In Defense of Ertel's Potential Vorticity and Its General Applicability as a Meteorological Tracer

EDWIN F. DANIELSEN

NASA Ames Research Center, Moffett Field, California

(Manuscript received 14 November 1988, in final form 22 January 1990)

ABSTRACT

A potential vorticity theorem and its two summary statements published by Haynes and McIntyre are challenged conceptually by equations, discussions and examples. The apparent simplification proposed by the authors to convert from a mass to volume integral, i.e., by cancelling density against the specific volume in the potential vorticity, changes the physical significance of the integrand. It no longer is the potential vorticity. The resulting mean for either a bulk Eulerian or Lagrangian system is then not analogous to a mixing ratio and therefore not independent of the broad spectrum of internal waves, the independence that makes Ertel's potential vorticity so valuable either as a stratospheric tracer or as a predictive or diagnostic, large scale, meteorological variable.

1. Introduction

In a recent publication, Haynes and McIntyre (1987) summarized a theorem they developed, in two extraordinary statements:

(i) There can be no net transport of Rossby–Ertel potential vorticity (PV) across any isentropic surface; and

(ii) PV can neither be created nor destroyed, within a layer bounded by two isentropic surfaces.

If (i) were correct, Ertel's potential vorticity could not function as a scalar meteorological tracer, analogous to and correlated with the mixing ratio of atmospheric trace constituents, except for isentropic transports. When Reed (1955) and Reed and Danielsen (1959) used Ertel's potential vorticity to identify a stratospheric origin for the air in the upper portion of active cold fronts, they certainly emphasized adiabatic transports. In the spring of 1960, however, an aircraft field experiment, including B57 and U2 aircraft, provided in situ measurements of specific radioactive isotopes to test its more general applicability as a tracer including diabatic in addition to adiabatic transports.

When these data were analyzed by Danielsen et al. (1962), including correlations of potential vorticity with each isotope (representing upper-polar and midtropical stratospheric origin as well as a tropical tropospheric origin) and distributions of each isotope plotted against potential temperature as the ordinate and potential vorticity as the abscissa, they recognized and stressed the importance of diabatic transports to stratospheric–tropospheric exchange. Clearly, the diabatic transports were not destroying the correlations with these isotopes of diverse origin.

Unlike trace chemical constituents, subject to photochemical sources and sinks, radioactive isotopes when corrected for their radioactive decay are unambiguously conserved. Therefore, if the air cools diabatically, the air and the isotopes move across the isentropic surfaces to lower values of θ with the isotopes maintaining a constant mixing ratio. Then, if the correlations with potential vorticity are qualitatively maintained, the potential vorticity must move diabatically with the air at approximately the same rate. Indeed, later it will be proved that a bulk system will move at the same mean rate when the mean gradient of diabatic cooling is zero, or, if it is not zero, when it is orthogonal to the absolute vorticity vector. Thus, with the aid of an example, (i) will be shown to be incorrect for a bulk system, just as it obviously is incorrect for an infinitesimal system.

If (ii) were correct, the conclusion stated by Haynes and McIntyre (1987) would logically follow; i.e., "Potential vorticity can be created or destroyed at those places, if any, where the layer [bounded by two isentropic surfaces] terminates laterally." Thus (ii) would essentially eliminate a stratospheric source for potential vorticity. The strong positive correlations with radioactivity and ozone of stratospheric origin (Danielsen 1968), and equally strong negative correlations with carbon monoxide and water vapor of tropospheric origin (Danielsen et al. 1987), would remain as major, unexplained, mysterious coincidences.

The reasons for the above conclusions are quite simple. Potential temperature surfaces in the stratosphere

Corresponding author address: Dr. Edwin F. Danielsen, 70093 Meadow View Road, Sisters, OR 97759.

range from 300 ± 10 K to 2000 ± 250 K. Of these surfaces only those at θ's < 320 K can terminate laterally; the overwhelming majority span the entire globe and consequently would have no potential vorticity source. Also, those that do intersect the earth's surface function statistically as a sink for potential vorticity, due to the negative vertical gradient of diabatic heating during the day. The positive gradient during the night, that generates potential vorticity in surface inversions, is hardly a viable source for the stratosphere.

Where then, and by what physical processes, are the observed stratospheric values of potential vorticity generated? Again, the answer is quite simple. The diabatic heating that generates the large static stability and gives the stratosphere its identity also generates the stratospheric values of potential vorticity. Of course, the elevated heat source simultaneously decreases the static stability and the potential vorticity in the air above it, in the mesosphere. Therefore, in a bulk system encompassing both the stratosphere and mesosphere the internal source and sink of potential vorticity resulting from the elevated heat source would tend to cancel.

But the integral of Haynes and McIntyre (1987) applies to any isentropic layers of finite dimensions, including one located in the stratosphere below the level of maximum mean heating or in the mesosphere above the level where the internal source or sink is only of one sign. Consequently, there is a conceptual error in their interpretation of the integrated form of Ertel's potential vorticity theorem.

