows by the values of $\partial A/\partial t$ and E-P flux divergence along the $68°N$ degree latitude on 590K level. In the lower latitude regions of maximum wave activity, the relation between wave drive and wave activity density is not as obvious as its northern counterpart. We see that the E–P flux divergence remains negative in the region $30 - 40°N$. According to transformed Eulerian mean formalism, this should lead to continuous decelerating of the zonal mean winds. As will be seen later, this is not well reflected by the fields of zonal mean winds. The cause of this lower latitude incoherence is not very clear at present. A possible reason might be that the lower latitude maxima are located in a different region, where different mechanisms dominate in the atmosphere. This region is approximately in the region of 'surf–zone', the

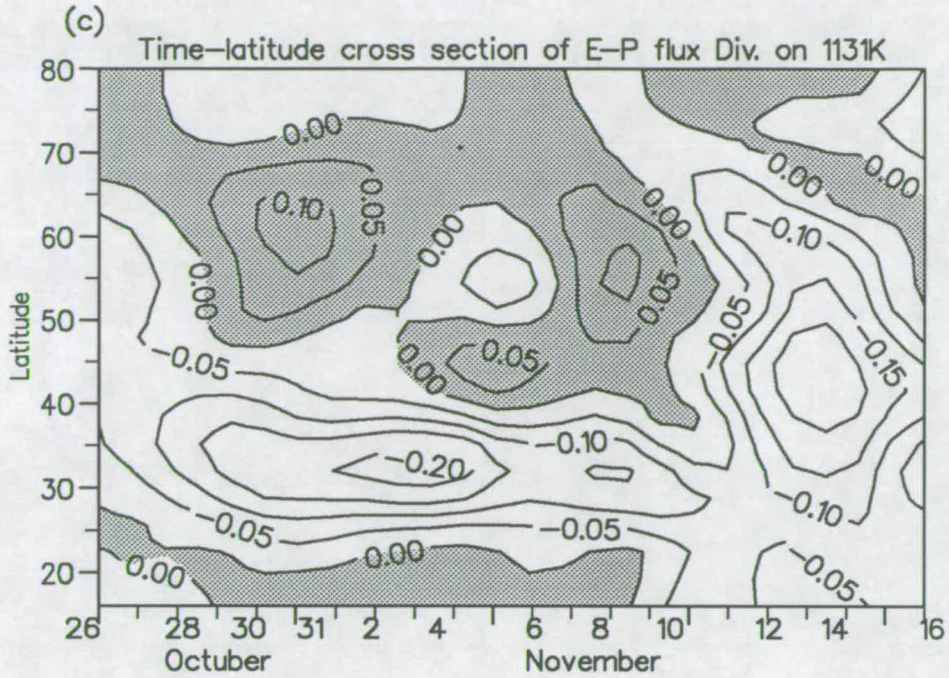


Figure 5.4c As figure 4a,b but on 1131K isentropic level . The positive area is shaded.

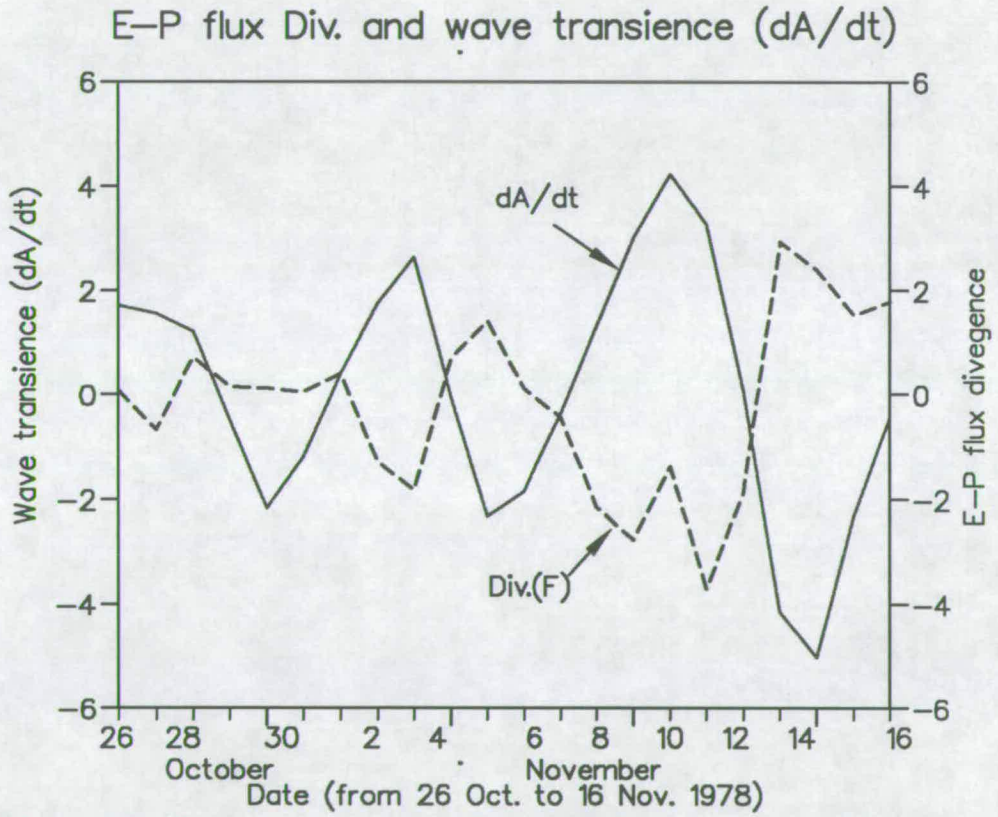


Figure 5.5 Wave transience and E–P flux divergence along 68 N latitude on 590K (near 30mb) level. Unit: mb*m/K. Solide line for wave transience dA/dt and dash line for E–P flux divergence.

main place of wave breaking and irreversible mixing to unresolvable scales. On account of mixing effects (which are necessary of finite amplitude) and the limitation of data for diagnosing small scale features, the wave forcing of zonal mean flow may not be as obvious as in northern regions or may no longer be resolved in the figure.

5.5.2 The effects on zonal mean flow

As indicated by Andrews (1987, 1989), the separation of the real flow into 'zonal mean' and 'eddy' or 'wave' can be rather artificial. Thus, a joint check of the variation of wave drive and zonal mean flow may be valuable.

Figure 5.6a shows the time evolution of zonal mean wind on $590°K$ level. Comparing with figure 5.2a we find that in mid-high latitudes when the wave is growing, the zonal mean is decreasing and vice versa, so that the maxima of wave activity are situated in the region where the zonal mean flow is at its minima. The two are in very good agreement, as is further confirmed by figure 5.6b showing two curves of wave activity density and zonal mean zonal wind along $68°N$ latitude circle on that level. Though we shall not expect complete quantitative correspondence between wave forcing and zonal mean flow as there are other terms such as Coriolis force etc. in the transformed Eulerian zonal momentum equation, the general

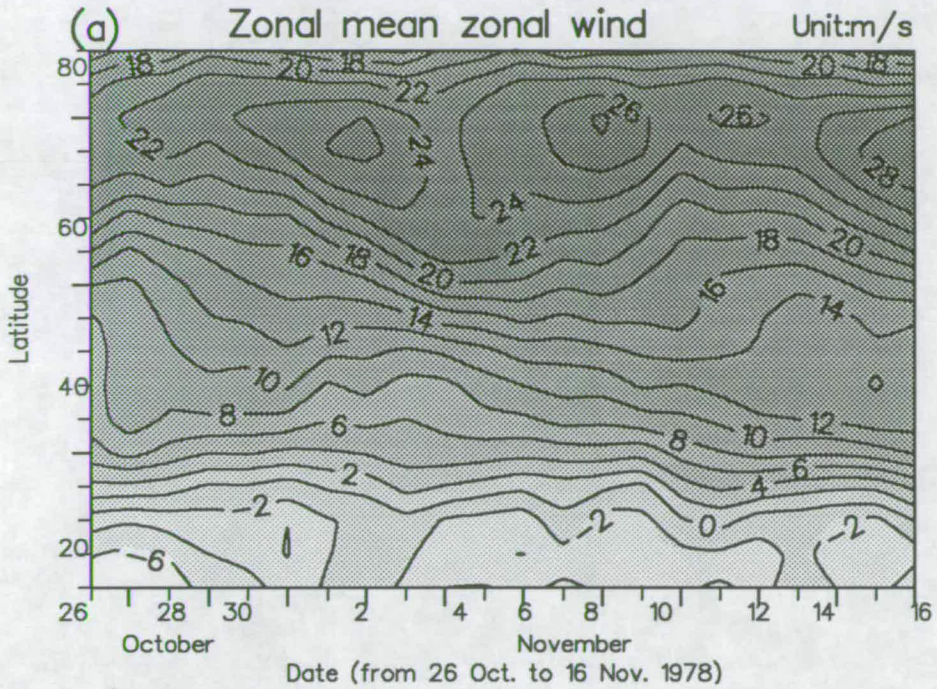


Figure 5.6a Time–latitude cross section of zonal mean zonal wind on 590K isentropic surface (near 30mb). Unit: m/s. Winds are derived from LIMS data as described in text.

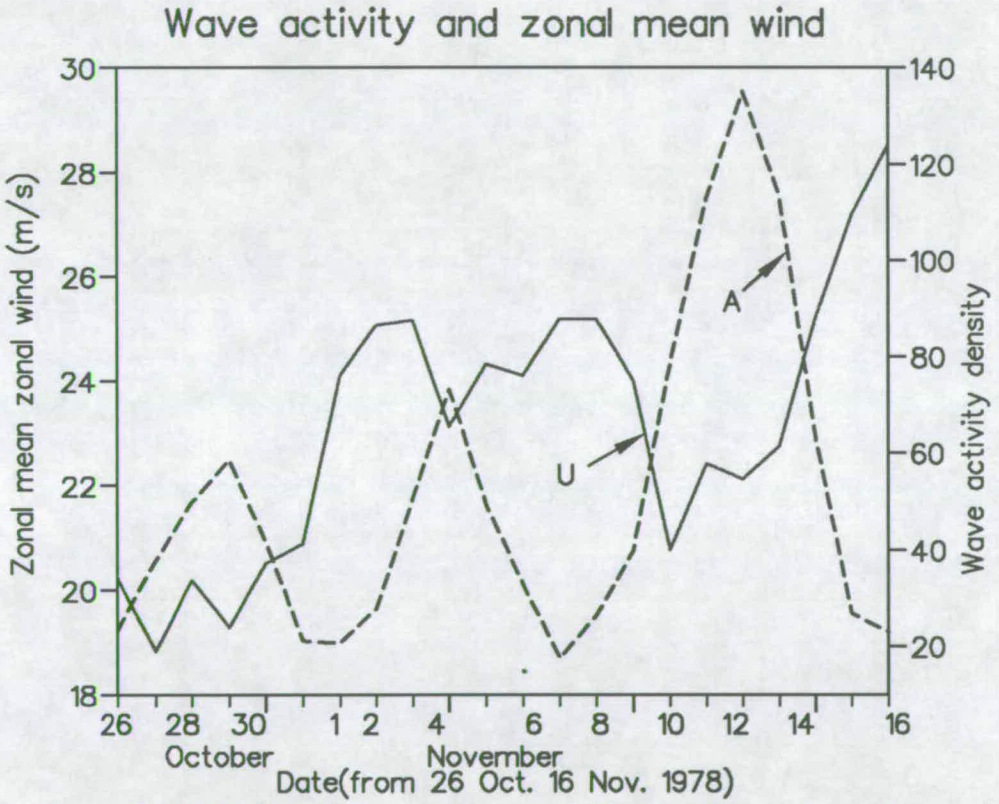


Figure 5.6b Relation between wave activity density and zonal mean zonal wind. The curves are along 68 degree latitude on 590K isentropic level. Unit for zonal wind is m/s; for wave activity density mb*m*s/K. Solid line:zonal mean zonal wind; dash line:wave activity density.

coherence between the two does reflect the interaction of wave and zonal flow and the applicability to this situation of the theorem which describes the interaction.

5.6 Correlation with ozone

Ozone, like PV is a quasi-conservative tracer in the middle and lower stratosphere, and has been successfully used to illustrate the processes of planetary wave breaking events(Leovy et al 1985) in the middle atmosphere. It will be shown that an equation analogous to (5.46) can be found in which PV is replaced by ozone mixing ratio as an approximately alternative measure of wave activity density. The reasoning and theoretical derivation of this substitution will be shown in the next chapter. This kind of density is then calculated and the results discussed below.

Figure 5.7 shows the time–latitude cross section of wave activity density calculated with ozone mixing ratio in the place of PV in equation 5.46 with $\overline{\sigma' u'}$ ignored, on the $590°K$ and $870°K$ isentropic surfaces (Figure 5.7a and b respectively). Since the gradient of ozone is in the opposite direction to the gradient of potential vorticity (Grose and Russel 1985, Leovy et al. 1985), the wave activity density defined by ozone should also have opposite sign to that derived from potential vorticity.

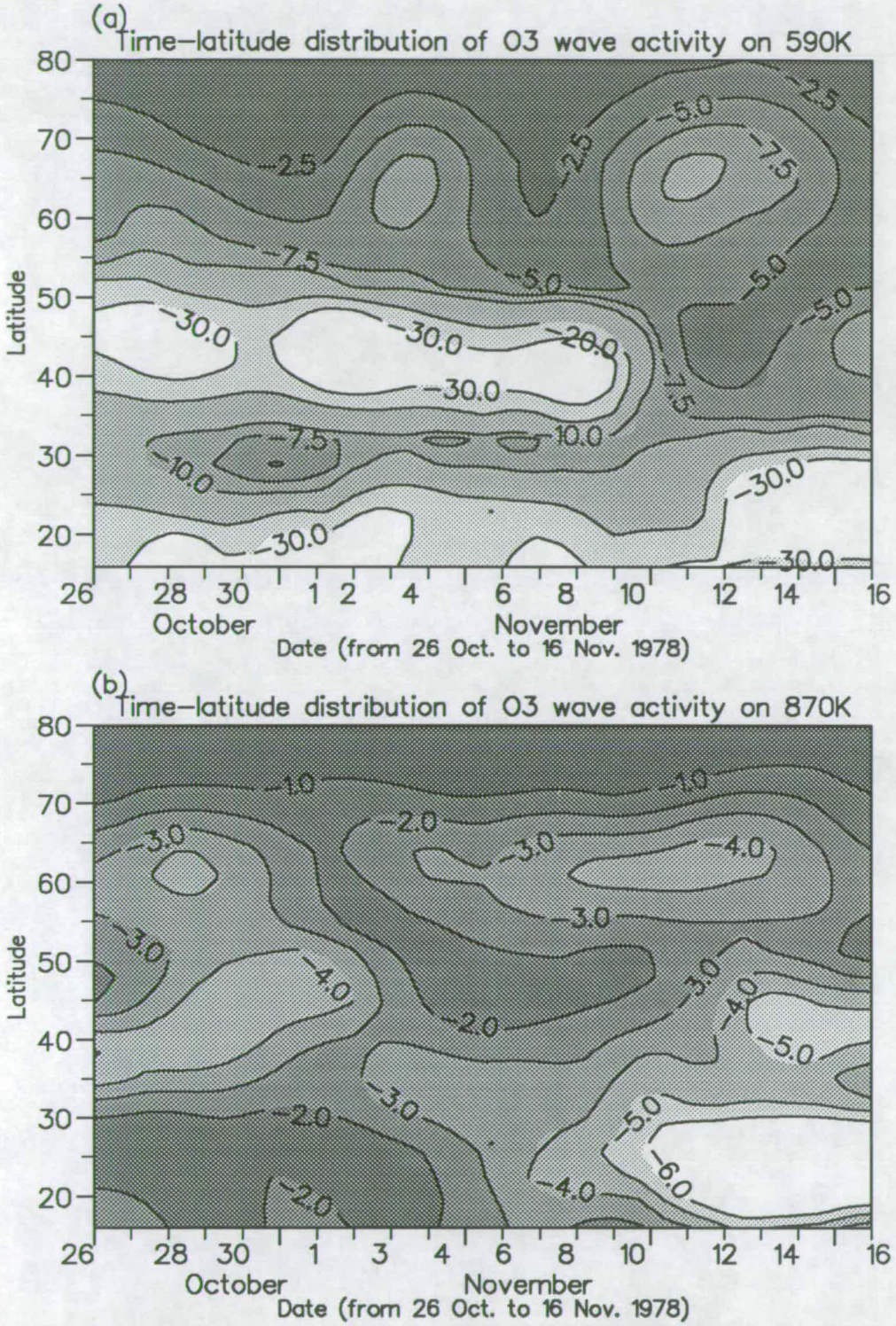


Figure 5.7a,b Time–latitude cross section of wave activity density defined by ozone mixing ratio in place of potential vorticity (see text for detail).
(a) on 590K isentropic level (near 30mb);unit:1.0E7mb*m*s*s*ppmv/K.
(b) on 870K isentropic level (near 10mb); unit: 1.0E6 mb*m*s*s*ppmv/K.

Comparing figure 5.7a,b with their counterparts figure 5.2a,b, we find that there are quite obvious correlations between the two definitions of wave activity density. For example, on the $870°K$ level the three shaded maxima of wave activity density in figure 5.2b approximately overlap the area of minima of figure 5.7b. These features can also be found in the northern region of $590°K$ level. Overall a negative correlation does exist between these two kinds of wave activity defined by different tracers. This correlation gives confidence that real features (as opposed to data-noise) are being diagnosed, and it emphasizes the dependence of tracer transport on the dynamics of wave activity. Similar arguments were used by Leovy et al (1984) and Butchart and Remsberg (1986) in the context of wave breaking and area change respectively observed in tracer distributions.

5.7 Balance analysis of GEP theorem

Equation 5.43 shows that if the wave amplitude is small enough the last term in the right hand side, $O(\alpha^3)$, can be neglected, and the three terms should be balanced by each other. For finite amplitude waves, the $O(\alpha^3)$ term will become more important as discussed by Marks (1989). The extension of the theorem to finite amplitude waves which is then needed has been developed by Killworth and McIntyre (1985), McIntyre and Shepherd (1987) and Haynes

(1988). This is beyond the scope of the present work. In our present case, as indicated in the previous section, the wave amplitude is small. Hence we shall expect the approximate balance among the three terms in 5.43.

We have seen a qualitative balance in the previous section. In this section we investigate to what extent a quantitative balance exists. The balance can be studied in terms of several quantities, such as the residuals directly from the summing up of the three terms of the theorem, the relative error of balance by dividing the absolute values of the residuals by the absolute value of the largest term of the three. To know how approximately important the nonconservative term contributes to the balance, we can drop it from the above two quantities and compare the results.

Figure 5.8a and 5.8d shows the time–latitude cross section of balance on $590°K$ level and $870°K$ level respectively. The balance is calculated directly from the theorem, i.e. the residual of the three terms $\partial A/\partial t + \nabla \cdot F - D$. Figure 5.8a shows that except the small patches (heavy shaded) locating at the area of maximum wave activity density, the balance is qualitatively reached with most residuals less than the order of each term itself. The similar phenomena can be found in $870°K$ level with the most residuals less than 0.2 (figure 5.8d), c.f. figure 5.4b where the individual term reaches 0.5.

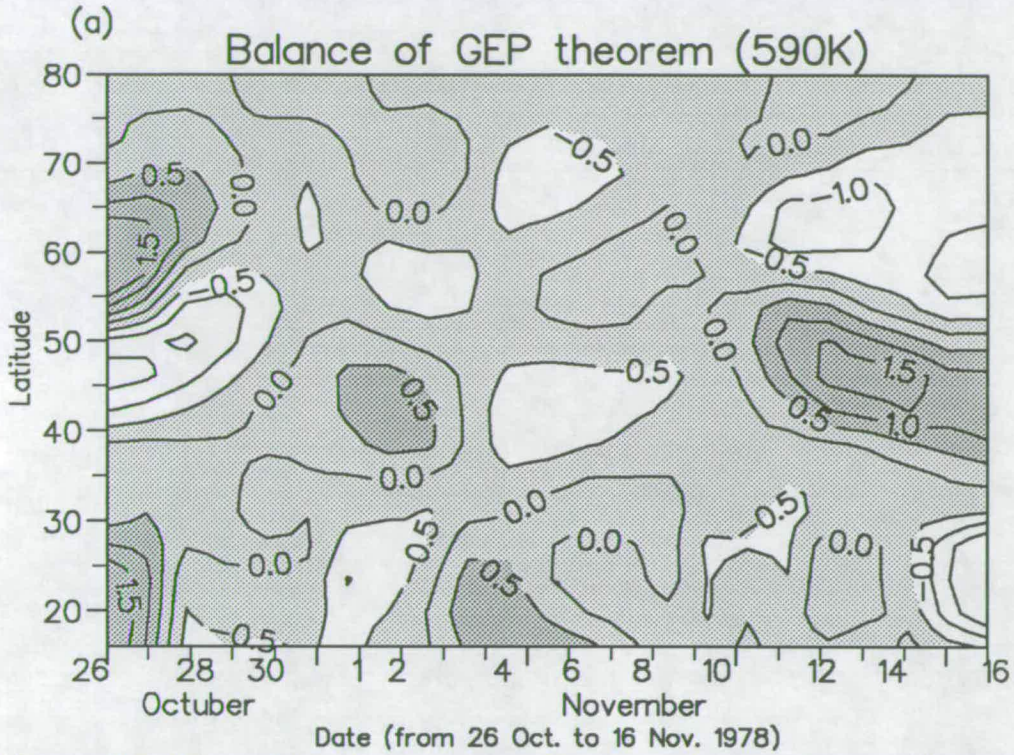


Figure 5.8a Time–latitude cross section of balance of generalied E–P flux theorem on 590 isentropic surface. The balance is calculated directly from the theorem,i.e. dA/dt+div(F)–D. Unit:mb*m/K.

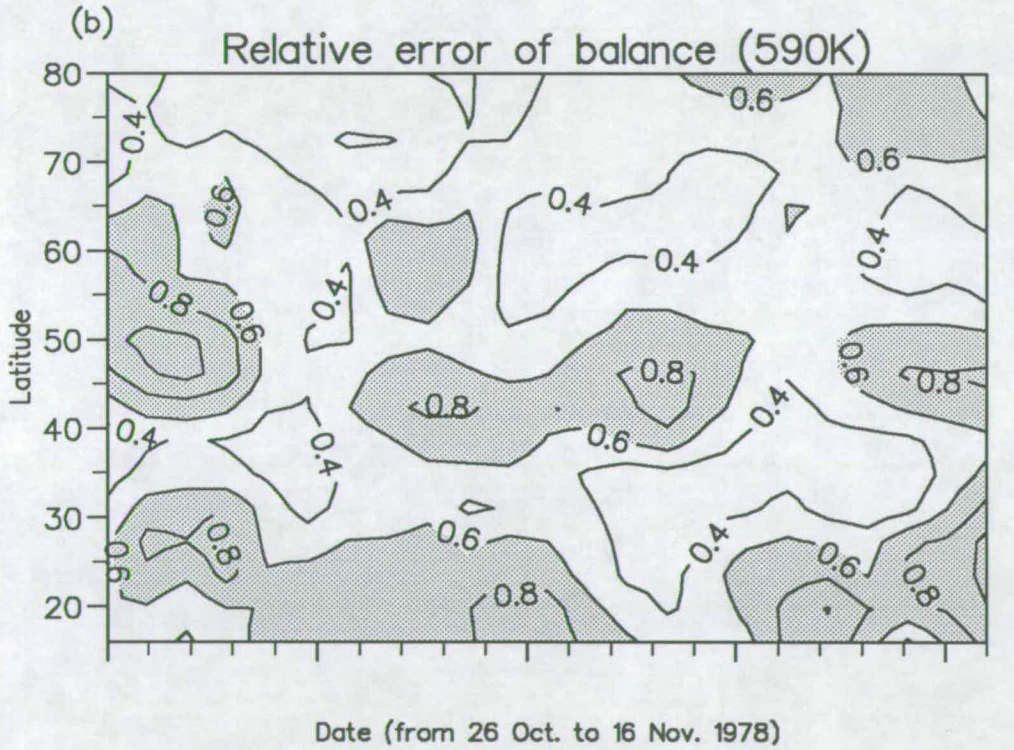


Figure 5.8b Relative error of the balance represented by fig.8a. The error is calculated by:|dA/dt+div(F)–D|/max{|dA/dt|,|div(F)| ,|D|} The areas greater than 0.6 are shaded and the numbers are dimensionless.

To see more concisely and know the role the dissipation term plays in the balance, figure 5.8b gives relative error of the balance, i.e. calculated from $|dA/dt + \nabla \cdot F - D|/max(|dA/dt|, |\nabla \cdot F|, |D|)$. Therefore if the error is small, the terms are balance each other. Otherwise the balance is not reached. Figure 5.8b shows that the balance is quite well reached in some areas with the error less than 0.4, but not reached in other areas with errors over 0.6 (shaded). Figure 5.8c is similar defined to figure 5.8b but with the non-conservative term dropped from the residuals of balance to see the role of the non-conservative plays in the balance. We see that the area with residuals larger than 0.6 (shaded) are more widespread than in figure 5.8b. A similar feature can be found in figures 5.8e,f, the equivalent of figure 5.8b,c but for 870°K isentropic level. Thus the dissipation term is important in balancing the wave transience and wave drive in the lower and middle stratosphere even in small waves. This shows that the three terms have the same importance in the theorem in accord with the result of Marks (1988).

At the high, 1131°K level, however, the significance of the dissipation term in the balance seems not so obvious or even adverse though it has the same magnitude. This may be due to the increasing effects of breaking gravity waves in the high stratosphere to lower mesosphere or due to small scale waves and nonlinear effects which are not explicitly expressed by the theorem. This is in agreement with the result of Marks(1988).

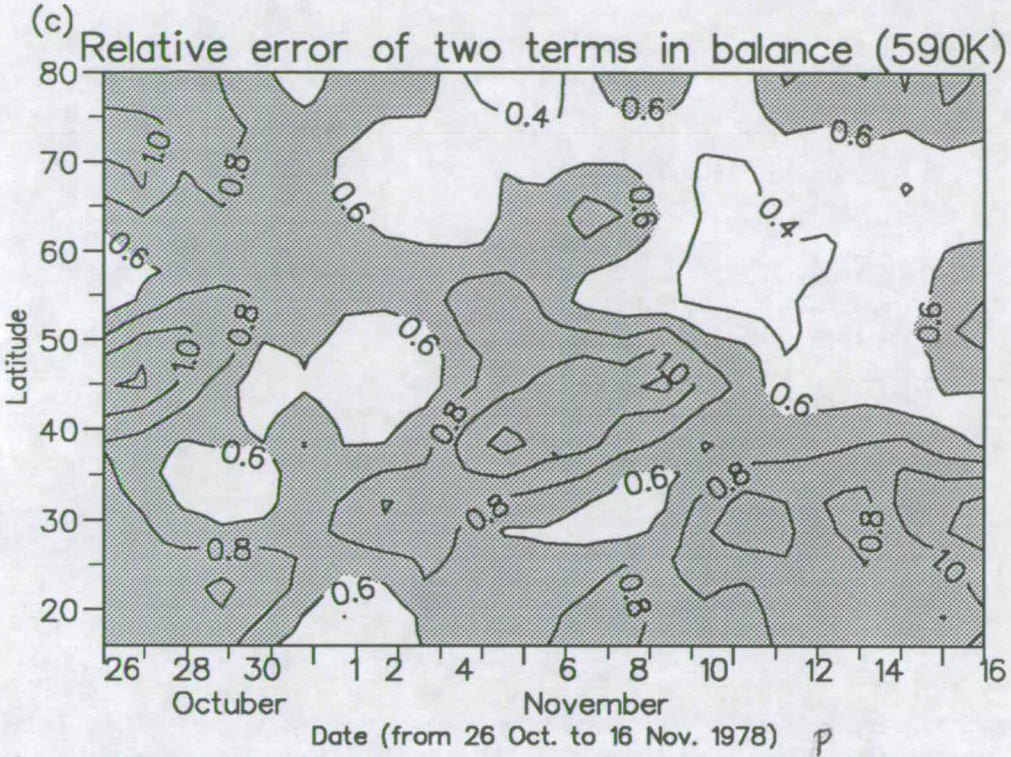


Figure 5.8c As 8b with the nonconservative term droped. The areas with values greater than 0.6 are shaded. Unit: dimensionless.

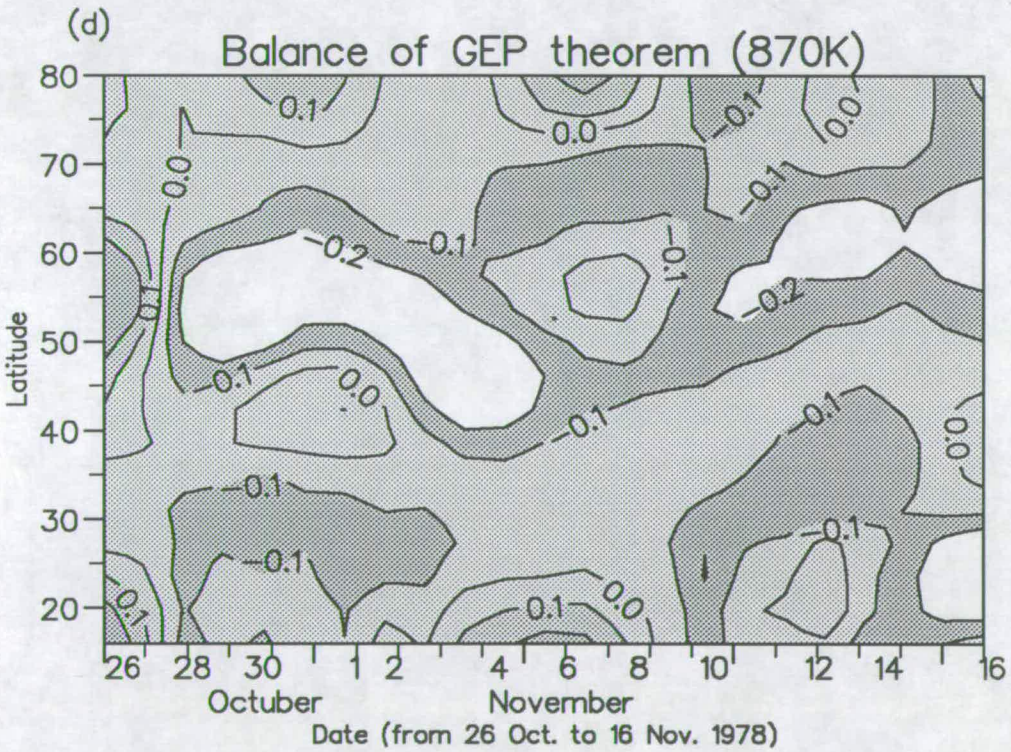


Figure 5.8d Time–latitude cross section of balance of generalized E–P flux theorem on 870K isentropic surface. The balance is calculated directly from the theorem, i.e. dA/dt+div(F)–D. Unit:mb*m/K.

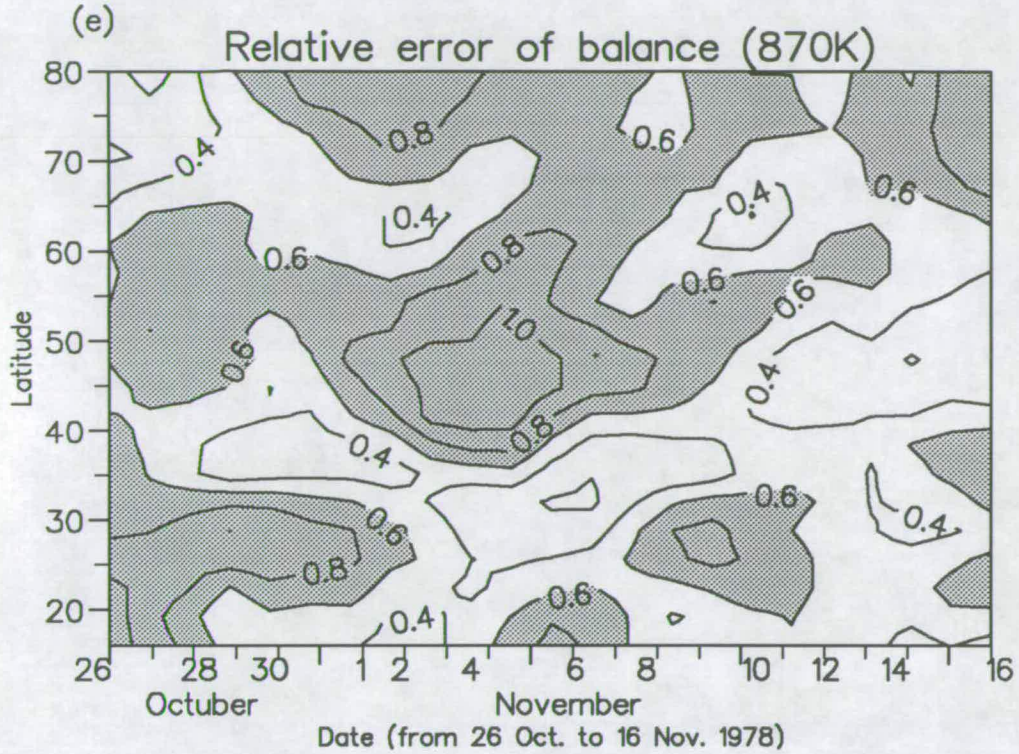


Figure 5.8e Relative error of the balance represented by figure8d on 870K isentropic level. Calculated by:$|dA/dt+div(F)-D|/max\{|dA/dt|, |div(F)|, |D|\}$ Areas greater than 0.6 are shaded and numbers are dimensionless.

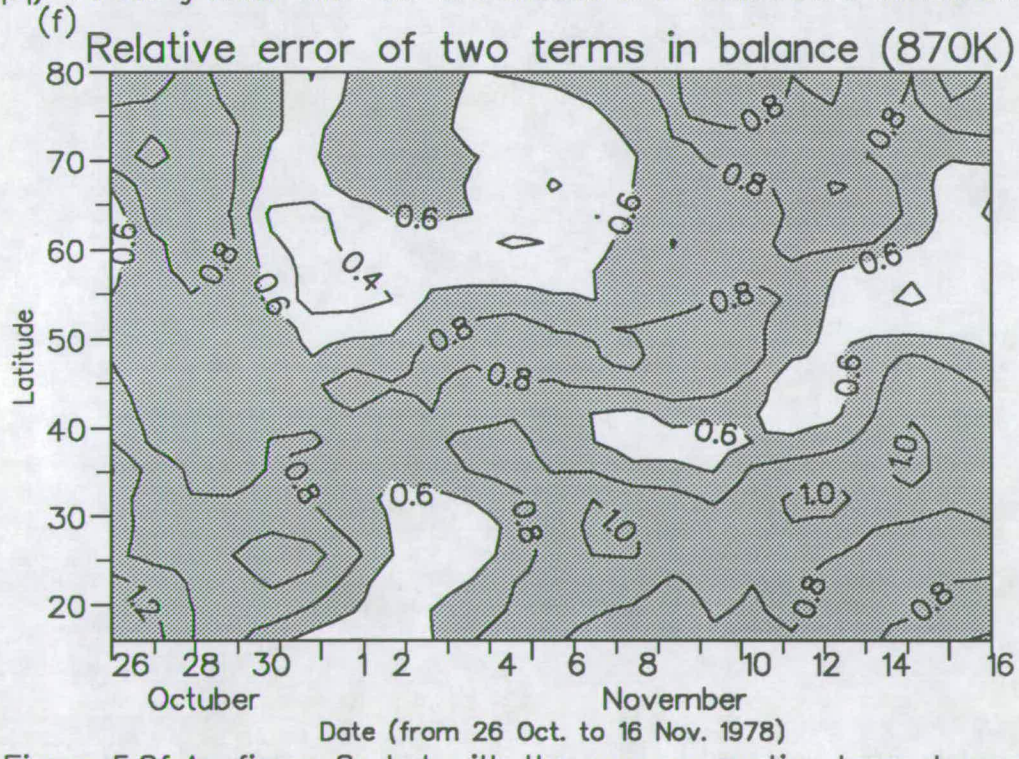


Figure 5.8f As figure 8e but with the nonconservative term dropped from the balance. Areas greater than 0.6 are shaded and numbers are dimensionless.

It is necessary to point out that the present work is carried out with daily data rather than monthly mean data used by Marks. This should be more suitable for the calculation of wave transience and thus the balance.

Generally the balance is qualitatively reached but quantitatively poor. This may reflect either the effects of small scales and nonlinear waves as mentioned above or the error in the calculation due to the limitations of the data.

In order to overcome the singularity problem and extend the present work to large wave amplitude periods, an attempt has been made to revise the gradients of Ertel potential vorticity by exploiting an idea of Kinnersley (personal communication). The basis of the revision is the idea that in the absence of eddies, the polar vortex of the basic state corresponds only to the seasonal variation and will be symmetrical the polar cap with no shift from the pole. Therefore it may be possible to replace the real gradients of PV ($\bar{P}_y$) in equation 5.46 by the gradients of a hypothetical basic state PV ($\tilde{P}_y$), where $\tilde{P}_y$ is obtained by a non-divergent readjustment of the potential vorticity so that it is symmetric and increasing with latitude. Unfortunately the region of maximum of wave activity seems to be located too far south. As the meaning of this revision and the resultant wave activity to dynamics is also not clear at present, the results are not discussed here.

5.8 Conclusion and discussion

In this chapter the terms of the GEP theorem in isentropic coordinates are calculated diagnostically using the LIMS daily data. As far as the author knows, this is the first attempt to apply the isentropic version of the GEP theorem to daily diagnostics study. In the process of carrying out the present work, a series of daily maps of latitude–height cross section of wave activity density, such as figures 5.1a, 5.3a,b, are obtained and daily evolution can be observed therefrom. Based on these maps and the other related analysis, the main conclusions can be drawn as:

1) The generalized E-P flux theorem, with definition of wave activity density expressed in Eulerian quantities given by Andrews (1987) based on the condition of small wave amplitude, is indeed applicable to the diagnostic study of stratospheric dynamics using satellite observed data when the wave amplitudes are small. The wave activity density and the wave transience have both a coherent time evolution and spatial distribution. The variations of wave activity are also reflected in the changes of zonal mean flow and give some indication of dynamics of wave-mean flow interaction. As the wave amplitude becomes larger, such as in the event of stratospheric warming, the wave activity density fails to show reasonable behaviour due to the change of sign of the gradient of zonal mean Ertel potential vorticity which leads to the occurrence of a singularity in this definition of wave activity density. Therefore to suit the complexity of dynamics in the real atmosphere, a further

calculation with finite amplitude version of the theorem may be necessary.

2) In the definition of the wave activity density the first term is about one order larger than the second and thus hardly able to affect the sign of the density in the present case. Of the three terms in the definition of dissipation, i.e. equation 5.47, the first term dominated over the third concerned in present work.

3) The balance of the three terms in GEP theorem is reached approximately qualitatively in lower stratosphere, but not quantitatively so. This may reflect the effects of small scale and nonlinear waves represented by $O(\alpha^3)$ terms in the theorem and not explicitly calculated here. The wave transience, wave drive and the nonconservative term have similar importance. In the waves of larger amplitude the balance is not applicable due to the failure of the calculation of wave activity density.

4) The substitution of ozone mixing ratio into the definition of wave activity density acts as a qualitative check for our calculation and also shows the relation between the dynamical tracer and chemical tracer.

It also possible to apply the calculation to the southern hemisphere which may help to investigate the difference of wave activity between two hemispheres. Since the zonal mean flow in the winter stratosphere of the southern hemisphere is stronger than its northern counterpart and the amplitude of the planetary waves

is generally smaller due to the lack of forcing by the orography from below, the version of the GEP theorem used in this chapter for small amplitude waves may be more suitable.