2. Ertel's potential vorticity theorem

In this case the conceptual difficulties are related to the contribution of the mass continuity equation to Ertel's potential vorticity. To clarify its contribution, it is constructive to examine the component equations whose sum reduces to Ertel's theorem and to identify those terms that cancel. Following Staley's (1960) vectorial derivation, but omitting a frictional force for reasons to be presented below, Ertel's theorem is the sum of three scalar equations:

$$(\mathbf{q}_a \cdot \nabla\theta)\frac{d\alpha}{dt} = (\mathbf{q}_a \cdot \nabla\theta)\alpha\nabla \cdot \mathbf{V} \qquad (2.1)$$

$$\alpha\mathbf{q}_a \cdot \frac{d\nabla\theta}{dt} = \alpha\mathbf{q}_a \cdot \nabla\frac{d\theta}{dt} - \underline{\alpha\mathbf{q}_a \cdot \nabla(\mathbf{V} \cdot \nabla\theta)} \qquad (2.2)$$

$$\alpha\nabla\theta \cdot \frac{d\mathbf{q}_a}{dt} = -\alpha(\nabla\theta \cdot \mathbf{q}_a)\nabla \cdot \mathbf{V} + \underline{\alpha\mathbf{q}_a \cdot \nabla(\mathbf{V} \cdot \nabla\theta)} \qquad (2.3)$$

The first (2.1) is the scalar $(\mathbf{q}_a \cdot \nabla\theta)$ multiplying the mass continuity equation. Here $\mathbf{q}_a$ is the absolute vorticity $(\nabla \times \mathbf{V} + 2\mathbf{\Omega})$ where ∇ is the three-dimensional gradient operator, $\mathbf{V}$ is the three-dimensional velocity of an infinitesimal air parcel, $\mathbf{\Omega}$ is the rotation vector of the earth, θ is the potential temperature and α is the specific volume of the air. The second (2.2) is derived from the gradient of the energy equation by interchanging the gradient and total derivative operators. This interchange includes ∇ operating on the scalar product $\mathbf{V} \cdot \nabla\theta$. Since the operation on $\nabla\theta$ has been incorporated in the total derivative on the lhs, $\underline{\nabla\theta}$ is held constant in the remaining term on the rhs. Hence the meaning of the underline. The third (2.3) is the scalar product of $\alpha\nabla\theta$ and the vorticity equation in which the baroclinic term $\nabla\theta \cdot \nabla\alpha \times \nabla p$ has been eliminated because it reduces to zero.

The sum is Ertel's potential vorticity theorem:

$$\frac{d}{dt}(\alpha\mathbf{q}_a \cdot \nabla\theta) = \alpha\mathbf{q}_a \cdot \nabla\frac{d\theta}{dt} \qquad (2.4)$$

or

$$\frac{dQ}{dt} = \alpha\mathbf{q}_a \cdot \nabla\frac{d\theta}{dt} \qquad (2.5)$$

as derived from Euler's equations of motion, with molecular heat conduction implicit but not explicitly described in the energy equation. If the more general equations are used, including the Navier–Stokes equations, there would be an additional term expressing molecular diffusion of vorticity. In the free atmosphere, however, it redistributes but does not create vorticity. The latter requires a torque and the molecular stress tensor, being symmetric, cannot exert a torque. For our purposes (2.5) is essentially complete. When we integrate over a bulk system free from a physical boundary and introduce the eddy or deviatory flux of potential vorticity across the conceptual boundaries of the bulk system we can neglect the effects of molecular diffusion. Any torque exerted on the bulk system by the environment as a result of deviatory mixing will be included in the surface integral and depend on the distribution of potential vorticity.

It is clear from (2.1) and (2.3) that potential vorticity is independent of three-dimensional divergence. Therefore potential vorticity is independent of sound waves and, more importantly, of adiabatic expansions and compressions associated with large vertical displacements. But as (2.5) indicates, conservation of potential vorticity is not limited to adiabatic processes. Nonconservation requires a gradient of diabatic heating colinear with the absolute vorticity vector. This extension toward conservation is especially significant in middle and lower stratosphere at high latitudes where radiative cooling and descent can be sustained as air passes over extensive, cold-cirrus cloud shields. With weak gradients of radiative cooling the potential vorticity is quasi-conserved, increasing its usefulness as a stratospheric tracer.

Finally, potential vorticity is independent also of a broad spectrum of transverse waves, ranging from relatively high frequency gravity to very low frequency baroclinic waves that generate and/or modulate extratropical weather. As these transverse waves propagate, the atmosphere undergoes an oscillating three-dimensional deformation. A cylindrical fluid element, bounded above and below by surfaces of constant θ, oscillates in shape from oblate to prolate to oblate, etc. As it expands horizontally, along the θ surfaces, it contracts vertically, moving the θ surfaces closer together.