Chapter 6

Tracer Wave Activity and Transport Effects

6.1 Introduction

As set out in chapter 1 and chapter 5, the generalized E-P flux theorem (GEP theorem hereafter) sets quantitative links among wave transience, wave driving of the zonal mean state and nonconservative processes. The expression of each term in Eulerian quantities allows us to calculate it diagnostically from observational data and provides a useful means of understanding the dynamical mechanism involved in the wave, zonal-mean flow interaction. On the other hand, because of the fundamental link between wave activity density and the parcel displacement (Edmon et al, 1980; Andrews 1987), namely that the wave activity density is proportional to the dispersion of parcels about their mean positions, i.e. $A \propto$

$\bar{P}_y\overline{\eta'^2}$, the GEP theorem also has important implications for tracer transport.

It is well known that there are similarities between dynamical tracers and chemical ones, especially between ozone and Ertel potential vorticity. As noted in chapter 2 and 4, Ertel potential vorticity is a conserved tracer under conditions of adiabatic, frictionless flow. In the lower to middle stratosphere, both ozone and potential vorticity are quasi-conserved for time scales of a week or so. Therefore potential vorticity has been used as a proxy for studying the transport of quasi-conserved trace species in this region of atmosphere. For instance, Hartmann (1976b) concluded from his study that the eddy transport of potential vorticity was consistent with observations of ozone transport and calculations conducted with general circulation models. Leovy et al. (1985) studied the transport of ozone using the 10mb daily map of ozone mixing ratio for the period October 1978 to May 1979. Their results show qualitative features which strongly support McIntyre and Palmer's (1983, 1984) hypothesis of planetary wave breaking based on an investigation of potential vorticity maps on the 850K isentropic surface. On this surface, both ozone and Ertel potential vorticity distributions display the signature of breaking waves with a tongue of high PV and low ozone being drawn clockwise out of the main polar vortex region and, meanwhile, a tongue of lower PV and high ozone being drawn from low latitudes poleward. Leovy et al. (1985) also suggested that observed differences in the ozone and potential vorticity arose from different

nonconservative processes affecting the two tracers and from their different gradients. Hence except for their differences in sources and sinks—including the special consequence of those differences noted by Haynes and McIntyre (1987) that there can be no net transport of PV across isentropic surfaces while ozone is not so restricted—the overall similarities between PV and chemical tracers such as ozone, water vapour etc. in dynamical behaviour, transport processes and life time well recognized from recent researches.

The generalized E-P flux theorem is based on Ertel potential vorticity and the mass continuity equation. Checking the procedures for deriving the generalized E-P flux theorem with each term expressed by Eulerian quantities (see e.g. Andrews 1987 and K.K.Tung 1988), we find that for every equation concerning PV and required for the derivation, there can be found a counterpart equation with PV replaced by tracer mixing ratio. Bearing in mind the well-recognized similar characteristics of PV and quasi-conservative chemical tracers in dynamical and transport processes, it is possible, as will be shown later, that an expression analogous to the GEP theorem for chemical tracers can be derived. Considering the relation between wave activity density, PV and parcel displacement, such an analogous expression and corresponding chemical tracer wave activity density may have implications for the parameterization of 3-dimensional transport in 2-dimensional models.

The main aim of this chapter is to present a derivation of an analogous expres-

sion to the GEP theorem for chemical tracers. The next section will give details of the derivation. Section 6.3 will discuss the simplification of such an analogous expression and analyse its implications and possible applications. The section 6.4 will show some results of an attempt to use the derived relation to calculate transport coefficients and the analysis and discussion of the results.

6.2 Derivation of the analogous expression of the generalized E-P flux theorem in isentropic coordinates

We start from the continuity equation for tracer mixing ratio:

$$\frac{D\chi}{Dt} = S \tag{6.50}$$

where χ is tracer mixing ratio; S is the net source denoting the sum of all production and loss processes. In the atmosphere, the source represented by S must include not only chemical production and loss but also the effects of turbulent diffusion by the unresolved scales of motion. For this reason alone tracers are never exactly conserved in the observational analysis of atmosphere data. However if the resolved scales include most of the tracer variance, the diffusion source can in many cases be neglected (Andrews et al. 1987). Attention is drawn to the

fact that the form of equation (6.50) is the same as the conservation equation for potential temperature and potential vorticity equation, with the S appropriately defined.

Isentropic coordinates have some of the conceptual advantages of generalized Lagrangian mean formalism, without many of the accompanying technical difficulties. For instance in the isentropic coordinates the vertical component of meridional circulation depends solely on diabatic heating. The present derivation of an analogous expression of GEP theorem is confined to the version for isentropic coordinates only. In the isentropic system, the vertical velocity is proportional to diabatic heating rate:

$$\frac{D\theta}{Dt} = Q \tag{6.51}$$

The tracer continuity equation of (6.50) in these coordinates can be expanded as:

$$\chi_t + u\chi_x + v\chi_y + Q\chi_\theta = S \tag{6.52}$$

The 'density' in this system is defined as $\sigma \equiv g^{-1}\frac{\partial p}{\partial \theta}$ as in chapter 5.

The linearized disturbance equation of (6.52) is:

$$\overline{D}\chi' + v'\overline{\chi}_y + Q'\overline{\chi}_\theta = S' + O(\alpha^2) \tag{6.53}$$

where $\overline{\chi}_t$,$\overline{v}$, $\overline{Q}$, $\overline{S}$ is assumed to be $O(\alpha^2)$ and α is the disturbance amplitude. $\overline{D} = \frac{\partial}{\partial t} + \overline{u}\frac{\partial}{\partial x}$ is the material derivative following the mean flow.

Multiply each term of equation (6.53) by $a\overline{\sigma}^2\chi'/\overline{\chi}_y$ and average along a latitude circle leading to:

$$\frac{1}{2}(a\overline{\sigma}^2\overline{\chi'^2}/\overline{\chi}_y)_t + \frac{a\overline{\sigma}^2}{\overline{\chi}_y}(\overline{\chi'v'}\overline{\chi}_y + \overline{\chi'Q'}\overline{\chi}_\theta) = a\overline{\sigma}^2\overline{\chi'S'}/\overline{\chi}_y + O(\alpha^2) \tag{6.54}$$

Following Andrews et al. (1987), we also define perturbation displacement quantities (η', q', γ') by:

$$\overline{D}(\eta', q', \gamma') = (v', Q', S') + O(\alpha^2) \tag{6.55}$$

From equation (6.53) and equation (6.55) we then obtain:

$$\chi' = -\eta'\overline{\chi}_y - q'\overline{\chi}_\theta + \gamma' + O(\alpha^2) \tag{6.56}$$

Writing

$$\hat{\mathbf{F}} = (\overline{\chi'v'}, \overline{\chi'Q'}) = (F^{(y)}, F^{(\theta)}) \tag{6.57}$$

and

$$\nabla\overline{\chi} = \left(\frac{\partial\overline{\chi}}{\partial y}, \frac{\partial\overline{\chi}}{\partial\theta}\right) = (\overline{\chi}_y, \overline{\chi}_\theta) \quad . \tag{6.58}$$

It can be shown after some manipulation (see appendix) that:

$$\hat{\mathbf{F}} \cdot \nabla\overline{\chi} = \overline{\chi'v'}\overline{\chi}_y + \overline{\chi'Q'}\overline{\chi}_\theta = -\nabla\overline{\chi} \cdot (\mathbf{K}^{(s)} \cdot \nabla\overline{\chi}) + \overline{v'\gamma'}\overline{\chi}_y + \overline{\gamma'Q'}\overline{\chi}_\theta \tag{6.59}$$

where

$$\mathbf{K}^{(s)} = \begin{pmatrix} K_{yy} & K_{y\theta} \\ K_{\theta y} & K_{\theta\theta} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} (\overline{\eta'^2})_t & (\overline{\eta' q'})_t \\ (\overline{\eta' q'})_t & (\overline{q'^2})_t \end{pmatrix} \tag{6.60}$$

is the symmetric 'diffusion' tensor in isentropic coordinates.

Replacing the second term on the left hand side of (6.54) with (6.59) gives:

$$\frac{\partial A^*}{\partial t} + \left(- \nabla \overline{\chi} \cdot (\mathbf{K}^{(s)} \cdot \nabla \overline{\chi}) \right) \frac{a\overline{\sigma}^2}{\overline{\chi}_y} = S^* + O(\alpha^2) \tag{6.61}$$

where A^* and S^* are defined as:

$$A^* = \frac{1}{2} a\overline{\sigma}^2 \overline{\chi'^2} / \overline{\chi}_y \tag{6.62}$$

$$S^* = a\overline{\sigma}^2 \overline{\chi' S'} / \overline{\chi}_y - \overline{v'\gamma'} + \overline{\gamma' Q'} \overline{\chi}_\theta / \overline{\chi}_y \tag{6.63}$$

A^* is analogous to A of the previous chapter with wave activity density defined by tracer mixing ratio in the place of potential vorticity and S^*, analogous to S, is the nonconservative term.

Equation (6.61) is the derived expression for chemical tracers which is analogous to the generalized E-P flux theorem. The first term and the nonconservative term are directly analogous to the corresponding terms in the generalized E-P flux theorem but the second term on the left is not in complete divergence form as in the original theorem.

The appearance of non-divergence form of the second term in the analogous expression is due to the difference between chemical tracers and dynamical tracers, say

between ozone mixing ratio and potential vorticity. The dynamical tracers such as potential vorticity and potential temperature are all derived quantities and have their 'definition' in terms of more basic dynamical quantities. This characteristic allows them to be further written as the form appearing in the generalized E-P flux theorem, whereas mixing ratio of chemical tracers are not derived quantities and hence have no 'definition', which prevents the attempt to write the second term in the analogous expression in full divergence form. However, taking into account some frequently used assumptions, a simplified version of the analogue expression can show us the link between the chemical wave transience and transport coefficients for 2-dimensional models. The simplification and implication of the analogous expression will be discussed in the next section.

6.3 The simplification of the analogous expression

If the eddies in consideration are adiabatic, i.e., $Q' = q' = 0$, then the $K_{\theta\theta}$, $K_{y\theta}$, $K_{\theta y}$ components in the diffusion tensor of equation (6.60) all vanish. The analogue expression (6.61) thus can be simplified as:

$$\frac{\partial A^*}{\partial t} + a\overline{\sigma}^2(-K_{yy}\overline{\chi}_y) = S_1^* + O(\alpha^2) \tag{6.64}$$

here

$$S_1^* = a\overline{\sigma}^2\overline{\chi' S'}/\overline{\chi}_y - \overline{v'\gamma'}$$

If we further define,

$$\hat{A} = \frac{1}{2}\overline{\chi'^2}/\overline{\chi}_y \tag{6.65}$$

$$S_1 = S_1^*/a\overline{\sigma}^2 \tag{6.66}$$

then equation (6.64) can be written as:

$$\frac{\partial \hat{A}}{\partial t} + (-K_{yy}\overline{\chi}_y) = S_1 + O(\alpha^2) \tag{6.67}$$

Equation (6.67) is the simplified analogous expression. Within the extent of linear assumptions this simplified analogous expression shows explicitly the relations among the transport coefficient, chemical wave transience and sources and sinks of the tracer in consideration. As the defined chemical wave activity deals only with deviation of chemical mixing ratio from its zonal mean in Eulerian form and has no need of knowledge of parcel displacement or trajectory, it is also possible to estimate the transport coefficients through equation (6.67) if the fields of mixing ratio and source (or sinks) field are provided.

Particularly if the tracer in consideration is quasi-conservative or its life time is much greater than dynamical time scale, then $\gamma' = S' = 0$ leading to $S_1 = 0$, equation (6.67) becomes:

$$\frac{\partial \hat{A}}{\partial t} + (-K_{yy}\overline{\chi}_y) = 0 \tag{6.68}$$

This gives directly the relation between wave transience and its transport effect on the mean field.

If $\chi_y \neq 0$ (if $\chi_y = 0$, the tracer is meridionally well mixed, $\hat{A}$ defined above is no longer applicable and equation 6.68 no longer necessary) then equation (6.68) also states that for steady, conservative waves $K_{yy} = 0$. This means that there is no net transport of the tracer by eddies. Taking into account the relation between transport by eddies and by mean circulations, the disappearence of eddies also impl the disappearence of the mean transport and the distribution of the tracer will be solely determined by photochemical or chemical source and sinks. This inference is the same as ~~that~~ from the nontransport theorem (Andrews et al. 1987).

Equation (6.67) and (6.68) are now established. We will further discuss their implications for transport processes in the following paragraph and section.

The eddy transport of trace species can result from several physical processes as well as chemical processes. The physical processes can contribute to transport of tracers from dispersion of air parcels, which in the presence of a mean gradient leads to mixing, and from eddy advection normal to the mean gradient, and from eddy induced mean circulation transport. The chemical contribution can result from chemical eddies as discussed by Smith (1988). The whole transport processes are

normally three dimensional and zonally asymmetric. Because present computing capabilities are insufficient to include full stratospheric chemistry in three dimensional models , zonally averaged two-dimensional models are still important tools for research in the interaction between dynamics and chemistry in the stratosphere. These two dimensional models do not explicitly include wave processes and deal only with zonal mean quantities. The eddy driving of wind and temperature mean state is expressed by E-P flux divergence. To tackle the tracer transport in these 2-dimensional models, we have to parameterize the real 3-dimensional processes by zonal mean quantities.

Parameterization of transport by eddy flux is commonly based on the assumption of flux-gradient relation, i.e.,

$$\overline{\mathbf{v}'\chi'} = -\mathbf{K} \cdot \nabla\overline{\chi} \tag{6.69}$$

where $\mathbf{v}'\chi'$ represents the meridional and vertical flux of the tracer χ, and $\mathbf{K}$ is a 2×2 matrix, often called the diffusion matrix. The matrix $\mathbf{K}$ is not limited to diffusive (down-gradient) transport though by name it seems to imply this. By various assumptions and approaches the relation expressed by (6.69) can represent both dynamical (dispersion) and chemical eddy transport (Rood and Schoeberl,1983; Plumb and Mahlman,1987; Smith, 1988). For example, by assuming that the transport processes are independent and additive, $\mathbf{K}$ can be split into several components caused by dispersion, divergence, chemical diffusion etc. (Andrews

1986; Smith 1988; Pyle and Rogers 1984; Rogers and Pyle 1980; Holton 1981; Matsuno 1980; Plumb 1979).

The current approaches to estimat the diffusion matrix **K** in equation (6.69) can be put into three categories by method. The first method addresses the problem directly through the assumed flux-gradient relation; the diffusion tensor is obtained by inversion of (6.69) by inserting the gradients and eddy flux statistics from independent tracer transport experiments of three dimensional circulation models. In this way Plumb and Mahlman (1987) produced the dispersive part of the transport matrix. Using their experiments of three-dimensional transport in a general circulation model, they also computed the meridional eddy transport of an idealized near conservative tracer by the flux-gradient relation. They calculated the symmetric (diffusive) part of the conservative transport matrix and included the asymmetric (advective) part in their zonal mean transport velocity.

The second method relies on the estimation of parcel displacement or trajectory to estimate transport tensors by their definitions, either dynamical or chemical. Lyjak (1987) calculated the three-dimensional trajectories of a large number of air parcels for a 10 day period, using winds derived from observations. By measuring dispersion as increase in the mean square distance of air parcels from their mean (zonally symmetric) latitude and height and using a definition like (6.60), he computed the transport tensor due to dispersion. Similar methods were used by

Kida (1983). Kinnersley (personal communication 1989) developed a particular way to estimate the parcel displacements on isentropic surfaces and used these displacements to derive the transport tensor for isentropic coordinates. Smith (1988) also used the parcel displacements derived by Lyjak and Smith (1987) as a starting point for calculating the contribution to the tracer transport matrix from chemical sources and sinks. Assuming steady, linearized production/loss, she managed to obtain transport coefficients resulting from several chemical processes, e.g. from simple dependence of loss rate on concentration, on wave structure of temperature and inter-dependence of two species. The last component offers a possibility of having a different transport coefficient for each tracer, as deemed necessary by Pyle and Rogers (1980). By calculating the transport matrix from LIMS data and comparing her resultant transport with the real transport revealed by LIMS observation, Smith was able to show that the chemical eddies play an important role in the transport processes, especially for ozone transport in the early winter of the LIMS year.

The third method for deriving K is based on the similarity between potential vorticity and chemical tracers. Tung (1987,1989) takes into account the fact that it is the planetary waves that are responsible for driving the atmosphere away from equilibrium. Ertel potential vorticity in isentropic coordinates is approximately conserved on a dynamical time scale of 1-2 weeks, and its transport

can be considered as representative of the flux of other quantities with similar life times. The meridional flux of potential vorticity is further related to E-P flux pseudo-divergence, the wave driving of the zonal mean state in isentropic coordinates. Assuming a flux-gradient relation between Ertel potential vorticity and its meridional flux, and showing that the meridional flux is further related to E-P flux pseudo-divergence by a generalized version of the geostrophic Taylor (1915) relation, the transport coefficient K_{yy} can then be derived by inversion of the equation between E-P flux pseudo-divergence and Ertel potential vorticity gradient (parameterized relation). Based on the planetary wave forcing of the zonal mean state, by obtaining E-P flux pseudo-divergence through computation of the residuals of zonal mean zonal momentum equation (diabatic circulation thus needed) and estimating the gradient of Ertel potential vorticity by mean circulation with chemical eddy effects neglected, Tung derived transport coefficients and claimed they were consistent with advective transport. Newman et al. (1986) applied the same method to quasi-geostrophic potential vorticity and also got a K_{yy}. His method and that of Tung (1987) contain both dispersive and chemical eddy contributions to transport due to inclusion of radiative or mechanical damping of the waves. Hitchman and Brasseur (1988) derived a method of computing the meridional potential vorticity flux associated with steady damped waves in their 2-dimensional model. They then use the flux-gradient relation to determine K_{yy} due to chemical eddy effects (damping) consistent with their model.

Most of the above methods require either the parcel displacements or trajectories, which are often difficult to determine from observational data in the middle atmosphere, or give identical transport coefficients for all tracers. By equation (6.68), however, it is possible, theoretically to deduce different dispersive transport coefficients for different tracers. If the chemical source/sink effect is included properly, it is also possible to derive the chemical eddy contribution and hence total coefficients for each tracer. The data can be either from observations or 3-dimensional model output. An attempt to apply equation (6.68) to LIMS observational data will be discussed in the next section.

6.4 Results and analysis

In the present study, an attempt is made to apply equation (6.68) to deduce transport coefficients. The general idea is that the monthly mean values of the gradients of tracer mixing ratio, $\bar{\chi}_y$, and the variance, $\overline{\chi'^2}$, are firstly estimated from LIMS data. The analogous wave activity is then derived from equation (6.65) and this leads to a knowledge of the analogous wave transience. K_{yy} is thus deduced by recasting equation (6.68).

The estimation of K_{yy} is confined to LIMS observational data. The Fourier coefficients of LIMS data are recombined to reconstruct a 4x15 degree latitude and

longitude grid of the daily ozone mixing ratio field. The zonal mean fields of ozone mixing ratio can be obtained directly from the original data. Figure 6.1 shows the zonal mean fields of ozone mixing ratio for November, December, January and February for the northern hemisphere. We see that the maximum is located above the equator at about 32km height for all four months and the meridional gradient is negative (equatorward) in most regions except near the lower and upper boundaries where it reverses. From November 1978 to February 1979 the gradient near the polar area (north of $60°N$) decreases continuously. On account of the existence of an area where the mean gradient changes signs and where the gradient is very small, the direct application of equation (6.68) will have problem of singularities in these areas. In such case, interpolation or extrapolation from the surrounding values are performed.

The daily gradient of zonal mean ozone mixing ratio is then produced by central differencing. Figure 6.2 gives the monthly mean of these gradient fields for the above mentioned months. We can see that the areas of positive gradient exist in all the four months in the subtropics below about 25km above about 38km with negative gradients elsewhere giving zero lines in the transition regions. The main negative area had two large centres in November, one at $30°N$ 32km and another at $45°N$ 40km. These two centres moved southward from November to January and merged into one in February at about $23°N$. The whole trend was toward

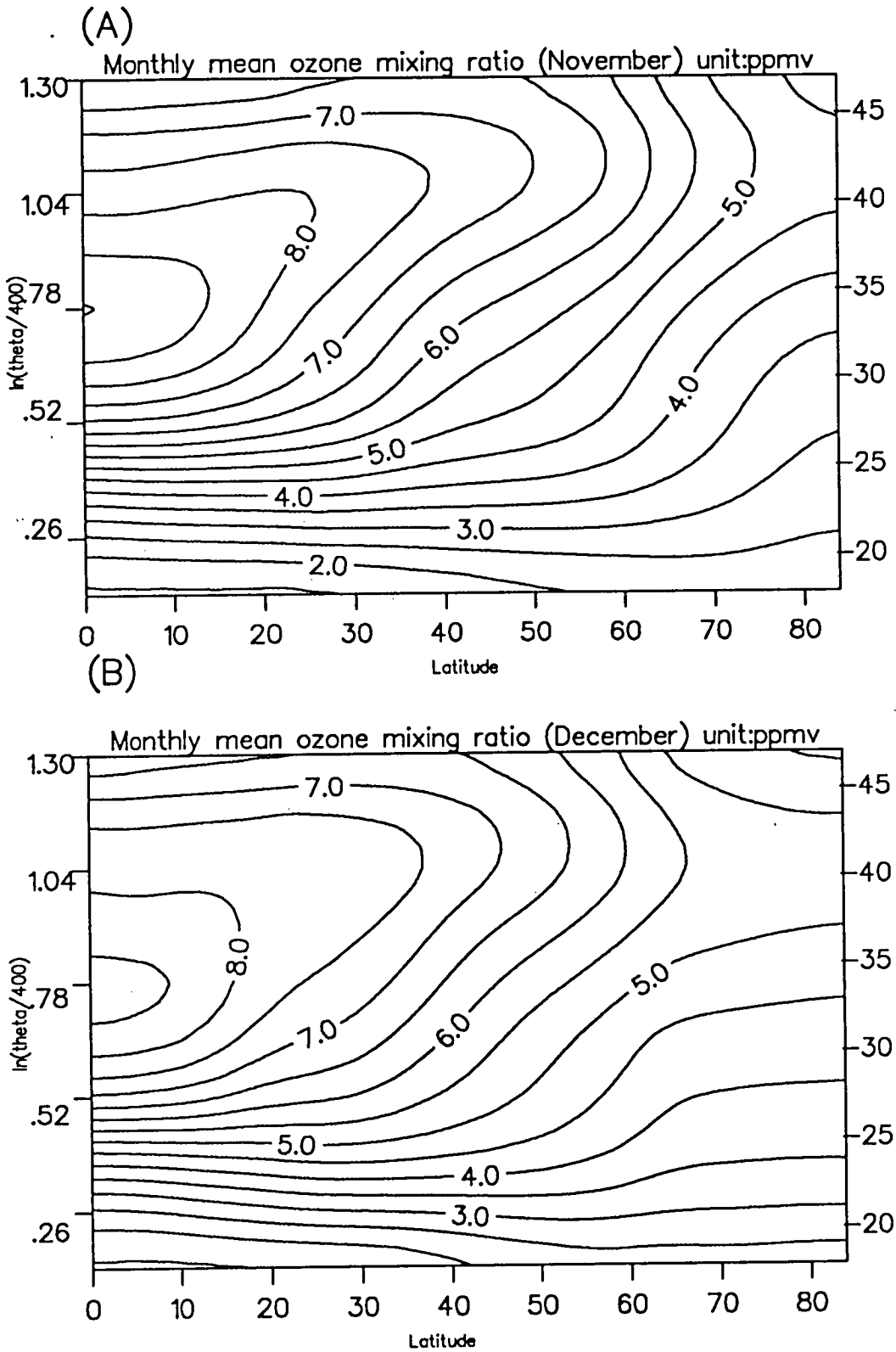
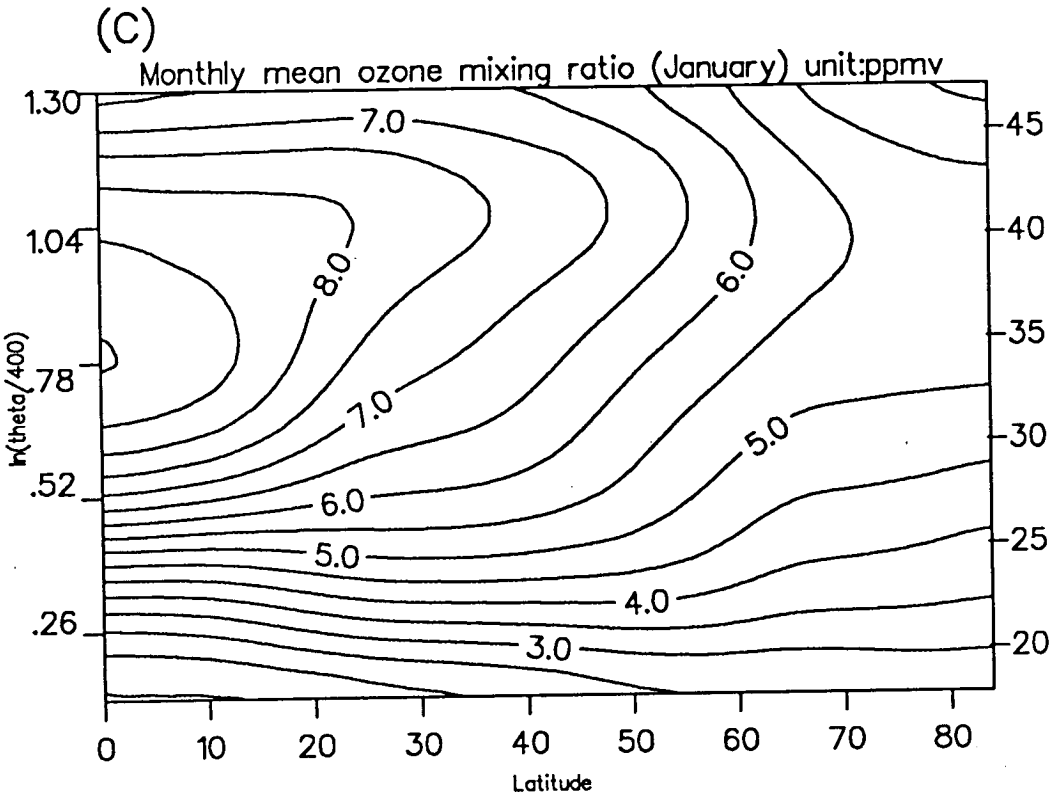


Figure 6.1A,B Monthly mean ozone mixing ratio with (A) for November and (B) for December. Unit:ppmv



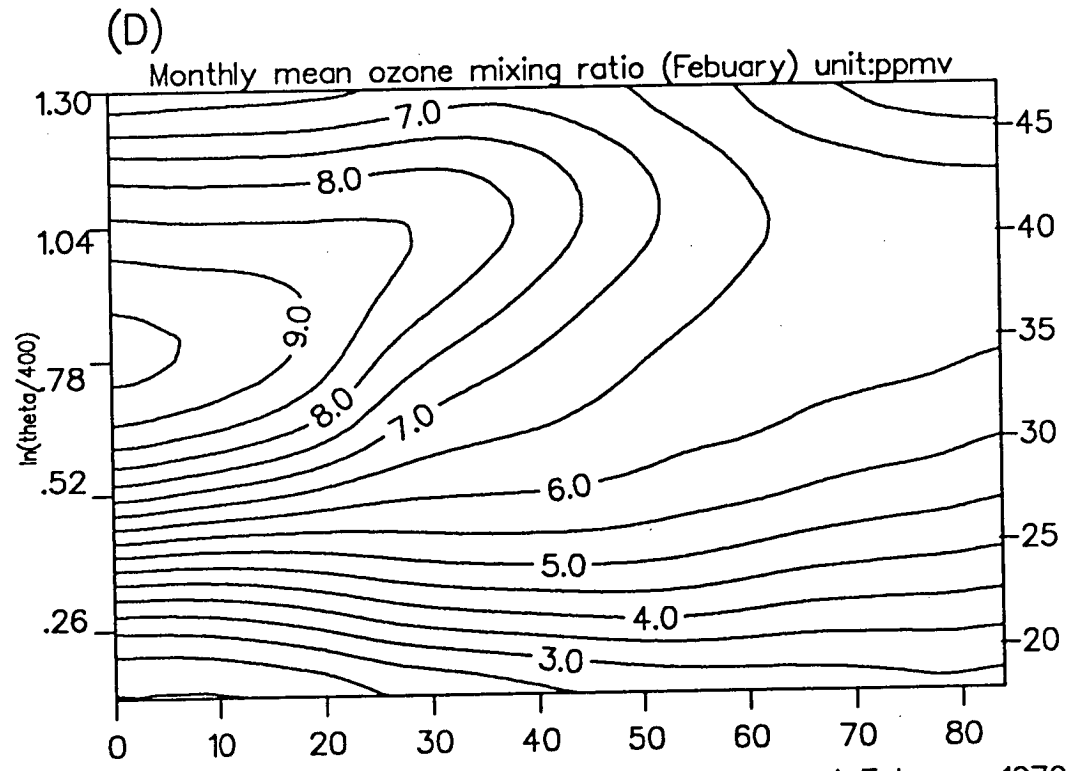


Figure 6.1C,D As figure 1A,B but for January and February 1979

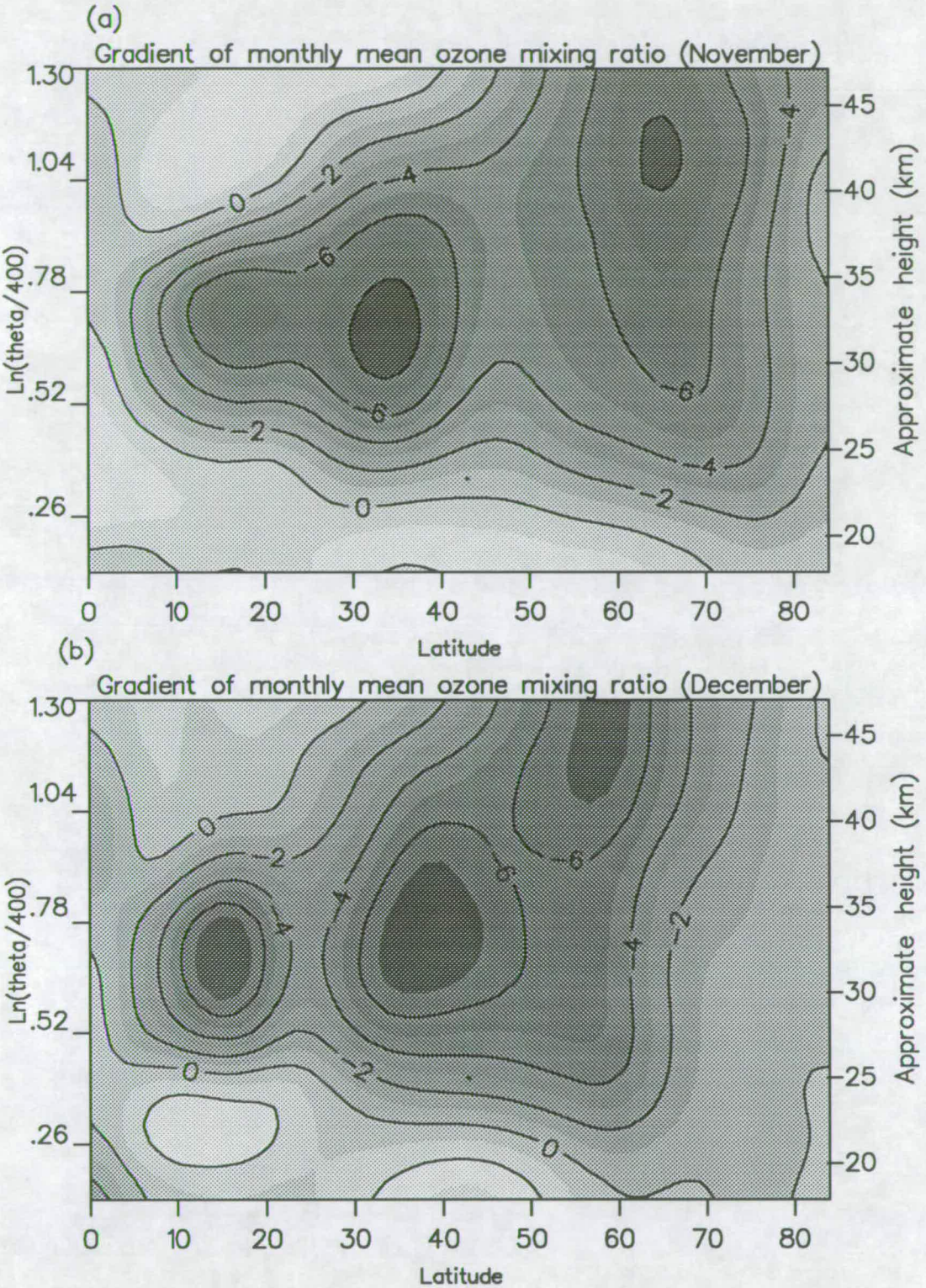


Figure 6.2a,b Monthly mean gradient of ozone mixing ratio. Unit: 1.0e−13ppv/m with (a) for November and (b) for December 1978.

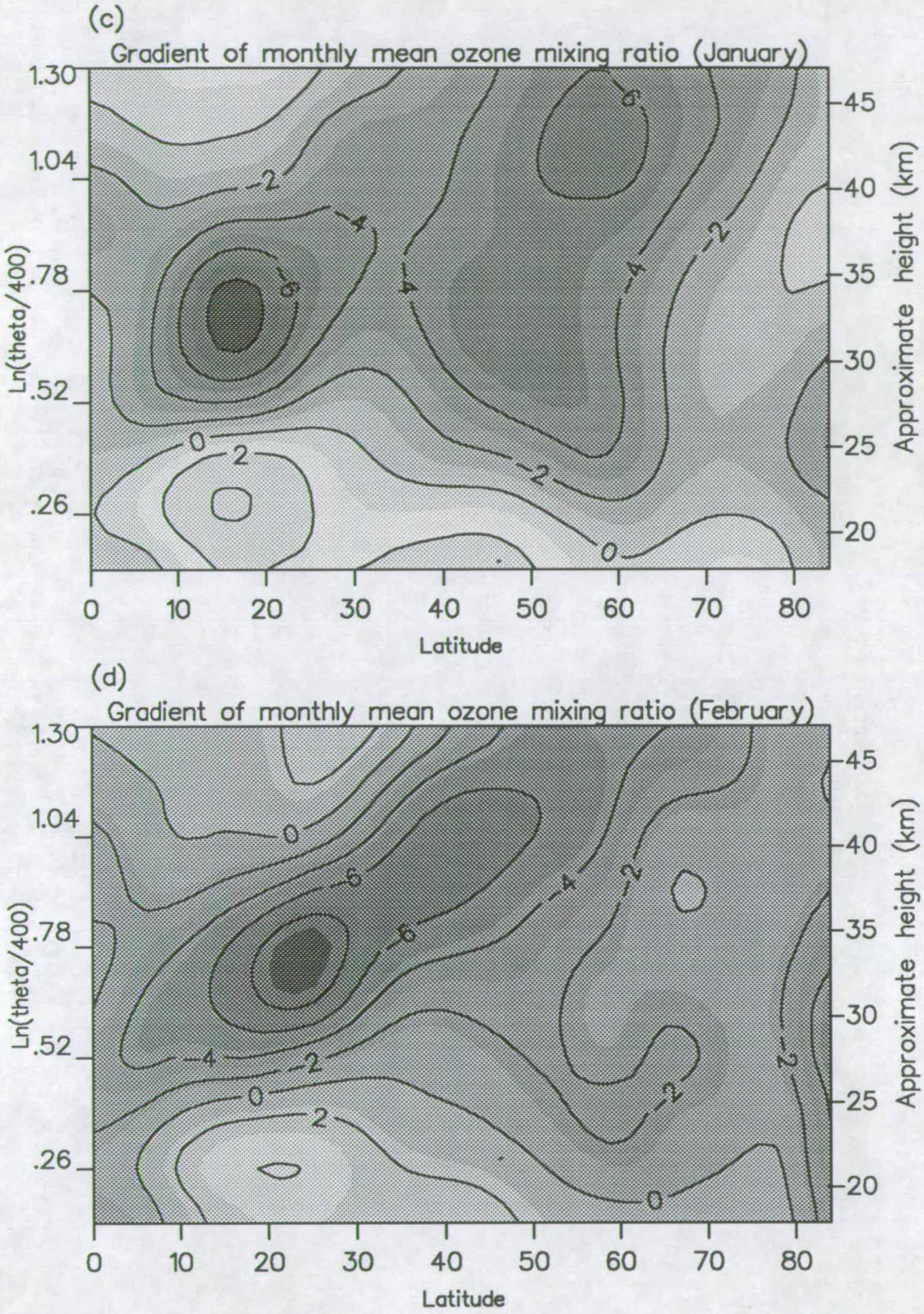


Figure 6.2c,d As figure 2a,b but for January and February 1979.

the weaker gradients in the mean ozone distribution, especially in the mid-high latitudes. This was consistent with the planetary waves transport of high ozone air from low latitudes into high latitudes given the fact that there were three stratosphere warmings (large wave breaking events) during the period.

To have a background of the general features of monthly mean tracer eddies or eddy variance of tracer, figure 6.3 shows the monthly mean of the daily values of $\overline{\chi'^2}/2$. There are two large centres throughout the four months. One is near $70°N$ 42km in November. Its position remained almost steady from November to January and moves southward and downward to near $62°N$ 38km in February. Its centre value decreases from November to December and increases to its strongest in January and decreases again in February. The other maximum is relatively weak in November located at $32°N$ and 35km, but it continues to strengthen and moves downward and northward during the four month period, reaching its strongest value and lowest position near $68°N$ 25km in February. This is also in accord with the planetary wave theory. During the stratospheric warmings when the strongest wave breaking event happens, the long and thin tongue of low latitude, ozone-rich air is drawn deeply into the pole and high latitude ozone-poor air moves into low latitudes. This naturally leads to large wave amplitudes in ozone mixing ratio and hence the high value of monthly averaged variance of ozone mixing ratio shown in figure 6.3d.