That these changes in shape are compatible with no change in the elements' potential vorticity can be seen most clearly by manipulating the basic equations somewhat differently than in (2.1) to (2.3). Instead of taking the gradient of the energy equation we take only its vertical derivative. Then using the partial derivative transformation from x, y, z to x, y, θ coordinates we can introduce $\nabla_H \cdot \mathbf{V}_\theta$, the horizontal divergence at constant θ, and eliminate $\nabla \cdot \mathbf{V}$ with the continuity equation to yield

$$\frac{d}{dt}\left[\ln\left(\alpha \frac{\partial\theta}{\partial z}\right)\right] = \nabla_H \cdot \mathbf{V}_\theta + \frac{\partial}{\partial\theta}\left(\frac{d\theta}{dt}\right)_\theta. \quad (2.6)$$

Assuming hydrostatic balance we obtain a static stability equation, with g the acceleration of gravity,

$$\frac{d}{dt}\left[\ln\left(-g \frac{\partial\theta}{\partial p}\right)\right] = \nabla_H \cdot \mathbf{V}_\theta + \frac{\partial}{\partial\theta}\left(\frac{d\theta}{dt}\right)_\theta. \quad (2.7)$$

Consistent with the hydrostatic assumption we can derive a corresponding vorticity equation from the horizontal equations of motion expressed in x, y, θ coordinates

$$\frac{d}{dt}(\ln[\zeta_\theta + f]) = -\nabla_H \cdot \mathbf{V}_\theta + [\zeta_\theta + f]^{-1} \frac{\partial \mathbf{V}}{\partial\theta} \times \nabla_H \left(\frac{d\theta}{dt}\right)_\theta \cdot \mathbf{k}. \quad (2.8)$$

Adding (2.7) and (2.8) eliminates the horizontal divergence at constant θ and returns us to the usual isentropic approximation to (2.5), with

$$Q \approx -g \frac{\partial\theta}{\partial p}[\zeta_\theta + f] \quad (2.9)$$

where $\zeta_\theta + f = (\nabla_H \times \mathbf{V}_\theta + 2\mathbf{\Omega}) \cdot \mathbf{k}$, and $\mathbf{k}$ is a vertical unit vector. The approximation is denoted because small terms involving the vorticity of vertical motion and the horizontal component of the earth's vorticity have been neglected.

In this isentropic form we can recognize that the potential vorticity is essentially the product of two stabilities derived from particle dynamics, i.e., static stability and inertial stability. Furthermore, we see from (2.7) and (2.8) that as one increases the other decreases. When the cylinder is oblate, static stability is a maximum and inertial stability (isentropic vorticity) is a minimum. For a low frequency wave these conditions are obtained in the ridge. If we follow the isentropic trajectory of this elemental cylinder as it advances from ridge to trough, isentropic convergence will change it to prolate with static stability decreasing to a minimum while inertial stability (isentropic vorticity) increases to a maximum.

The propagating transverse waves' isentropic convergence appears to convert static stability to vorticity. In this sense the former functions as a potential for the latter. As (2.8) clearly shows, however, for a given convergence, vorticity amplification depends on the vorticity but not on static stability. Nevertheless, static stability does play a critical "potential" role.

As isentropic convergence decreases the static stability, the probability increases that a small vertical shear can initiate vertical mixing. Thus as evident by (2.7), increasing the static stability by a vertical gradient of diabatic heating, increases the potential for vorticity amplification without a turbulent breakdown and vertical mixing. Indeed, we use the average static stability, i.e., its resistance to vertical mixing, to distinguish stratosphere from troposphere.

Applying the same reasoning to the divergent phase of the wave cycle, we can distinguish the cyclonic from the anticyclonic stratosphere by its greater resistance to horizontal, isentropic, mixing. According to (2.8) absolute vorticity at constant θ can be increased by a horizontal gradient of diabatic heating. In the lower stratosphere where $\partial U/\partial\theta$ is negative, the combined effects of heating at low latitudes and cooling at high latitudes will rotate horizontal vorticity into the vertical and, at the same time, smaller values of potential vorticity will be advected upward as a result of the heating while larger values are moved diabatically downward by the cooling.

Therefore, the gradients of potential vorticity, both vertical and horizontal components, are directly related to gradients of diabatic heating and to transport processes connecting source to sink. Most significantly for diagnostic and trace chemical analyses, potential vorticity is a unique meteorological scalar analogous to χ_i, the mixing ratio of trace constituent i. Unlike θ and like χ_i, surfaces of equal Q can fold due to velocity shears, many being wave induced, without the folding initiating instability and mixing. This process that produces a laminar structure in Q and χ_i is essential to maintaining the correlations between $\bar{Q}$ and $\bar{\chi}_i$ gradients and their corresponding deviations from the local means.

3. Relevance of analogy to mixing ratio

The analogy to χ_i is evident in the conservation equations: in total derivative, elemental form

$$\frac{dQ}{dt} = \alpha\mathbf{q}_a \cdot \nabla \frac{d\theta}{dt} = S_Q \qquad \frac{d\chi_i}{dt} = S_i \quad (3.1)$$

and in partial derivative, flux form

$$\frac{\partial}{\partial t}(\rho Q) + \nabla\cdot\rho Q\mathbf{V} = \mathbf{q}_a\cdot\nabla\frac{d\theta}{dt} = \rho S_Q$$

$$\frac{\partial}{\partial t}(\rho\chi_i) + \nabla\cdot\rho\chi_i\mathbf{V} = \rho S_i. \quad (3.2)$$

As stressed by Haynes and McIntyre (1987), since $\nabla\cdot\mathbf{q}_a = 0$ the upper left equation can be written also as

$$\frac{\partial}{\partial t}(\rho Q) + \nabla\cdot\left(\rho Q\mathbf{V} - \mathbf{q}_a\frac{d\theta}{dt}\right) = 0. \quad (3.3)$$

In the above equations S_i is the source per unit mass of constituent i. Note that in the flux form, the product ρQ is not potential vorticity. Just as $\rho\chi_i = C_i$, the concentration, is not independent of compression and expansion, the product ρQ also is not independent.