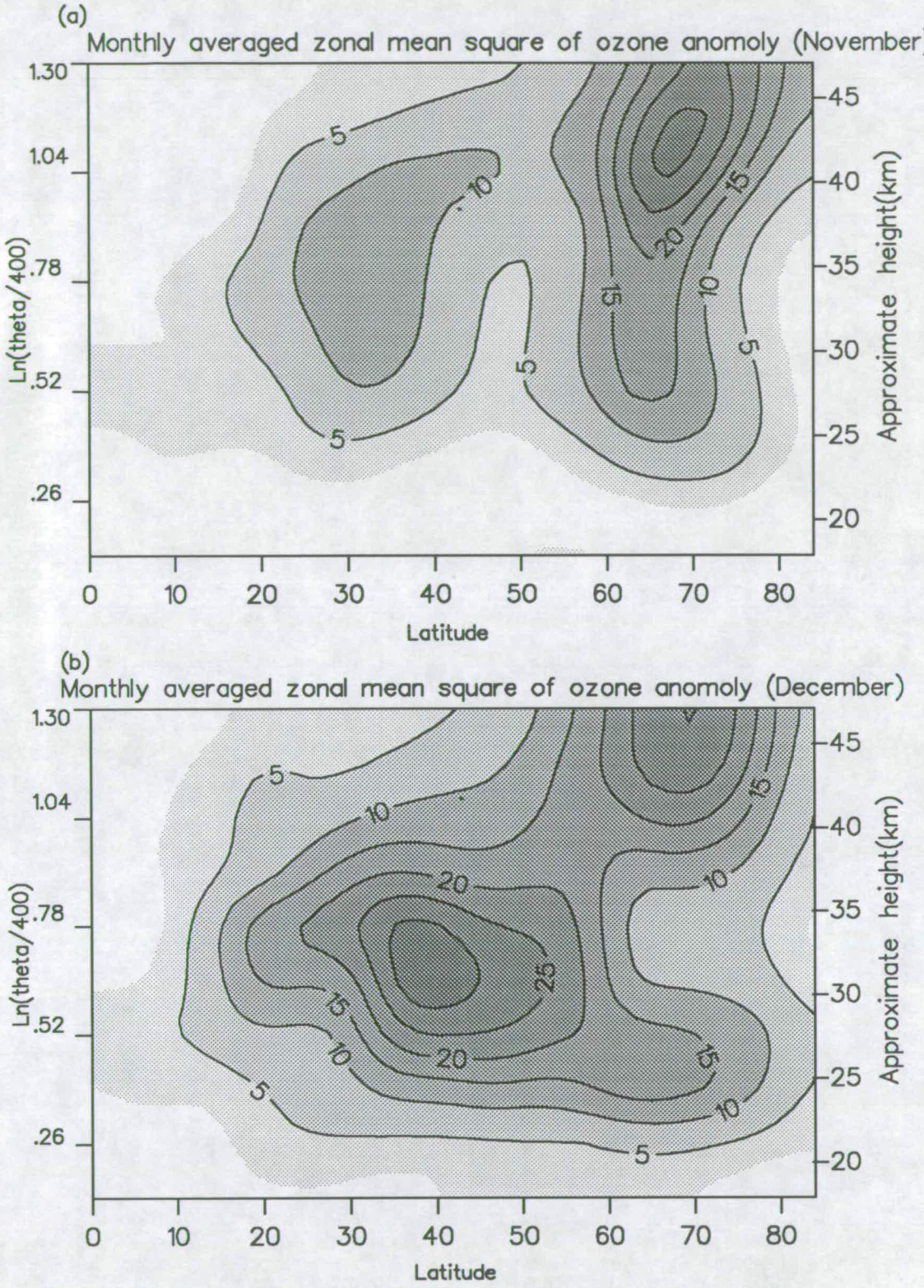
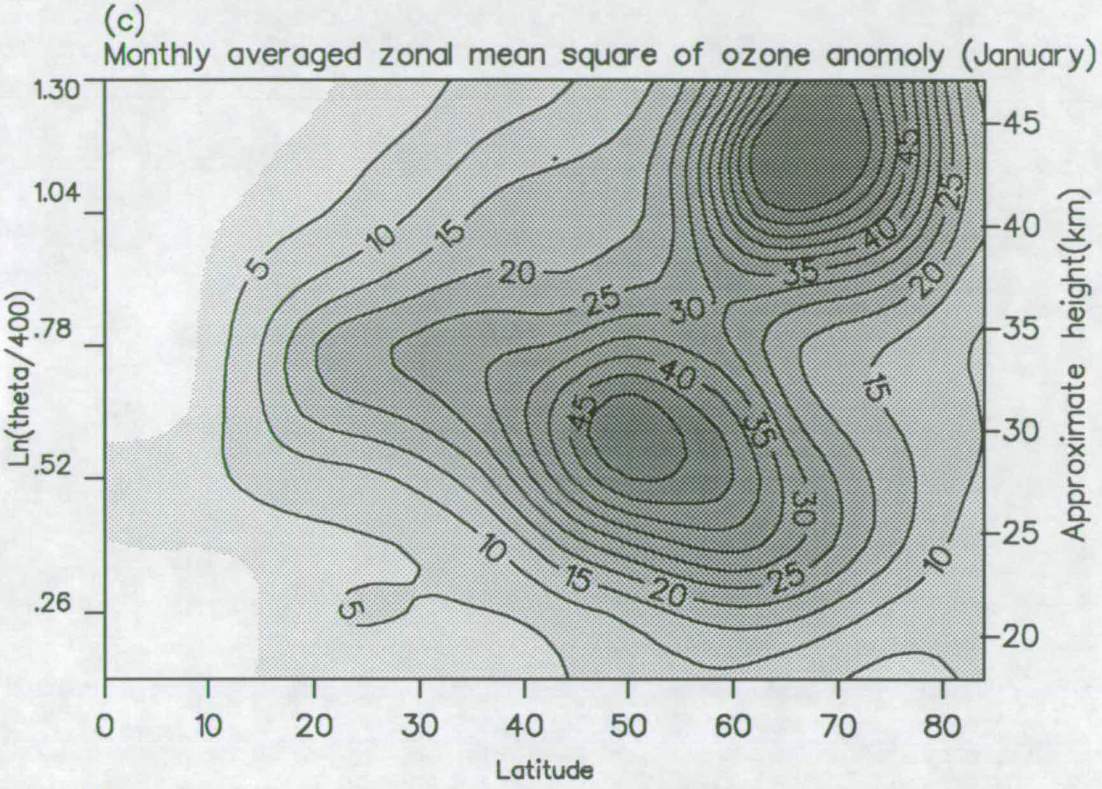


Figure 6.3a,b Monthly averaged zonal mean variance of ozone mixing ratio. with (a) for November and (b) for December 1978. Unit: 1.0e−14(ppv)**2



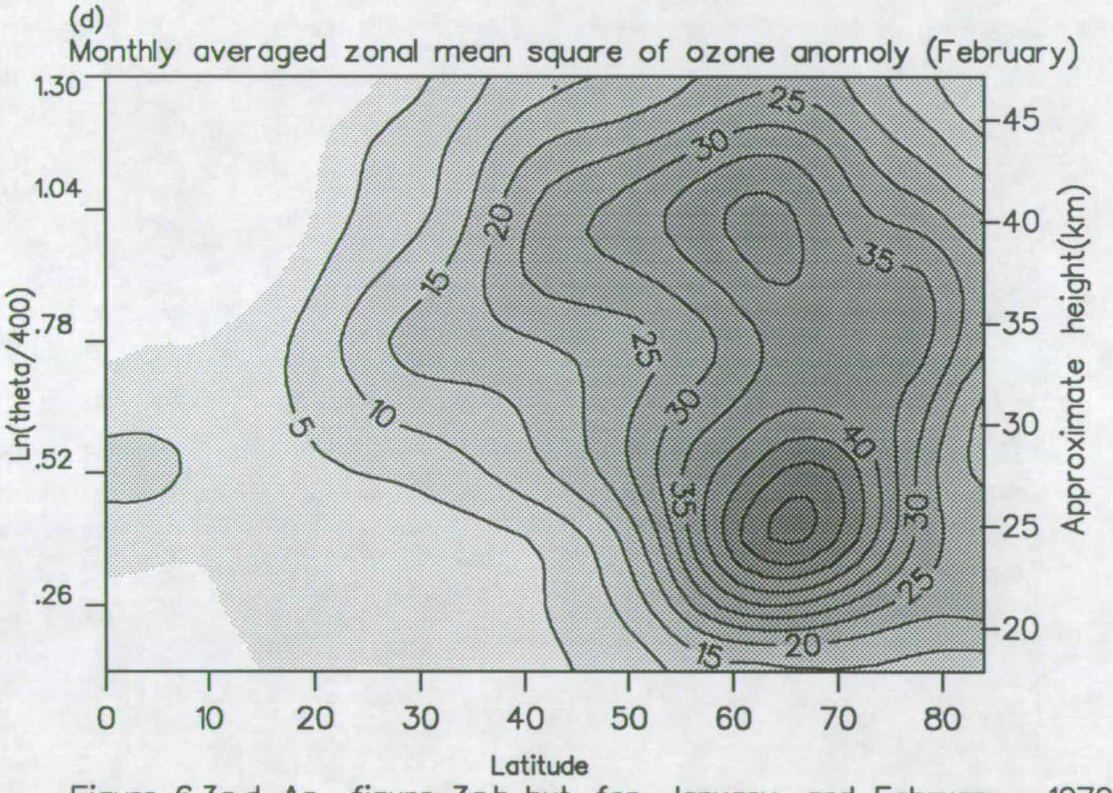


Figure 6.3c,d As figure 3a,b but for January and February 1979.

Assuming conservative and adiabatic conditions and integrating over a time period $\triangle t = t_2 - t_1$, the pure dispersive transport coefficients may be deduced from equation (6.68), i.e.

$$K_{yy} = \frac{A^*|_{t_2} - A^*|_{t_1}}{\triangle t \overline{\bar{\chi}_y}^{\triangle t}} \approx \frac{\overline{\chi'^2}|_{t_2} - \overline{\chi'^2}|_{t_1}}{2\triangle t(\overline{\bar{\chi}_y}^{\triangle t})^2} \tag{6.70}$$

This expression gives an approximation to estimate the pure dispersive transport coefficient by tracer wave transience. The K_{yy} obtained by this approximation and parameterized eddy flux as well as some intermediate products will be presented and discussed in the following paragraphs.

Figure 6.4 gives the difference field of zonal eddy variances of ozone mixing ratio, an intermediate product for the calculation of K_{yy} from equation (6.70), where t_2 is taken as the last day of a month and t_1 the last day of the month before. Figure 6.4a is the difference for November 1978 figure 6.4b for January 1979. In November the differences of eddy variances in the lower stratosphere are positive with two maxima located at $70°N$, 27km and $30°N$, 32km respectively. The negative values are prevailing above 35km with a patch extending to lower levels in midlatitudes. In January, however, the distribution is different with positive values occuping almost all the area north of $40°N$ and the negative values are on the southern side. There are two maxima in the positive region, one is around 62°, 25km and the other around $76°N$, 40km. Equation (6.70) shows that the signs in the difference field will determine the signs in the field of transport coefficients in

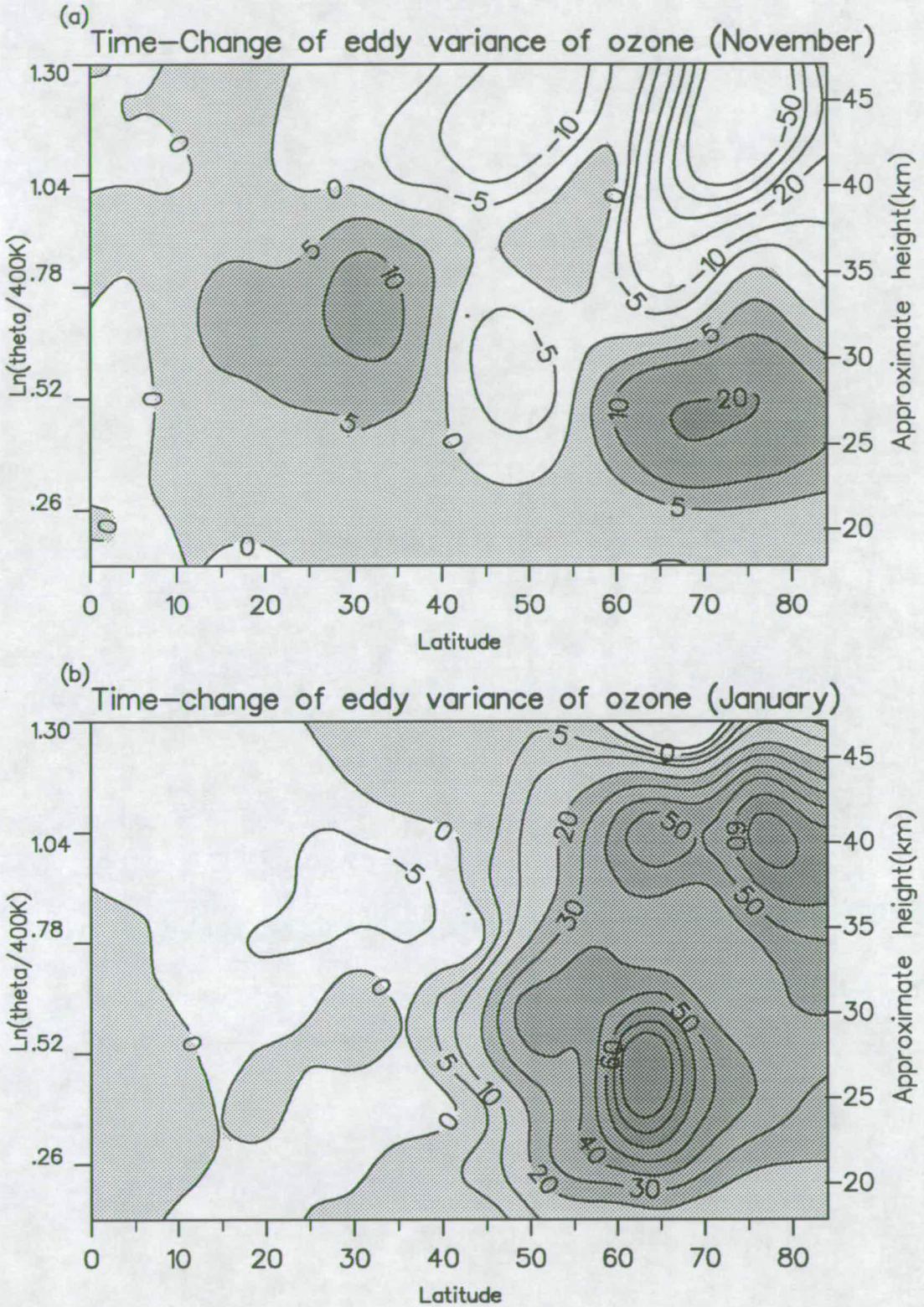


Figure 6.4 Difference of eddy variance of ozone mixing ratio. (a) for November 1978 and (b) for January 1979. Unit:1.0e–14 ppv**2. (see text for detail).

most areas. The large area of negative values will lead to a negative area in the transport coefficients as will be seen below.

The above calculations were carried out on the rate of change of the eddy variance, $\overline{\chi'^2}$, over a time span of a month. It seems possible to integrate equation (6.68) on a daily basis and then average the daily values of K_{yy} to give the monthly mean transport coefficients. However such calculations on a daily basis were found to be rather noisy and were not adopted. The fields of figure 6.4 represent the real wave transience of the observed ozone fields. Since the photochemical timescale of O_x family is about a week in the lower stratosphere (Brasseur and Solomon, 1984), similar to the transport timescale, the distribution of ozone is controled by both dynamical transport and chemical processes in the lower stratosphere. Ozone is considered quasi-conservative only on time scales of the order of a week or less (Butchart and Remsberg 1986), therefore the fields represented by figure 6.4 may be caused by all processes in the whole dynamic-radiative-chemical system. The factors which contribute to this deviation are all the terms in the equation (6.61), plus others which may arise from finite amplitude effects.

Figure 6.5 shows the resultant K_{yy} from (6.70). In agreement with figure 6.4, the transport coefficients are positive in the lower stratosphere in November and in almost all the areas north of $40°N$ in January. The maximum in January is around $75°N$, 30 to 37km.

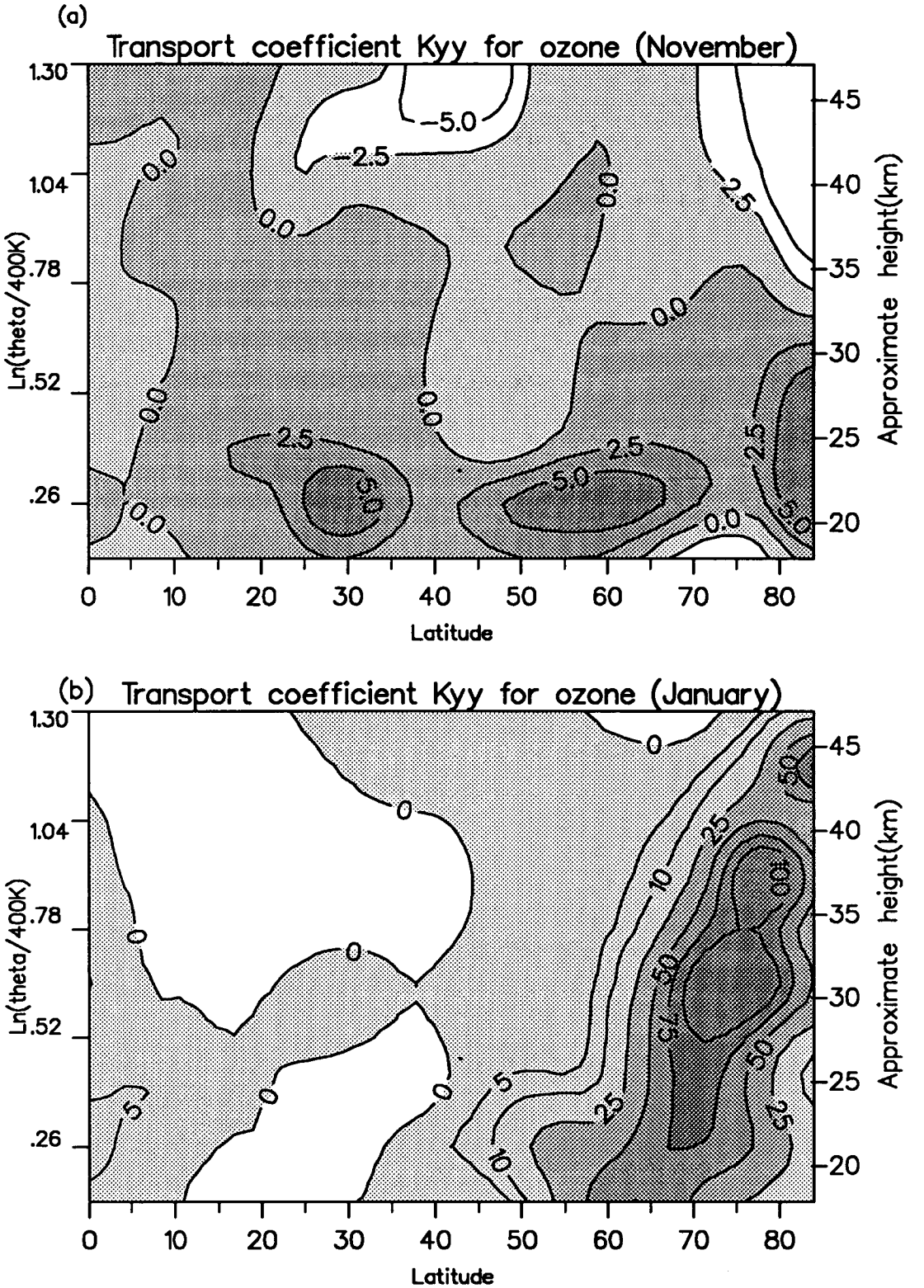


Figure 6.5 Transport coefficients derived from the derived expression for chemical tracers. Ozone is used in this figure. (a) for November 1978, (b) for January 1979. Unit: 1.0e5.

These K_{yy} fields will be discussed together with the parameterized flux field in the forthcoming paragraph. In practice the main problem is the existence of a large area of negative K_{yy}; while this is theoretically permitted, it may cause model instability.

Figure 6.6 shows the parameterized ozone flux by multiplying transport coefficients in figure 6.5a,b with their monthly mean fields of the gradient respectively. For comparison figures 6.6c,d give the real observed eddy ozone flux calculated from LIMS observational data by Smith(1988). There are agreements between the parameterized flux and the real flux. The main positive flux areas of the real flux in high latitudes at low levels in both months are roughly produced, with the magnitude and position in January very similar to those real flux, though the magnitudes in November are only half of those real values. The reversal of the real flux from positive in lower levels to negative in higher levels in November are also roughly produced, although the transition appears lower and the magnitude of the negative values larger. The most conspicious feature is that the negative values in low to mid-latitudes in January is too large and extends too far north at very low levels and that the vertical reversal of the flux in January is only very weakly reproduced. Another problem is that in November the high level negative area is split and there is a small patch of negative in the low level around $45 - 50°N$ in figure 6.6a which seems spurious. It should be borne in mind that besides

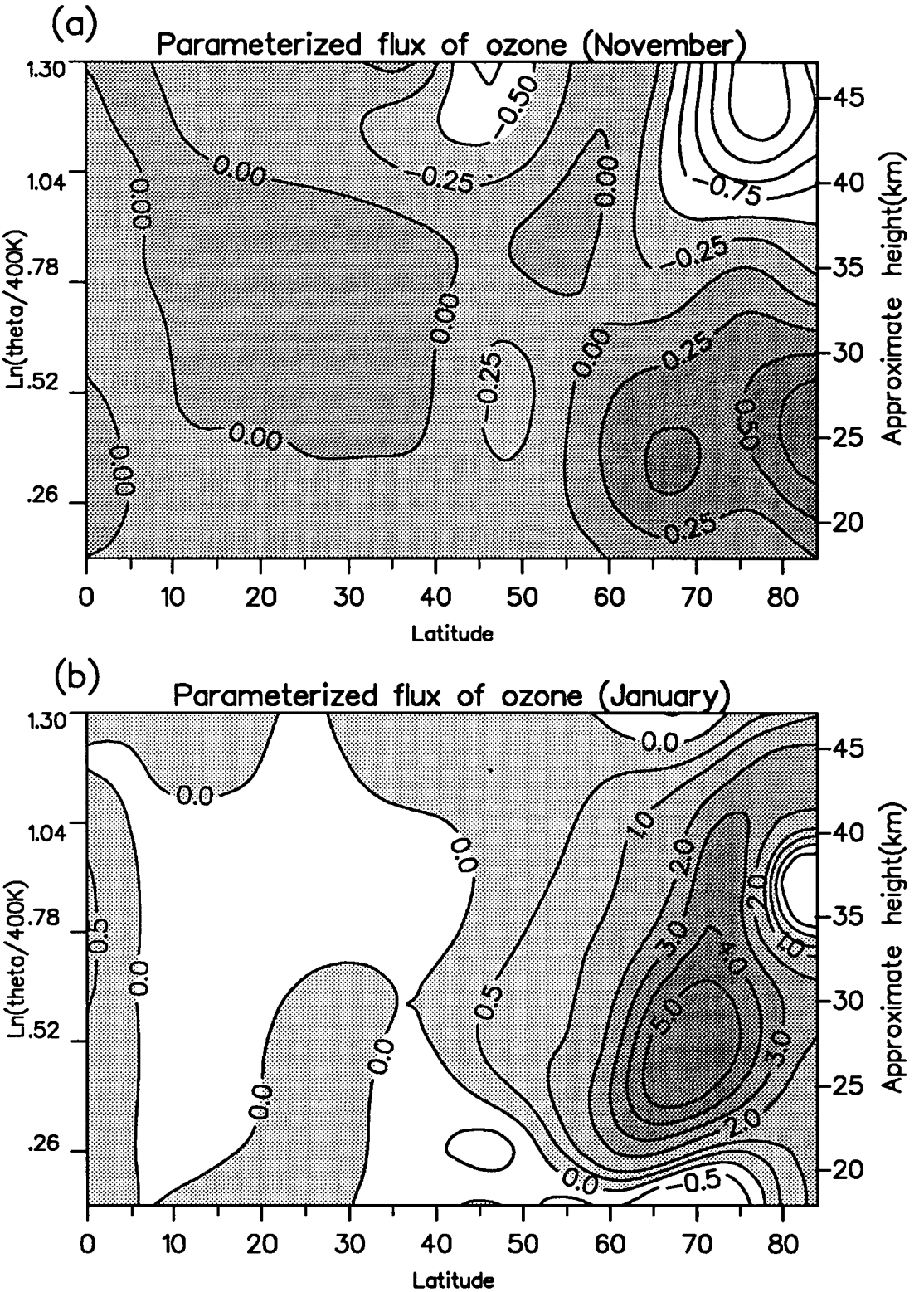


Figure 6.6 Parameterized ozone flux using transport coeffients in figure 6.5. (a) for November 1978, (b) for January 1978. Unit: ppmv (m/s).

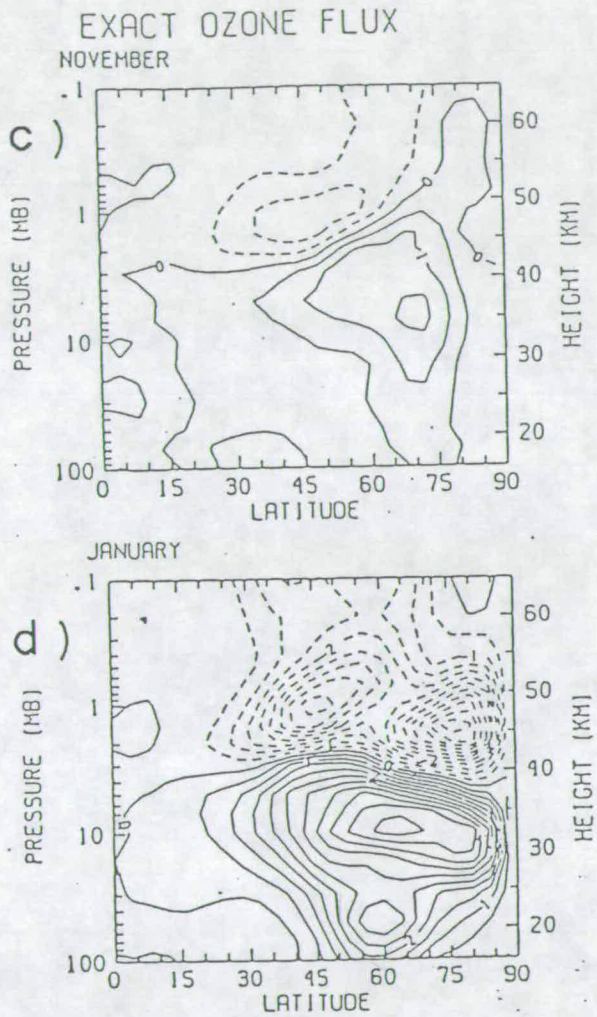


Figure 6.6: Observed meridional eddy ozone transport, calculated from observed waves for (c) November, (d) January. Units are ppmv (m/s). [From Smith, 1988] The interval of contours is 0.5.

dispersion the exact flux includes all processes acting in the atmosphere, such as the contribution from chemical sources and sinks and the diabatic effects, whereas the present calculation considers only the contribution by the effects of dispersive transport approximated by wave transience. To reproduce, to some approximate extent the real flux, we need to take into account all major processes, which is however beyond the present scope.

In order to contrast and to understand better the contribution of dispersion to transport processes, Smith's (1988) parameterization of flux by chemical eddies (sources and sinks) is shown in figure 6.6e,f.

She reproduced the reversal of the meridional flux but failed to produce the correct magnitude of the positive flux area in the low level which she attributed to possible effects of dispersion. From the wave breaking theory, the dispersion is expected to be strongest in regions of weak zonal wind. In January and February 1979 three stratospheric warmings occurred and these would increase the dispersive contribution to transports. The fact that Smith's chemical eddy parameterization is worse in the later winter than the early winter is a further evidence of the importance of dispersive effects.

It is necessary to emphazise that the quotation of Smith's results of parameterized flux here is not to compare the present scheme to hers or to justify either of them,

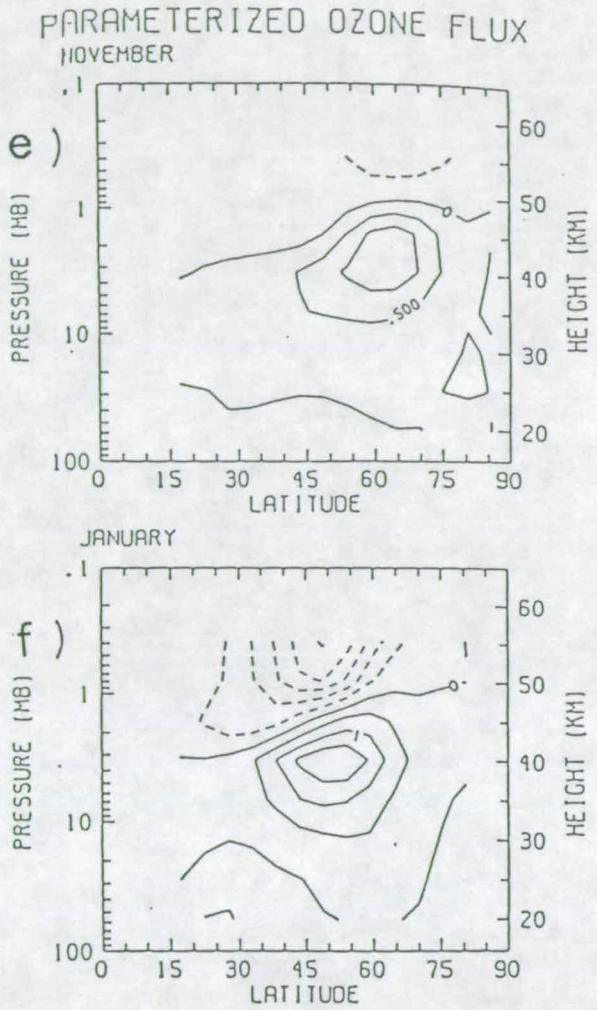


Figure 6.6: Meridional chemical eddy ozone transport calculated from the parameterization $\overline{v'\mu'} = K_c \cdot \nabla\bar{\mu} + K_T \cdot (T/\theta)\nabla\bar{\theta}$, where K_c and K_T are transport coefficients caused by chemical sources and sinks and diabatic heating respectively. (e) For November, (f) January. Unit: ppmv (ms^{-1}). [From Smith 1988]

since the two schemes are based on completely different premises, in which Smith considered mainly the effects of chemical eddies (sources and sinks), whereas the present scheme concentrates on the effects of dispersion only assuming tracers are conservative and waves are adiabatic. The purpose of discussing the present result in conjunction with hers' is to show that the dispersive part of the transport may give better parameterization in some regions of the stratosphere where transport due to dispersion may dominate. The good parameterization may therefore presumably be the one which combines the two schemes. The result here therefore can be thought of as supplementary to those of Smith's in the understanding of transport processes.

It needs to be mentioned that equation (6.68), from which the above results are obtained, is also based on the condition of small wave amplitudes. As already discussed in chapter 5, it can not be used in the periods of large wave amplitudes such as those periods of stratospheric warmings. However the present estimation is concentrated on monthly mean quantities and emperically equation (6.68) can be used to deduce K_{yy}. Moreover the above discussions show that the approximation to dispersive coefficients by wave transience in the form of equation (6.70) is very good, at least in this case. This approximation does not need the knowledge of parcel displacement and is able to give different coefficients for different tracers.

Nitric acid HNO_3 has a fairly long chemical lifetime (several days near 30km

and longer below, Brasseur and Solomon, 1984) and is removed by transport into the troposphere where it dissolves in water droplets and is rained out. It can be considered as quasi-conservative as ozone on time scales of the order of a week or less (Butchart and Remsberg 1986) and its transport coefficients can also be estimated in the same way as ozone. Figures 6.7 to 6.9 give the results equivalent to those of figures 6.4 to 6.6 but HNO_3 is substituted for ozone in the calculation. Figure 6.8 shows that the areas of positive K_{yy} are dominating. It is different from the K_{yy} for ozone. Figure 6.10a,b shows the observed real flux of HNO_3 produced by Smith from LIMS data. Comparison of figure 6.9 with figure 6.10 indicates that though the parameterized magnitude is similar to that of the real flux, the distributions are very poor. The November parameterized flux is opposite in sign and distribution to the real flux and in January the parameterized positive flux is too high in position and the area of negative flux too weak. The parameterized negative area is not on the top of the positive area as in the real flux. Since the lifetime scale of nitric acid decreases quickly with height above 30km (Brasseur and Solomon 1984) and there are uncertainties about its chemistry in high latitudes (Smith 1988), the assumptions of adiabatic and conservative conditions in the present estimation may be not fit and cause the large errors in the parameterized flux. Nitric acid, therefore, may be not an ideal species for transport studies.

It is found that the estimation of the transport coefficients are sensitive to the way

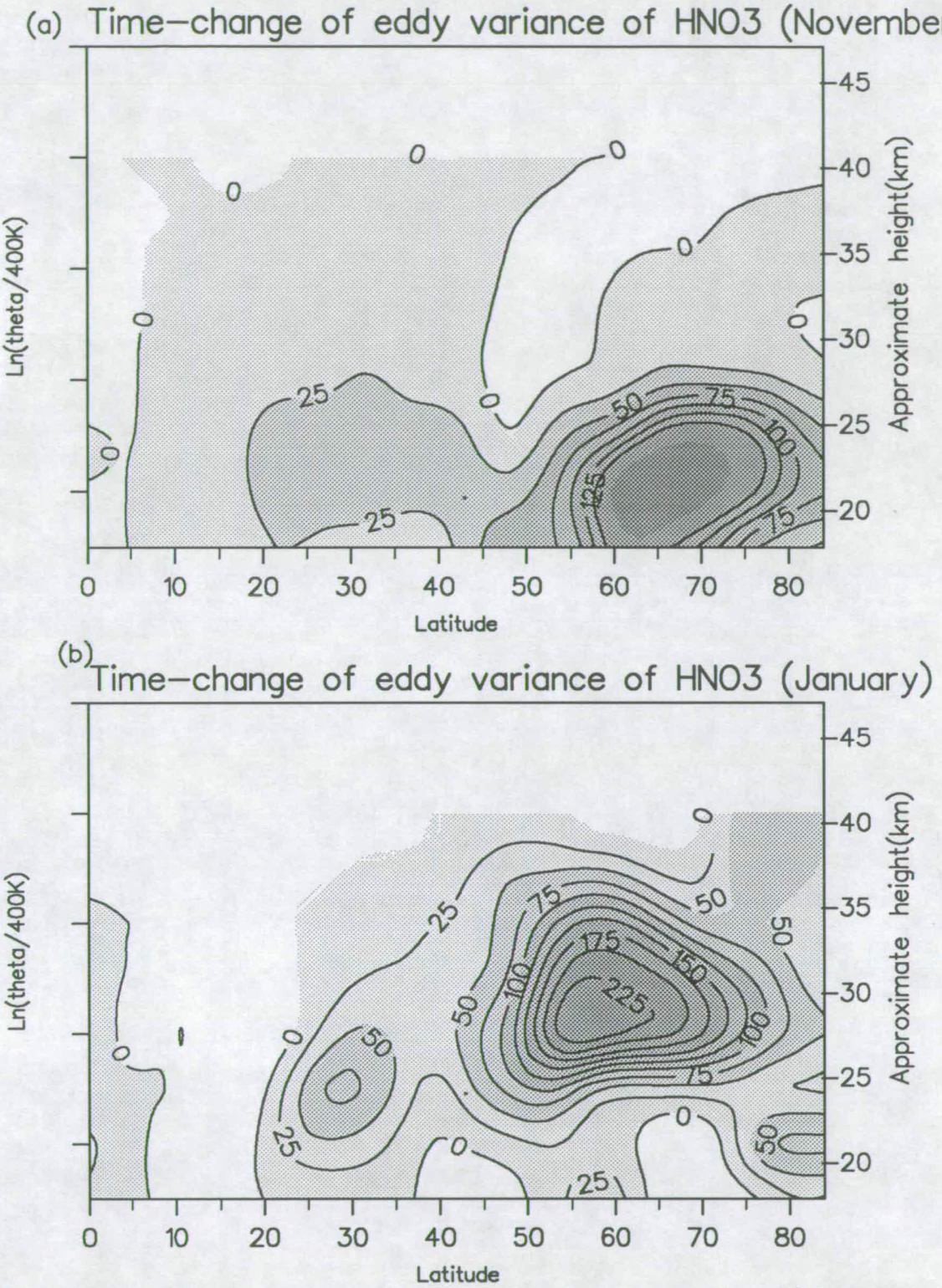


Figure 6.7 Difference of eddy variance of HNO3 mixing ratio. (a) for November 1978, (b) for January 1979. Unit: 1.0e-14 ppv**2. (see text for details).

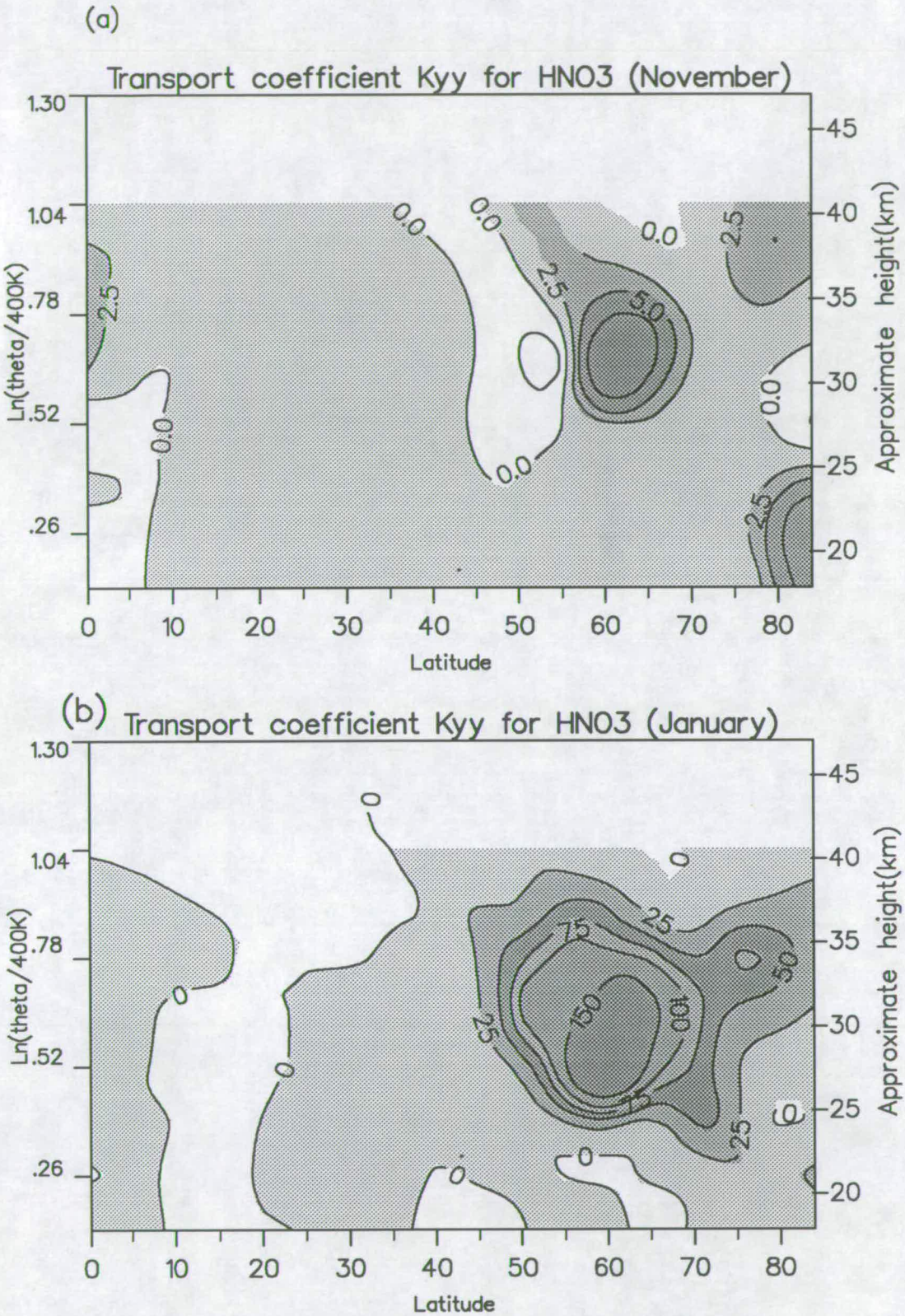


Figure 6.8 Transport coefficients for HNO3. As figure6.5 but ozone is replaced by HNO3 as chemical tracer. Unit: 1.0e5 m**2/s.