In fact, (2.1) and (2.3) have been included here to demonstrate that the three-dimensional divergence term in the vorticity equation would remain uncancelled if the continuity equation is not incorporated in the derivation of potential vorticity as recommended in section 4 of Haynes and McIntyre (1987). Finally, (2.6)–(2.8) have been included to demonstrate a major, conceptual pitfall when x, y, θ coordinates are used. Quoting from Haynes and McIntyre (1987), "That is, (2.4) is the form required if we wish to discuss potential vorticity as a tracer and to quantify its transport from place to place. . . ."

Their (2.4) is the x, y, θ approximation to our (3.3) with ρ and ∇ replaced by $\sigma \equiv -g^{-1}(\partial p/\partial\theta)$ and ∇_H, respectively. But the product $\sigma Q = \zeta_\theta + f$. Indeed, their (2.4) is then identical to our (2.8), i.e., the vorticity equation in isentropic coordinates. Clearly, we do not want to use the wave dependent vorticity in tracer analyses.

This discussion brings us to the crux of the problem. To use potential vorticity in atmospheric analyses we have no choice but to integrate over appropriately large bulk systems. Whether we use it as a tracer in Lagrangian studies or as a correlated meteorological scaler in Eulerian studies, we must define the bulk system so that $\bar{Q}$, the bulk mean value, preserves the analogy to $\bar{\chi}_i$.

A conceptually simple method, applicable to both Lagrangian and Eulerian integrals, is discussed in section 4, along with specific examples of both types of integrals. The method will lead us, directly, to include changes in $\bar{Q}$ and $\bar{\chi}_i$ due to small-scale mixing, a conservative process, in addition to changes produced by sources and sinks, nonconservative processes.

4. Conservation equations for moving bulk systems

A concise and convenient formula for computing the total derivative of a spatial integral is the generalized Leibnitz formula for differentiating a triple integral (Bird et al. 1960). It is repeated here for the convenience of the reader:

$$\frac{D}{Dt}\iiint_V f(\mathbf{r},t)dV = \iiint_V \frac{\partial f}{\partial t}(\mathbf{r},t)dV + \iint_A f(\mathbf{r},t)\mathbf{V}_B(\mathbf{r},t)\cdot d\mathbf{A}. \quad (4.1)$$

The total derivative on the lhs is written with upper case D's to emphasize that it refers to a bulk system and to distinguish it from the elemental derivative, written here in the customary lower case. It is arbitrary, of course, until the velocities $\mathbf{V}_B(\mathbf{r}, t)$ of the system's boundary are defined. For example, a rigid, completely open system, moving through the fluid with a velocity $\mathbf{V}_S = \mathbf{V}_B$ permits the linear expansion $D/Dt = \partial/\partial t + \mathbf{V}_S\cdot\nabla$. As a special case, when $\mathbf{V}_S = 0$, the Eulerian integral is a stationary, open system. At the opposite extreme one can define a closed system by setting $\mathbf{V}_B = \mathbf{V}(\mathbf{r}, t)$. This system, closed to mass exchange, would maintain a constant mass and if the velocities are density weighted it would move with the velocity of the center of mass, i.e.,

$$\mathbf{V}_S = \mathbf{V}_{CM} = \frac{\iiint_V \rho\mathbf{V}dV}{\iiint_V \rho dV}. \quad (4.2)$$

Although conceptually simple and attractive, a closed system is physically unrealistic for atmospheric applications. The necessary density of observations and degree of resolution are not available and, if they were, the system would deform and distort much too rapidly.

A rational alternative proposed by Danielsen (1968) and used by Danielsen and Hipskind (1980) and Danielsen (1981) is to define a local mean, as a continuous variable, by averaging over a mass M large enough to be compatible with the observed velocities, thus

$$\bar{\mathbf{V}}(\mathbf{r},t) = \frac{\iiint_V \rho\mathbf{V}dV}{M} \quad (4.3)$$

and to set $\mathbf{V}_B(\mathbf{r}, t) = \bar{\mathbf{V}}(\mathbf{r}, t)$. As will be shown, this system also maintains a constant mass and moves with the velocity of the center of mass but it is open to exchanges with its environment by the smaller scale velocity deviations $\mathbf{V}'$. It also can deform due to gradients of $\bar{\mathbf{V}}(\mathbf{r}, t)$ but more slowly and less drastically than the closed system.