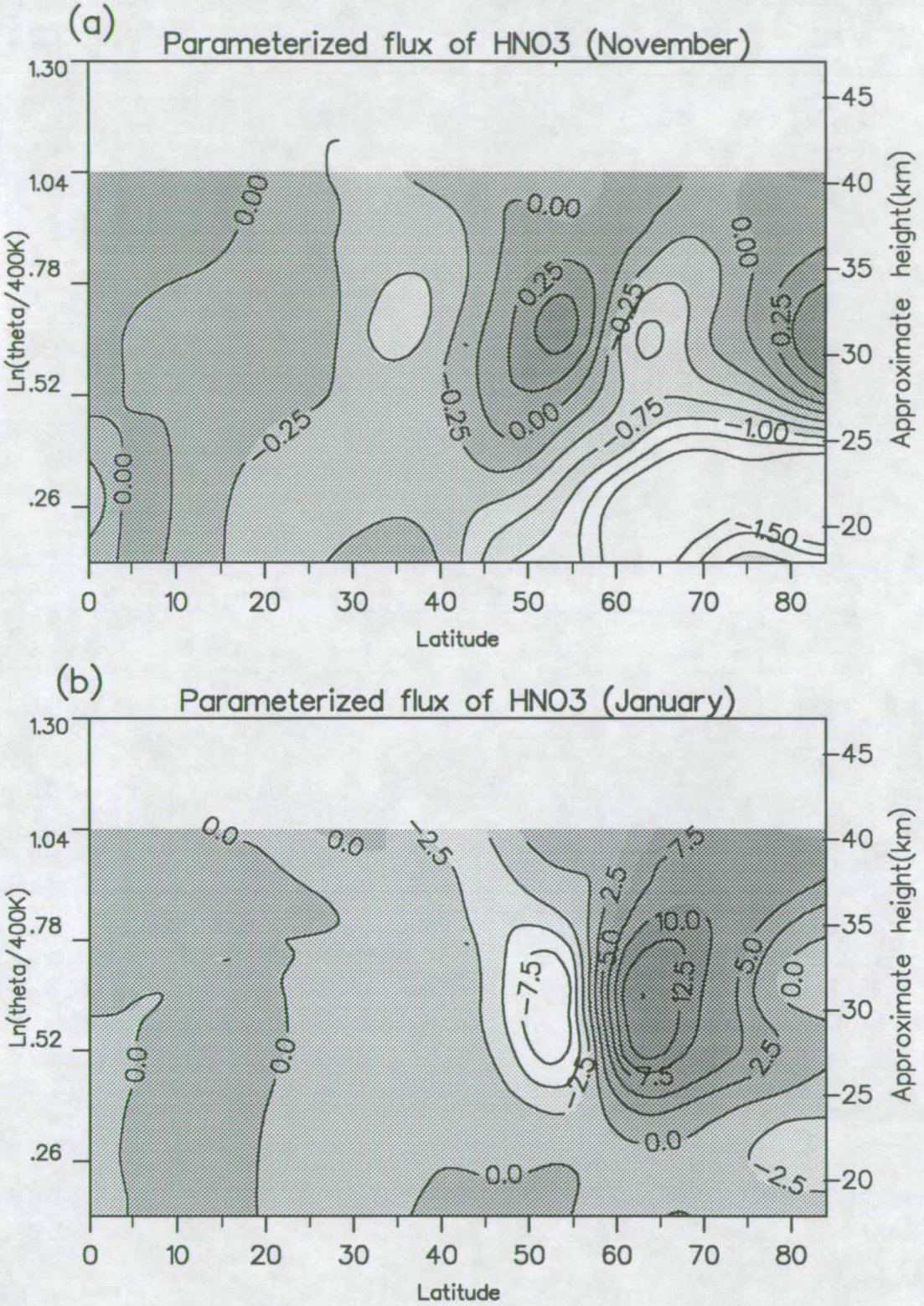


Figure 6.9 Parameterized wave flux of HNO3 using coefficients in figure 6.8 (a) for November 1978, (b) for January 1979. Unit: ppmv (m/s).

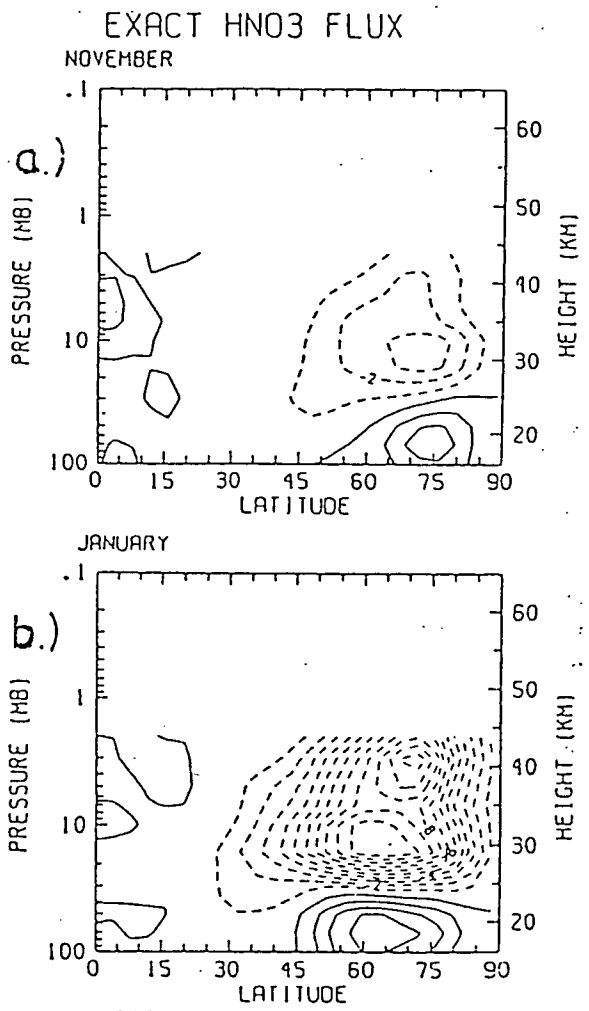


Figure 6.10: Observed meridional eddy HNO3 transport, calculated from observed waves for (a) November, (b) January. Units are ppmv (m/s). Contour interval is 1. [From Smith, 1988]

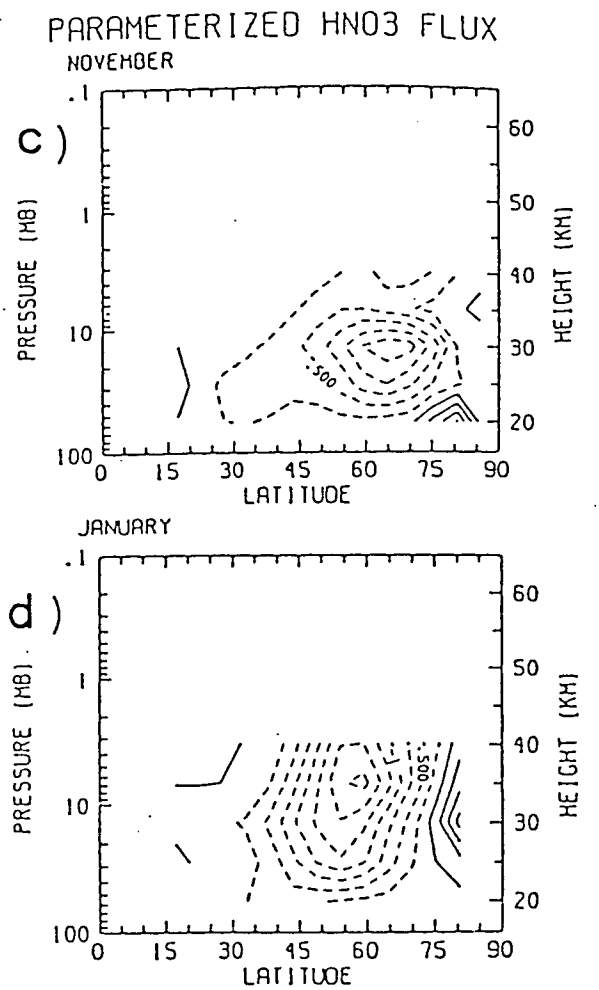


Figure 6.10: Meridional chemical eddy HNO3 transport calculated from the parameterization $\overline{v'\mu'} = K_c \cdot \nabla\bar{\mu} + K_T \cdot (T/\theta)\nabla\bar{\theta}$, where K_c and K_T are transport coefficients caused by chemical sources and sinks and diabatic heating respectively. (c) For November, (d) January. Unit: ppmv (ms^{-1}). Contour interval is 0.25. [From Smith 1988]

in which the monthly mean gradients of tracer mixing ratio are derived. In particular, using gradients obtained from fields of monthly-mean mixing ratio is not quite the same in estimating the transport coefficients and the parameterized flux as using gradients obtained from monthly-average of the daily gradients. There is also the possibility that the discrepancies between our parameterized flux and the real flux may arise from inadequacies in the small amplitude assumption in the theory itself. This inadequacy has already been discussed in chapter 5.

In summary, the derived expression and the related tracer wave transience are analogous to the generalized E–P flux theorem and the related quantities. Applying them to approximate and derive the dispersive transport coefficients and to parameterize tracerwave flux showed good agreements with the real flux when ozone mixing ratio is used. The method can reflect the importance of dispersion in late winter at lower levels and the magnitude is very similar to the real flux. Since only the dispersion effects approximated by tracer wave transience are considered here, there are several problems in the transport coefficients obtained. The main one is the existence of large areas of negative coefficients, for example in November and January. Theoretically the negative coefficients are allowed because our calculation of K_{yy} is purely due to dispersion and under the assumption of linearity or small amplitude. In these conditions, the diffusion coefficients defined like $K_{yy} = \frac{1}{2}(\overline{\eta'^2})_t$ represent reversible dispersion of fluid parcels: when parcels

disperse, that is the parcels travel away from their equilibrium latitude, then $\overline{\eta'^2}$ increases with time, $K_{yy} \geq 0$; otherwise when parcels return to their equilibrium latitude, $\overline{\eta'^2}$ decreases with time leading to $K_{yy} \leq 0$. Therefore it is hard to judge the obtained K_{yy} alone by their signs. However, practically, in the two dimensional model the negative K_{yy} often causes problems for convergence of calculations and much more care is needed to treat it (Tung 1989). This hinders the application of the above deduced K_{yy} to model experiments.

Another problem is the discrepancy between the parameterized flux and the real flux as discussed above, especially for the flux of nitric acid. This may be caused by our neglect of nonlinearity and of nonconservative terms in the equation (6.68) and by the adiabatic assumption. The chemical eddies, according to the work of Smith(1988), may play an important role in the contribution to transports and in reality our calculation on a monthly mean basis is far longer than the chemical life time and the real atmosphere is surely affected by diabatic heating in such a period.

Another potential problem is that the use of observational data like LIMS itself may be dubious. As indicated by Yang and Tung (1989) and several others (e.g. Rosefield et al.,1987; Kiehl and Solomon,1986; Shine,1989), there are large imbalances in the calculated net radiative heating rates due to the errors in the observed data for ozone and possibly for temperature. Such errors may induce large uncertainties in our previous calculation of such highly derived quantities.

We believe that a better data set, for instance ~~those from a~~ model produced ozone like that used by Yang and Tung (1989) where the ozone and temperature field are compatible and mass conserved, and the inclusion of chemical wave effects in the calculation of K_{yy} by equation (6.67) (or even K^s by 6.61), may produce realistic and applicable transport coefficients, provided the linear theory is adequate in most periods of tracer transport, and the simple flux-gradient relation is able to characterize the transport. The next chapter will indicate some possible extensions and improvements of the present study for future investigations.

Chapter 7

Summary and Concluding Remarks

This thesis has presented a diagnostic study of some aspects of the stratospheric dynamics by applying recently developed ideas and theory to satellite derived data. The scope includes the estimation of various methods used to diagnose winds from satellite derived height fields, the investigation of the influence of divergence of wind on the polar vortex and wave breaking processes, the wave activity and wave-mean-flow interactions from the view point of the generalized E–P flux theorem and the chemical wave properties and their implications for tracer transport.

In the following paragraphs the main conclusions from this thesis will be summarized and potential further extensions of the work indicated.

It has been recognized recently that the geostrophic wind is insufficient in

accuracy for stratospheric dynamical study. In order to diagnose more accurate winds from satellite derived height fields, an approximation scheme named "quasi-geostrophic wind", was introduced and compared with some commonly used methods, i.e. geostrophic, quasi-gradient, linear and nonlinear balance. The winds diagnosed by each scheme from either the satellite observed data or model-output height fields were compared with the corresponding 'true' winds and the errors from each scheme were analysed. The geostrophic approximation has been found to give considerable errors in winds in cyclonically curvating areas and in the eddy momentum flux derived from it. The quasi-geostrophic winds have the smallest error of all the schemes tested and as a result have been adopted in the other studies of this thesis. The Rossby number is confirmed to be crucial for all the approximations. When Rossby number is small, all approximations give better winds than they may give in the large Rossby number case. When Rossby number is large every approximation has considerable errors. A possible further extension to improve the wind approximation in the large Rossby number case and at high altitude is to incorporate drag effects in the zonal momentum equation, as applied by Marks (1989) to the climatological monthly mean data.

The influence on the polar vortex of the effects of the divergence of winds was investigated in chapter 4 of this thesis. It has recently been highlighted and accepted that the area enclosed by potential vorticity contours on isentropic

surfaces be used as an 'index' to understand the detailed dynamics involved in wave-breaking and erosion of the polar vortex. With the effect of wind divergence on the area neglected, Butchart and Remsburg (1986) concluded that the area change is a result of nonconservative processes and/or irreversible mixing to unresolvable scales. The investigation here shows that in the ordinary condition of the stratospheric winter circulation with no occurrence of a warming, their conclusion is true, with wind divergence playing a negligible role. However before the appearance of stratospheric warmings the divergence of the winds contributes significantly to the change of the area and thus, together with nonconservative processes and/or mixing to unresolvable scales, has non-negligible effects on the polar vortex. In the future studies of vortex erosion, when stratospheric warmings are involved, the influence on the polar vortex of the divergence of winds should be taken into account.

By exploiting the newly developed version (Andrews 1987) of the generalized Eliassen-Palm flux theorem in isentropic coordinates, the dynamical behaviour of the stratosphere was diagnosed using satellite derived data. It was found that the version of the theorem based on small amplitude waves is applicable to the diagnostic study of stratospheric dynamics when the waves in the real atmosphere are indeed small. The wave activity density, an important concept in the theory and the theorem, when so calculated can reflect the true wave behaviour in the

real atmosphere and its variations are reflected in the changes of zonal mean flow in a good qualitative agreement with the theorem. The wave activity density and wave transience, E–P flux divergence, together with wave dissipation show coherent time and spatial distributions. The balance among the terms in the theorem is only approximately reached in the lower stratosphere and this indicates the importance of nonlinear effects neglected in the linear premise. The wave transience, wave drive and nonconservative term have similar importance in the balance. The balance is improved when the radiative dissipation term is included. Since the effects of drag by processes such as gravity waves and friction are not considered both in diagnosing the winds and in the analysis of the balance, one of the extensions of this work in the future could be the inclusion of such processes in both estimation of winds and analysis of the balance by means of parameterization as proposed by Marks (1989) (also see Lindzen 1981; Holton and Wehrbein 1980).

As present application of the theorem is carried out only in the Northern Hemisphere, an obvious question is: how applicable is the theorem in the other hemisphere? Since the circumpolar zonal circulation in the winter Southern Hemisphere is much stronger and the amplitude of the planetary waves there is smaller than its northern counterpart, due to the lack of orographic forcing from below, is the theory based on the small amplitude condition more suitable for diagnostic study there? Such a study certainly has the potential to give a comparison of the

dynamics of waves in the two hemispheres.

Moreover as the version of the theorem used in the present investigation is strictly based on the small amplitude of waves, there are practical difficulties in applying it to look into the large wave amplitude cases as in the periods of stratospheric warmings. Other future work could be the use of those finite amplitude versions of the theorem (as given by Haynes, 1988) to diagnostic studies in much disturbed periods.

In chapter 6, based on the same ideas as are used in deriving the generalized E–P flux theorem, an expression analogous to the theorem but for chemical tracers was derived theoretically, without additional assumptions, and a corresponding wave activity defined by chemical tracers such as ozone was presented. This analogous expression shows the relation between the tracer transport and the wave transience; after some simplifications this gives an essentially new method of estimating the transport coefficients. In the middle atmosphere good agreement was found between wave activity defined in terms of ozone mixing ratio and that defined in terms of potential vorticity and this reflects some relation between the dynamical and trace gas processes. The tracer-dependent transport coefficients resulting from dispersion of fluid parcels alone were estimated by tracer wave transience and the parameterized fluxes compared with the real fluxes. It was found that the parameterized flux from the present study can reflect the large influences

of planetary waves on the tracer transport in the lower stratosphere in late winter. On the other hand, there are some discrepancies between the parameterized flux and the real flux, especially for nitric acid, and there is practical difficulty in the direct use of the derived transport coefficients in the parameterization of transport effects in model studies. These are due to the appearance of large areasof negative values in the monthly mean cofficients. Since the estimation of the transport coefficients is on the monthly mean base which is longer than the tracer chemical life time, the effects of chemical eddies (sources and sinks) and the effects of diabatic heating may be the main reason for the discrepancies. The errors could also be caused by the limit of linear theory, or by the errors in the data or by calculating schemes that may hinder our estimation of wave transience. In future work, by properly considering the effects of chemical eddies and diabatic heating as well as using data of good quality, the analogue expression for chemical tracers could be used to understand further the relations between chemical and dynamical tracers and processes, and to derive a set of satisfactory tracer-dependent transport coefficients. We hope to see such work appear in the near future.

Appendix A

Derivation of equation 6.59

Define that

$$F \cdot \nabla \bar{\chi} \equiv \overline{\chi' v'} \bar{\chi}_y + \overline{\chi' Q'} \bar{\chi}_\theta \tag{A.71}$$

Replacing χ' in above by equation 6.56 gives:

$$\begin{aligned}
\overline{\chi' v'} \bar{\chi}_y + \overline{\chi' Q'} \bar{\chi}_\theta &= \bar{\chi}_y(-\overline{v'\eta'} \bar{\chi}_y - \overline{v'q'} \bar{\chi}_\theta + \overline{v'\gamma'}) \\
&\quad + \bar{\chi}_\theta(-\overline{\eta' Q'} \bar{\chi}_y - \overline{q' Q'} \bar{\chi}_\theta + \overline{\gamma' Q'}) \\
&= \bar{\chi}_y[-\bar{D}(\frac{1}{2}\overline{\eta'^2}) \bar{\chi}_y - \frac{1}{2}\bar{D}\overline{\eta' q'} \bar{\chi}_\theta - \frac{1}{2}\bar{D}\overline{\eta' q'} \bar{\chi}_\theta + \overline{v'\gamma'}] \\
&\quad + \bar{\chi}_\theta[-\frac{1}{2}\overline{\eta' \bar{D} q'} \bar{\chi}_y - \bar{D}(\frac{1}{2}\overline{q'^2}) \bar{\chi}_\theta - \frac{1}{2}\overline{\eta' \bar{D} q'} \bar{\chi}_y + \overline{Q'\gamma'}] \\
&= \bar{\chi}_y[-\bar{D}(\frac{1}{2}\overline{\eta'^2}) \bar{\chi}_y - \frac{1}{2}(\overline{\eta' q'})_t \bar{\chi}_\theta + \frac{1}{2}\bar{D}\overline{q'\eta'} \bar{\chi}_\theta - \frac{1}{2}\bar{D}\overline{\eta' q'} \bar{\chi}_\theta + \overline{v'\gamma'}] \\
&\quad + \bar{\chi}_\theta[-\frac{1}{2}(\overline{\eta' q'})_t \bar{\chi}_y + \frac{1}{2}\overline{q' \bar{D} \eta'} \bar{\chi}_y - \bar{D}(\frac{1}{2}\overline{q'^2}) \bar{\chi}_\theta - \frac{1}{2}\overline{\eta' \bar{D} q'} \bar{\chi}_y + \overline{Q'\gamma'}] \\
&= \bar{\chi}_y[-K_{yy} \bar{\chi}_y - K_{y\theta} \bar{\chi}_\theta + \frac{1}{2}(\overline{\eta' Q'}) \bar{\chi}_\theta - \frac{1}{2}(\overline{v' q'}) \bar{\chi}_\theta + \overline{v'\gamma'}]
\end{aligned}$$

$$
\begin{aligned}
& +\bar{\chi}_\theta[-K_{\theta y}\bar{\chi}_y - K_{\theta\theta}\bar{\chi}_\theta + \frac{1}{2}(\overline{v'q'})\bar{\chi}_y - \frac{1}{2}(\overline{\eta'Q'})\bar{\chi}_y + \overline{\gamma'Q'}] \\
= \; & \bar{\chi}_y[-K_{yy}\bar{\chi}_y - K_{y\theta}\bar{\chi}_\theta - \frac{1}{2}(\overline{v'q'} - \overline{\eta'Q'})\bar{\chi}_\theta + \overline{v'\gamma'}] \\
& +\bar{\chi}_\theta[-K_{\theta y}\bar{\chi}_y - K_{\theta\theta}\bar{\chi}_\theta + \frac{1}{2}(\overline{v'q'} - \overline{\eta'Q'})\bar{\chi}_y + \overline{\gamma'Q'}] \\
= \; & -\bar{\chi}_y(K_{yy}\bar{\chi}_y + K_{y\theta}\bar{\chi}_\theta) - \bar{\chi}_\theta(K_{\theta y}\bar{\chi}_y + K_{\theta\theta}\bar{\chi}_\theta) + \overline{v'\gamma'}\bar{\chi}_y + \overline{\gamma'Q'}\bar{\chi}_\theta \\
= \; & -\nabla\bar{\chi}\cdot\left(\underline{K}^{(s)}\cdot\nabla\bar{\chi}\right) + \overline{v'\gamma'}\bar{\chi}_y + \overline{\gamma'Q'}\bar{\chi}_\theta
\end{aligned}
$$

Equation 6.60 is taking into account in the above derivation.