Substituting $f = \rho$ in (4.1), replacing the partial derivative by the mass divergence and using Gauss's theorem to convert this term from a volume to a surface integral yields

$$\frac{DM}{Dt} = \iint_A \rho(\mathbf{V}_B - \mathbf{V}) \cdot d\mathbf{A} \quad (4.4)$$

setting $\mathbf{V}_B = \bar{\mathbf{V}}$, as stated above, and expressing $\mathbf{V} = \bar{\mathbf{V}} + \mathbf{V}'$ (4.4) reduces to

$$\frac{DM}{Dt} = -\iint_A \rho\mathbf{V}' \cdot d\mathbf{A} = -\iiint_V \nabla \cdot \rho\mathbf{V}' dV$$
$$= -\nabla \cdot \iiint \rho\mathbf{V}' dV = 0 \quad (4.5)$$

because the mass weighted mean of $\mathbf{V}'$ is zero by definition. This bulk system, closed to all scales of motion larger than that defining $\bar{\mathbf{V}}$ is open to exchanges of momentum, energy, potential temperature, potential vorticity, etc., by the smaller scale deviatory motions. Therefore, it is a bulk or large scale generalization of the elemental, material system, often called a Lagrangian system, which is open to exchanges by the molecular motions.

To determine the diabatic transports we set $f = \rho\theta$ in (4.1). Equating the total (material) derivative of θ to its partial derivatives, multiplying this equation by ρ and adding the continuity equation multiplied by θ yields

$$\frac{\partial}{\partial t}(\rho\theta) + \nabla \cdot (\rho\theta\mathbf{V}) = \rho \frac{d\theta}{dt}. \quad (4.6)$$

After substituting (4.6) into (4.1) and converting the divergence term from a volume to a surface integral, it can be written as

$$M \frac{D\bar{\theta}}{Dt} + \bar{\theta} \frac{DM}{dt}$$
$$= -\iint_A \rho\theta'\mathbf{V}' \cdot d\mathbf{A} + \iiint_V \frac{d\theta}{dt} \rho dV \quad (4.7)$$

and reduced to (noting that $DM/Dt = 0$)

$$\frac{D\bar{\theta}}{Dt} = \overline{\frac{d\theta}{dt}} - \iint_A \frac{\rho\theta'\mathbf{V}' \cdot d\mathbf{A}}{M}. \quad (4.8)$$

With (4.8) we can separate the change in $\bar{\theta}$ due to diabatic and adiabatic effects. The change due to diabatic heating or cooling is the mass-weighted mean of the elemental, material change. The change due to adiabatic mixing is, of course, evaluated statistically at the boundary where the smaller scale deviatory motions can exchange θ with its environment. Here it is assumed that the mixing is predominantly adiabatic even when the deviatory velocities are initiated by nonuniform diabatic heating. From (4.8) it should be obvious that the air can move diabatically relative to the θ surfaces. Now, can the potential vorticity move with the air when the air moves diabatically?

To answer this question we substitute $f = \rho Q$ into (4.1). Starting with Ertel's theorem in total derivative form (3.1) we convert first to the partial derivative form of (3.2). The integration yields

$$M \frac{D\bar{Q}}{Dt} + \bar{Q} \frac{DM}{Dt} = -\iint_A \rho Q'\mathbf{V}' \cdot d\mathbf{A}$$
$$+ \iiint_V \left(\alpha \mathbf{q}_a \cdot \nabla \frac{d\theta}{dt}\right) \rho dV \quad (4.9)$$

which reduces to (again noting that $DM/Dt = 0$)

$$\frac{D\bar{Q}}{Dt} = \overline{\alpha \mathbf{q}_a \cdot \nabla \frac{d\theta}{dt}} - \iint_A \frac{\rho Q'\mathbf{V}' \cdot d\mathbf{A}}{M}. \quad (4.10)$$

Similarly, for the change in the mean mixing ratio of species i,

$$\frac{D\bar{\chi}_i}{Dt} = \bar{S}_i - \iint_A \frac{\rho \chi_i'\mathbf{V}' \cdot d\mathbf{A}}{M} \quad (4.11)$$

again, S_i is the source or sink of species i per unit mass.

From (4.8) and (4.10) we can readily conclude that $\bar{Q}$ will move with the air when the mass-weighted mean of the elemental source term $\overline{\alpha \mathbf{q}_a \cdot \nabla(d\theta/dt)} = 0$ despite the fact that $d\theta/dt \neq 0$. Under these conditions, closely approximated in the middle and lower-middle stratosphere where the vertical gradients of heating and cooling are small, $\bar{Q}$ is conserved and $\bar{Q}$ moves as does a conserved trace constituent with the center of mass of the moving bulk air parcel.

Also, from (4.10) and (4.11) we can appreciate the importance of the adiabatic mixing across the moving boundary of the bulk system. When $\nabla\bar{Q}$ and $\nabla\bar{\chi}_i$ are approximately colinear the small scale mixing will tend to maintain the positive or negative correlations between $\bar{Q}$ and $\bar{\chi}_i$. The systematic decrease in $\bar{Q}$ and in the mean mixing ratios of radioactive isotopes and ozone in the layer of stratospheric air extruded during tropopause folding is consistent with the adiabatic mixing expressed in (4.10) and (4.11); see Danielsen et al. (1987). A diabatic transport to colder values of θ, however, is essential to explain the frequently observed positive $\bar{Q}$ and $\bar{\chi}_{O_3}$ anomalies in the stratosphere on the cyclonic side of the jet. For an example, note Figs. 12 and 13 of Danielsen (1968). In the associated discussion, the diabatic cooling was attributed to radiational cooling over the cold cirrus cloud shields. This diabatic transport is consistent with (4.8), (4.10) and (4.11) if the vertical gradient of radiational cooling is small. Thus the observations support the theory that Ertel's potential vorticity can be transported across isentropic surfaces with the air and with the ozone mixing ratio.

Furthermore, (4.8) and (4.10) indicate that potential vorticity is being created when $\overline{\alpha \mathbf{q}_a \cdot \nabla(d\theta/dt)} > 0$ as in the upper extratropical stratosphere of the Northern Hemisphere during the spring. Then, $\bar{\mathbf{q}}_a$ is predominantly a horizontal vector pointing towards the North Pole and $\nabla(\overline{d\theta}/dt)$ is predominantly a vertical vector pointing upwards, but the angle between them is <90 deg. During the same season in the Southern Hemisphere the angle is >90 deg and the negative potential vorticity of that hemisphere's stratosphere is being created. The diabatic generation of potential vorticity in the stratosphere where the isentropic surfaces do not terminate laterally definitely contradicts statement (ii) of Haynes and McIntyre (1987) and the conclusion they logically derive from it. One might claim, however, that (4.8) and (4.10) represent an indirect proof that (i) and (ii) are incorrect, but not a direct proof, because of differences in the definition of the bulk system and the use of (3.2) rather than (3.3) in the integration.

Using (3.3) instead of (3.2) in (4.1) is a difference that makes no difference. The resulting integral

$$\frac{D\bar{Q}}{Dt} = \iint_A \frac{\frac{d\theta}{dt}\mathbf{q}_a \cdot d\mathbf{A}}{M} - \iint_A \frac{\rho Q \mathbf{V}' \cdot d\mathbf{A}}{M} \quad (4.12)$$

is different from (4.10) in appearance, but changing the mass-weighted volume integral of the diabatic heating source to a surface integral does not change the result. It merely clarifies that an internal heat source produces a self-cancelling source and sink, as discussed above with respect to the stratosphere and mesosphere. A net production or destruction depends on how the diabatic heating is distributed over the boundary surface of the bulk system.

Now it will be shown that changing the definition of the bulk system, the reference system, also will not change the result. After all, it should not, since the definition of the bulk system is arbitrary and the physical process can not depend on an arbitrary choice of a reference system. The system used by Haynes and McIntyre (1987, section 4, p. 835) is neither Eulerian nor Lagrangian. Although moving it does not follow the center of mass of a system of constant mass, therefore the change in mass is of critical importance. Also, their bulk system is rather special in that it implies a mass change solely due to diabatic processes, no explicit description of the adiabatic exchanges between the system and its environment is included. Although physically unrealistic, such a system is conceptually possible and in this case focuses attention on the diabatic effects.

Proceeding by substituting (3.3) into (4.1), the integration yields

$$\frac{D}{Dt}\iiint_V \rho Q dV = \iint_A \left[\rho Q(\mathbf{V}_B - \mathbf{V}) + \frac{d\theta}{dt}\mathbf{q}_a\right] \cdot d\mathbf{A} \quad (4.13)$$

which is the same as their (4.5) except for the differences in symbolic notation. Since their bulk system is bounded above and below by two isentropic surfaces (two constant potential temperature surfaces) the upper and lower differential area vectors can be expressed as

$$d\mathbf{A} = \pm \frac{\nabla\theta}{|\nabla\theta|} dA \quad (4.14)$$

with the plus sign for the upper and minus sign for the lower vectors. These two boundaries are assumed to move with a velocity $\mathbf{V}_\theta$. The lateral boundary is considered fixed in space, so if the top, bottom and sides are to remain connected $\mathbf{V}_\theta$ must have no component normal to the lateral boundary. They assume the same condition by choice, not by necessity. When (4.14) and $\mathbf{V}_B = \mathbf{V}_\theta$ are substituted into (4.13) the integrand for the upper and lower boundaries can be written as

$$\left[(\mathbf{q}_a \cdot \nabla\theta)(\mathbf{V}_\theta - \mathbf{V}) \cdot \nabla\theta + \frac{d\theta}{dt}\mathbf{q}_a \cdot \nabla\theta\right] \frac{\pm}{|\nabla\theta|} \quad (4.15)$$

which vanishes identically since

$$\frac{d\theta}{dt} = -(\mathbf{V}_\theta - \mathbf{V}) \cdot \nabla\theta. \quad (4.16)$$

Up to this point we agree with their statement below their Eq. (4.5), "It follows that the surface integral in Eq. (4.5) need be evaluated only over δV_s (the lateral boundary), the contribution from δV_θ (the upper and lower boundaries) vanishing identically." However, we definitely disagree with their following statement, "In other words, there is no flux of PV across isentropic surfaces."