Equation 6.59 is thus verified.

Bibliography

Al-Ajmi D. N., Harwood R. S. and Miles T. 1985. A sudden warming in the middle atmosphere of the southern hemisphere. *Quart. J. R. Meteor. Soc. Vol. 111 pp 359-389*

Andrews D. G. 1989 Some comparison between the middle atmosphere dynamics of the southern and northern hemisphere. *PAGEOPH Vol. 130 pp 213-232*

Andrews D. G. 1987 On the interpretation of the Eliassen-Palm flux divergence. *Quart. J. Roy. Meteor. Soc. Vol. 113 pp 323-338*

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

Andrews D. G. 1983. A finite-amplitude Eliassen-Palm theorem in isentropic coordinates. *J. Atmos. Sci. Vol. 40 pp 1877-1883*

Andrews D. G. 1985. Wave,mean-flow interaction in the middle atmosphere. *Adv. Geophys. Vol. 28A pp 249-275*

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

Andrews D. G. and McIntyre M. E. 1976b. Planetary waves in horizontal and

vertical shear: asymptotic theory for equatorial waves in weak shear. *J. Atmos. Sci. Vol. 33 pp 2049-2053*

Andrews D. G. and McIntyre M. E. 1978a. Generalized Eliassen-Palm and Charney-Drazin theorems for waves on axisymmetric mean flows in compressible atmosphere. *J. Atmos. Sci. Vol. 35 pp 175-185*

Andrews D. G. and McIntyre M. E. 1978b An exact theory for nonlinear waves on a Lagrangian-mean flow. *J. Fluid Mech. Vol 89 pp 609-646*

Andrews D. G., Mahlman J. D. and Sinclair R. W. 1983. Eliassen-Palm diagnostics of wave-mean flow interaction in the GFDL "SKYHI" general circulation model. *J. Atmos. Sci. Vol. 40 pp 2768-2784*

Anne K. Smith, Lyjak V. L. and Gille J. C. 1988 The eddy transport of nonconserved trace species derived from satellite data. *J. Geophys. Res. Vol. 93 pp 11103-11122*

Austin J. and Tuck A. F. 1985 The caculation of stratospheric air parcel trajectories using satellite data. *Quart. J. Roy. Meteor. Soc. Vol. 111 pp 279-307*

Barnett J. J. and Corney 1985a. Middle atmosphere reference model derived from satellite data. *Handbook MAP Vol. 16*

Barnett J. J. and Corney M. 1985b. Planetary waves. *Handbook MAP Vol. 16*

Boville B. A. 1987. The Validity of the geostrophic approximation in the winter

stratosphere and troposphere. *J. Atmos. Sci. Vol. 44 pp 443-457*

Boyd J. P. 1976 The noninteraction of waves with the zonally averaged flow on a spherical earth and the interrelationships of eddy fluxes of energy, heat and momentum. *J. Atmos. Sci. Vol. 33 pp 2285-2291*

Brasseur G. and Solomon S. 1984 *Aeronomy of the Middle Atmosphere* Reidel, Dordrecht, Netherland.

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

Charney J. G. and Drazin 1961 Propogation of planetary-scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res. Vol. 66 pp 83-109*

Clough, S. A., Grahame and O'Neill A. 1985 Potential vorticity in the stratosphere derived using data from satellites.*Quart. J. Roy. Meteor. Soc. Vol. 111 pp 335-358*

Dickinson R. E. 1968a On the exact and approximate linear theory of vertically propagating planetary Rossby waves forced at a sperical lower boundary. *Mon. Weath. Rev. Vol. 96 pp 405-415*

Dickinson R. E. 1968b Planetary Rossby waves propagating vertically through weak westerly wind wave guides. *J. Atmos. Sci. Vol. 25 pp 984-*

1002

Dickinson R. E. 1969a Vertical propagation of planetary Rossby waves through an atmosphere with Newtonian cooling. *J. Geophys. Res. Vol. 74 pp 929-938*

Dickinson R. E. 1969 Theory of Planetary wave-zonal flow interaction. *J. Atmos. Sci. Vol. 26 pp 73-81*

Dunkerton T. J. 1978 On the mean meridional mass motions of the stratosphere and mesosphere. *J. Atmos. Sci. Vol. 35 pp 2325-2333*

Dunkerton T. J. 1981a On the inertial stability of the equatorial middle atmosphere. *J. Atmos. Sci. Vol. 38 pp 2354-2364*

Dunkerton T, J, 1981 Wave transience in a compressible atmosphere. Part 2: Transient equatorial waves in the quasi-biennial oscillation. *J. Atmos. Sci. Vol. 38 pp 298-307*

Dunkerton T. J. and Delisi 1986 Evolution of potential vorticity in the winter stratosphere of January-February 1979 em J. Geophys. Res. Vol. 91 pp 1199-1208

Elson L. S. 1986. Agostrophic motion in the stratosphere from satellite observations. *J. Atmos. Sci. Vol. 43 pp 409-418*

Eliassen A. and Palm E. 1961. On the transfer of energy in stationary mountain waves. *Geofys. Publ. Vol. 22(3) pp 1-23*

Ertel H. 1942. Ein neuer hydrodynamischer Wirbelsatz. *Meterol. Z. Vol. 59 pp 271-281*

Edmon H. J. Jr., Hoskins B. J. and McIntyre M. E. 1980 Eliassen-Palm cross section for the troposphere. *J. Atmos. Sci. Vol 37 pp 2600-2616*

Fisher, M 1987 The Met.O.20 Stratosphere-Mesosphere Model. *Meteorological Office, Met.O.20, DCTN 52*

Garcia R. R. and Hartmann D. L. 1980 The role of planetary waves in the maintenance of the zonally averaged ozone distribution of the upper stratosphere. *J. Atmos. Sci. Vol. 37 pp 2248-2264*

Garcia R.R. and Solomon S. 1983 A numerical model of the zonally averaged dynamical and chemical structure of the middle atmosphere *J. Geophys. Res. Vol 88C pp 1379-1400*

Garcia R. R. 1987 On the mean meridional circulation of the middle atmosphere. *J. Atmos. Sci. Vol. 44 pp 3599-3609*

Gille J. C., Russell M. *III*, Bailey P. L.,Gordley L. L., Remsberg E. E., Lienesch J. H.,Planet W. G.,House F. B., Lyjak L. V. and Beck S. A. 1984. Validation of temperature retrievals obtained by the limb infrared monitor of the stratosphere (LIMS) experiment on Nimbus 7. *J. Geophys. Res. Vol. 89 pp 5147-5160*

Gillie J.C. and Lyjak L.V. 1986 Radiative heating and cooling rates in the middle

atmosphere. *J. Atmos. Sci. Vol. 43 pp 2215-2229*

Grose W. L. and Russell M. *III* 1985 The use of isentropic potential vorticity in conjunction with quasi-conserved species in the study of stratospheric dynamics and transport.

Haigh J. D. 1984. Radiative heating of the lower stratosphere and the distribution of ozone in two dimensional model. *Quart. J. Roy. Meteor. Soc. Vol. 110 pp 167-185*

Haltiner G. J. and Williams R. T. 1980 *Numerical Prediction and Dynamical Meteorology.* Second ed. John Wiley and Sons,Inc. pp53-69

Hartmann D. L., Mechoso C. R. and Yamazaki K. 1984 Observations of wave,mean-flow interaction in the southern hemisphere. *J. Atmos. Sci. Vol. 41 pp 351-362*

Hartmann D. L. 1976 The dynamical climatology of the stratosphere in the southern hemisphere during late winter 1973. *J. Atmos. Sci. Vol. 33 pp 1789-1802*

Haynes P. H. and McIntyre M. E. 1987 On the Evolution of vorticity and potential vorticity in the presence of diabatic heating and frictional or other forces. *J. Atmos. Sci. Vol. 44 pp 829-841*

Haynes P. H. 1988 Forced, dissipative generalizations of finite-amplityude wave-avtivity conservation relations for zonal and nonzonal basic flows.*J.*

Atmos. Sci. Vol. 45 pp 2352-2362

Hitchman M. H. and Brasseur 1988 Rossby wave activity in a 2-D model: Closure for wave driving and meridional eddy diffusivity. *J. Geophys. Res. Vol. pp*

Holton J. R. 1986 Meridional Distribution of stratospheric trace constituents.*J. Atmos. Sci. Vol. 43 pp 1238-1242*

Holton J. R. 1986 A dynamically based transport parameterization for one-dimensional photochemical models of the stratosphere. *J. Geo. Res. Vol. 91 D2 pp 1681-2686*

Holton J. R. 1981 An advective model for two-dimensional transport of stratospheric trace species. *J. Geophys. Res. Vol. 86 pp11989-11994*

Holton J. R. 1980 The dynamics of sudden stratospheric warmings. *Annu. Rev. Earth Planet. Sci. Vol. 8 pp 169-190*

Holton J.R. and Wehrbein 1980 A numerical model of the zonal mean circulation of the middle atmosphere. *Pure Appl. Geophys., Vol. 118 pp 284-306*

Holton J. R. 1979 *An Introduction to Dynamic Meteorology. Academic Press,Inc.* Second ed. pp94–97 and pp173-181

Hoskins B. J., McIntyre M. E. and Robertson A. W. 1985 On the use and significance of isentropic potential vorticity maps. *Quart. J. R. Meteor.*

Soc. Vol. 111 pp 877-946

Hsu C.-P. F. 1981 A numerical study of the role of wave-wave interactions during sudden stratospheric warmings. *J. Atmos. Sci. Vol. 38 pp 189-214*

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

Juckes M. N. and O'Neill A. 1988 Early winter in the northern stratosphere. *Quart. J. Roy. Soc. Vol. 114 pp 1111-1125*

Karoly D. J. and Hoskins B. J. 1982 Three dimensional propagation of planetary waves. *J. Meteo. Soc. Japan Vol. 60 pp 109-122*

Kida H. 1983 General circulation of air parcels and transport characteristics derived from a hemispheric GCM. Part 2: Very long-term motions of air parcels in the troposphere and stratosphere. *J. Meteor. Soc. Japan. Vol. 61 pp 510-523*

Kiehl J. T. and Solomon S. 1986 On the radiative balance of the stratosphere. *J. Atmos. Sci. Vol. 43 pp 1525-1534*

Killworth P. D. and McIntyre M. E. 1985 Do Rossby-wave critical layers absorb,reflect or over-reflect?. *J. Fluid Mech. Vol. 161 pp 449-492*

Kinnersley J. 1989 Personal communication.

Labitzke K.,Barnett J. J. and Edwards B. 1985 Atmospheric structure and its

variation in the region 20 to 120 km. Draft of a new reference middle atmosphere. *Handbook for map Vol. 16 SCOSTEP Univ. of Illinois, Urbana-Champaign*

Leovy C. B., Sun C-R. Hitchman M. H., Remsberg E. E., Russell *III*, Gordley L. L., Gille J.C. and Lyjak L. V. 1985 Transport of ozone in the middle stratosphere:evidence for planetary wave breaking *J. Atmos. Sci. Vol. 42 pp230-244*

Lin B-D 1982 The behaviour of winter stationary planetary waves forced by topography and diabatic heating *J. Atmos. Sci. Vol. 39 pp 1206-1226*

Lindzen R. S. 1981 Turbulence and stress owing to gravity wave and tidal breakdown. *J. Geophys Res., Vol. 86 pp 9707-9714*

Lyjak L. V. 1987 Diffusion coefficients calculated from satellite data, *Transport Process in the Middle Atmosphere*, edited by G. Visconti and R. Garcia, pp 343-352 D.Reidel, Hingham,Mass.

Lyjak L. V. and Smith A. K. 1987 Lagrangian mean circulations in the stratosphere. *J. Atmos. Sci. Vol. 44 pp 2252-2266*

Marks C.J. 1989 Some Features of the Climatology of the Middle Atmosphere Revealed by Nimbus 5 and 6*Jour. Atm. Sci. Vol. 46 pp2485-2508*

Marks C. J. 1987 (See Page 219)

Matsuno T. 1970 Vertical propogation of stationary planetary waves in the winter northern hemisphere. *J. Atmos. Sci. Vol. 27 pp 871-883*

Matsuno T. 1971 A dynamical model of the stratospheric sudden warming. *J. Atmos. Sci. Vol. 28 pp 1479-1494*

Matsuno T. 1980 Lagrangian motion of air parcels in the stratosphere in the presence of planetary waves. *Pure Appl. Geophys. Vol. 118 pp189-216*

Matsuno T. 1984 Dynamics of minor stratospheric warmings and preconditioning. 'Dynamics of the Middle Atmosphere', *eds. Holton J. R. and Matsuno T pp 333-351. Terrapub, Tokyo*

McIntyre M. E. 1987 Dynamics and tracer transport in the middle atmosphere: an overview of some recent development. 'Transport processes in the middle atmosphere'. *NATO ASI Series ed. G. Visconti and R. Garcia*

McIntyre M. E. and Shepherd T. G. 1987 An exact local conservation theorem for finite amplitude disturbances to non-parallel shear flows,with remarks on Hamiltonian structure and on Arnol'd's stability theorems. *J. Fluid Mech. Vol. 181 pp 527-565*

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

McIntyre M. E. and Palmer T. N. 1984 The 'surf zone' in the stratosphere *J. Atmos. Terr. Phys. Vol. 46 pp 825-849*

McIntyre M. E. 1982 How well do we understand the dynamics of stratospheric warmings?. *J. Meter. Soc. Japan Vol. 60 pp 37-65*

McIntyre M. E. 1980a An introduction to the generalized Lagrangian-mean description of wave, mean-flow interaction. *Pure Appl. Geophys Vol. 118 pp 152-176*

McIntyre M. E. 1980b Toward a Lagrangian-mean description of stratospheric circulations and chemical transports. philos. Trans. R. Soc. London Ser. A296 pp129-148

McIntyre M. E. 1981 On the wave momentum myth. *J. Fluid Mech. Vol. 106 pp 331-347*

Murgatroyd R. J. 1982 Recent progress in studies of the stratosphere *Quart. J. R. Meter. Soc. Vol. 108 pp 271-312*

Newman P. A., Schoeberl M. R. and Plumb R. A. 1986 Horizontal mixing coefficients for two-dimensional chemical models calculated from National Meteorological Society data. *J. Geophys. Res. Vol. 91 pp 7919-7924*

O'Neill A. and Pope V. D. 1988 Simulations of linear and nonlinear disturbances in the stratosphere. *Quart. J. R. Meteor. Soc. Vol. 114 pp 1063-1110*

O'Neill A. and Youngblut C. E. 1982. Stratospheric warmings diagnosed using the transformed Eulerian-mean equations and the effect of the mean

state on wave propagation. *J. Atmos. Sci. Vol. 39 pp 1370-1386*

Palmer T. N. and Hsu C.-P. F. 1983 Stratospheric sudden collings and the role of nonlinear wave interactions in preconditioning the circumpolar flow. *J. Atmos. Sci. Vol. 40 pp 909-928*

Palmer T. N. 1982 Properties of the Eliassen-Palm flux for Planetary scale motions. *J. Atmos. Sci. Vol. 39 pp 992-997*

Palmer T. N. 1981 Diagnostic study of a wavenumber-2 stratospheric sudden warming in a transformed Eulerian-mean formalism. *J. atmos. Sci. Vol. 38 pp 844-855*

Palmer T. N. 1981b aspects of stratospheric sudden warmings studied from a transformed Eulerian-mean viewpoint. *J. Geophys. Res. Vol. 86 pp 9679-9687*

Pawson S. and Harwood R.S. 1989 (See page 219).

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

Plumb R. A. 1985 On the three-dimensional propagation of stationary waves. *J. Atmos. Sci. Vol. 42 PP 217-229*

Plumb R. A. and Mahlman J. D. 1987 The zonally Averaged Transport characteristics of the GFDL general circulation/transport model. *J. Atmos. Sci. Vol. 44 pp 298-327*

Pyle J. A. and Rogers 1980 Stratospheric transport by stationary planetary waves

- the importance of chemical processes. *Quart. J. Roy. Meter. Soc. Vol. 106 pp421-446*

1987
Randel W. J. The Evaluation of winds from geopotential height data in the stratosphere. *J. Atmos. Sci. Vol. 44 pp 3097-3120*

Reed R. J. and German 1965 A contribution to the problem of stratospheric diffusion by large-scale mixing. *Mon. Weather Rev. Vol. 93 pp 313-321*

Robinson W. A. 1986. The application of the quasi-geostrophic Eliassen-Palm flux to the analysis of stratospheric data. *J. Atmos. Sci. Vol. 1986 pp 1017-1023*

Rogers C. F. and Pyle J. A. 1984 Stratospheric tracer transport: A modified diabatic circulation model. *Quart. J. Roy. Meteor. Soc. Vol. 110 pp 219-237*

Rood R. B. and Schoeberl M. R. 1983 A mechanistic model of Eulerian, Lagrangian mean, and Lagrangian ozone transport by steady planetary waves. *J. Geophys. Res. Vol. 88 pp 5208-5218*

Rosenfield J. E., Schoeberl M. R. and Geller M. A. 1987 A computation of the stratosphere diabatic circulation using an accurate radiative transfer model. *J. Atmos. Sci. Vol. 44 pp 859-876*

Scherhag R. 1952 Die explosionsartige stratospharenerwarmung des Spatwinters

1951/1952. *Ber. Deut. wetterdieustes Vol. 6 pp 51-63*

Schoeberl M. R. and Geller M. A. 1977. A calculation of the structure of stationary planetary waves in winter. *J. Atmos. Sci. Vol. 34 pp1235-1255*

Shine K. 1989 Sources and sinks of zonal momentum in the middle atmosphere diagnosed using the diabatic circulation. *Quart. J. Roy. Meteor. Soc. Vol. 115 pp 265-292*

Shine K. The middle atmosphere in the absence of dynamical heat fluxes. *Quart. J. Roy. Meteor. Soc. Vol. 113 pp 603-633* World Meteorological Organization 1986 Global ozone research and monitoring project Report no. 16 Vol. 114 pp 241-343

Simmons A. J. 1974 Planetary scale disturbances in the polar winter stratosphere. *Quart. J. R. Met. Soc. Vol. 79 pp 342-366*

Smith, A. K. 1988 (see Anne K. Smith et al. in page 206).

Strobel D. F. 1978 Parameterisation of the atmospheric heating rate from 15-120km due O_2 and O_3 absorption of solar radiation. *J. Geophys. Res. Vol. 83 pp 6225-6230*

Taylor G. I. 1915 Eddy motion in the atmosphere. *Phil. Trans. Roy. Soc. London, A215, pp1-26*

Tung K.K. 1984 Modelling of Tracer Transport in the Middle Atmosphere. *Dynamics of the Middle Atmosphere. Ed J.R.Holton and T.Matsuno. D.Reidel Publishing Company.*

Tung K.K. 1986 Nongeostrophic Theory of Zonally Averaged Circulation. Part 1: Formulation *Jour. Atm. Sci. Vol. 43 pp2600-2618*

World Meteorological Organization 1986 Global ozone research and monitoring project Report no. 16

Yang H., Tung K. K. and Olaguer 1989 Nongeostrophic Theory of zonally averaged circulation Part *II*:Eliassen-Palm flux divergence and isentropic mixing coefficient. *J. Atmos. Sci.*

Marks, C. J. 1987: The seasonal evolution of the middle atmosphere as revealed by PMR/SCR data. Memorandum 87.3, Dept. Atmospheric, Occeanic and planetary physics, Clarendon Laboratory, Oxford. U. K., 47 PP.

Pawon S. and Harwood R. S. 1989 Monthly-mean diabatic circulations in the stratosphere. Quart. J. Roy. Meteorol. Soc., Vol. 115, PP807-840.