There is a flux equal to $\rho Q(\mathbf{V} - \mathbf{V}_\theta)$ but the contribution of this flux to the integral is cancelled mathematically by the diabatic heating that produces the flux. Therefore, if the horizontal flux divergence is zero, a special but not impossible case suitable for illustrating the error in (ii), (4.13) reduces to

$$\frac{D}{Dt}\iiint_V (\rho Q) dV = 0 \quad (4.17)$$

but to preserve the analogy to a mixing ratio we convert ρdV to dM, and expand (4.17) as

$$\frac{D}{Dt}(M\bar{Q}) = \bar{Q}\frac{DM}{Dt} + M\frac{D\bar{Q}}{Dt} = 0 \quad (4.18)$$

or as

$$\frac{D\bar{Q}}{Dt} = -\frac{\bar{Q}}{M}\frac{DM}{Dt}. \quad (4.19)$$

Therefore, potential vorticity can be created or destroyed within a layer bounded by two isentropic surfaces if the mass in the same bounded system decreases or increases respectively. An example is presented in Fig. 1.

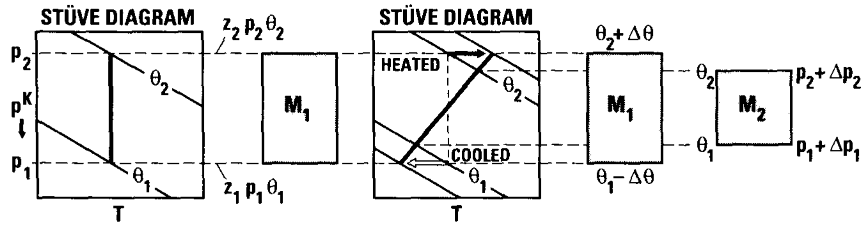

FIG. 1. Potential vorticity generation by vertical gradient of diabatic heating. M_1 remains constant, since $\bar{T}$ is constant, but $M_2 < M_1$ because θ_2 and θ_1 surfaces move relative to the air. See text.

5. Examples and summary

On the left in Fig. 1, a portion of a Stüve diagram illustrates an isothermal layer extending from z_1, p_1 and θ_1 to z_2, p_2 and θ_2. With no loss in generality we can assume that the air in this layer is stationary, in hydrostatic balance, that θ varies only in the vertical and $\mathbf{q}_a$ reduces to $2\Omega_z$, the vertical component of the earth's vorticity. If this layer is heated isobarically at the top and cooled isobarically at the bottom, increasing the static stability but not changing the mean temperature, as illustrated in the central Stüve diagram, the mass in the layer and the layer thickness remains unchanged. Also, the air remains stationary so there can be no change in the vorticity.

Therefore, if as in (4.10) we consider a bulk system of constant mass the potential vorticity has increased because the numerator has increased. Expressing $\alpha\partial\theta/\partial z$ by $-g\partial\theta/\partial p$,

$$Q\,(\text{final}) = g\,\frac{(\theta_2 - \theta_1 + 2\Delta\theta)2\Omega_2}{p_1 - p_2}$$

$$> Q\,(\text{initial}) = g\,\frac{(\theta_2 - \theta_1)2\Omega_z}{p_1 - p_2}.$$

Conversely, if as in (4.19) we consider the mass bounded by the same θ_1 and θ_2 surfaces, it has decreased because θ_2 has moved diabatically to higher pressure ($\Delta p_2 > 0$) and θ_1 has moved to lower pressure ($\Delta p_1 < 0$). Again, the potential vorticity has increased because the denominator has decreased.

$$Q\,(\text{final}) = g\,\frac{(\theta_2 - \theta_1)2\Omega_z}{p_1 - p_2 + \Delta p_1 - \Delta p_2}$$

$$> Q\,(\text{initial}) = g\,\frac{(\theta_2 - \theta_1)2\Omega_z}{p_1 - p_2}.$$

The example in Fig. 1 clearly demonstrates the generation of potential vorticity in a system of constant mass and a system bounded by two isentropic surfaces. Thus, it is sufficient to demonstrate that (ii) of Haynes and McIntyre (1987) is incorrect.

Another simple example is sufficient to demonstrate also that (i) is incorrect. In Fig. 2 a diabatic displacement is emphasized with the potential vorticity remaining constant as the bulk system, initially isothermal, is cooled isothermally such that $[(\partial/\partial z)(d\theta/dt)] = 0$. To provide a realistic example we consider a mid-stratospheric bulk system in the Antarctic vortex subject to small, but persistent, radiative cooling. Its horizontal area is equivalent to 3° of latitude squared and its depth is close to 2 km, corresponding to an initial pressure difference of 20 mb. Note that the absolute vorticity and potential vorticity are negative in the Southern Hemisphere.

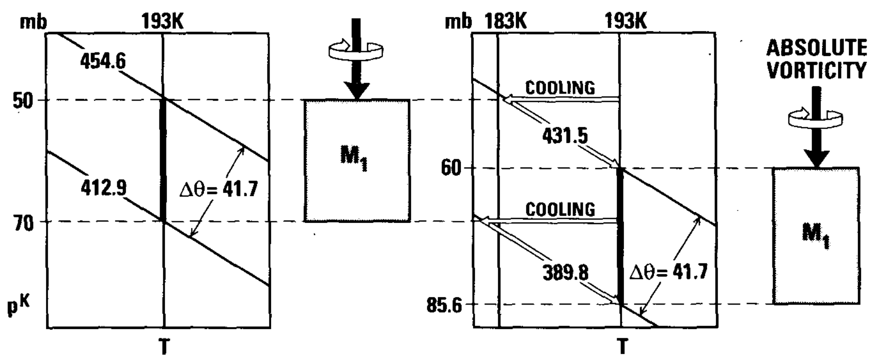

FIG. 2. Conservation of potential vorticity when bulk system cools isothermally in Southern Hemispheric, stratospheric vortex. The net change in θ also is expressed as a diabatic, isobaric cooling followed by an adiabatic compressive warming.

TABLE 1. Numerical values associated with isothermal cooling illustrated in Fig. 2 and final values, assuming same initial values, for the second case (not illustrated).

	Mass (Kg)	Volume (M3)	Area (M2)	ΔZ (M)	V/M (M3/Kg)	$\Delta\theta/\Delta z$ (deg M^{-1})	$-\mathbf{q}_a$ (1/s)	$-\bar{Q}$ (M2 deg^{-1} Kg^{-1})
	1.E13	1.E14	1.E11	1.E3	1.0	1.E-2	1.E-4	1.E-5
First case (see Fig. 2)—$\partial/\partial z(d\theta/dt) = 0$								
Initial	2.2676	2.1131	1.1111	1.9018	9.3189	2.1932	1.5000	3.0657
Final	2.2676	1.7433	0.8678	2.0089	7.6880	2.0763	1.9206	3.0657
% change	0	−19.24	−24.72	+5.48	−19.24	−5.48	+24.72	0
Second case—$\partial/\partial z(dT/dt)_p = 0$								
Final	2.2676	1.7609	0.9259	1.9018	7.7657	2.0817	1.8000	2.9098
% change	0	−18.23	−18.23	0	−18.23	−5.22	+18.23	−5.22

Two sets of final conditions are presented in Table 1. Both involve net diabatic cooling and descent, with the upper boundary of the bulk system moving to 60 mb, but they differ in that the first has a uniform vertical rate of cooling as measured by $d\theta/dt$ while the second has a uniform $(dT/dt)_p$. The latter is introduced because the uniform 23.1 K cooling in θ is equivalent to slightly less than 10 K cooling at the 50 mb level and slightly more than 10 K at the 70 mb level. Conversely, when $\partial/\partial z(dT/dt)_p = 0$, $\partial/\partial z(d\theta/dt) < 0$; there is a potential vorticity sink. Although the duration of cooling is not critical to these examples a reasonable period would be two to three weeks.

It is clear from Fig. 2 and Table 1 that there can be a net transport of potential vorticity across isentropic surfaces. In the first case the potential vorticity is conserved, in the second it decreases by a small percentage. Certainly one interesting and relevant aspect of these examples is the large increase in magnitude of absolute vorticity in the descending bulk parcel. This increase could be offset by a decrease in static stability and/or a decrease in specific volume. Here we see the importance of the latter. The bulk system is both compressed and deformed as a result of the diabatic cooling. The deformation with horizontal contraction and vertical dilation, the latter opposing the three-dimensional compression, depends on the gradient of $d\theta/dt$. It decreases as the negative gradient increases in magnitude.

In the upper polar stratosphere, below the stratopause, the diabatic cooling rates and their vertical gradients do increase significantly during late summer and fall seasons (see Kiehl and Solomon 1986). Thus upper stratospheric air descends with the magnitude of its potential vorticity decreasing as the magnitude of its absolute vorticity increases. Polar easterlies reverse to westerlies and intensify with the sustained cooling. In a recent paper, Danielsen and Houben (1988) discuss the effects of this downward diabatic transport on the distributions of trace constituents in the Antarctic vortex and examine their relationship to the ozone-hole.

With the equations and examples presented above, we conclude:

1) There can be a net transport of potential vorticity across isentropic surfaces.

2) Potential vorticity can be generated or destroyed in a bulk system bounded by two isentropic surfaces.

3) The result does not depend on the arbitrary choice of a bulk reference system.

REFERENCES

Bird, R. B., W. E. Stewart and E. N. Lightfoot, 1960: *Transport Phenomena.* John Wiley and Sons, p. 780.

Danielsen, E. F., 1968: Stratospheric–tropospheric exchange based on ozone and potential vorticity. *J. Atmos. Sci.,* **25,** 502–518.

——, 1981: An objective method for determining the generalized transport tensor for two-dimensional Eulerian models. *J. Atmos. Sci.,* **38,** 1319–1339.

——, and R. S. Hipskind, 1980: Stratospheric–tropospheric exchange at polar latitudes in summer. *J. Geophys. Res.,* **83,** 393–400.

——, and H. Houben, 1988: Dynamics of the Antarctic stratosphere and implications for the ozone-hole. Anthropogene Beeinfluss und der Ozonschicht, DECHEMA, [Theodor-Heuss-Allee 25, Postfach 97 01 46, D6000 Frankfurt am Main 97], 191–242, ISBN 3-926959-0405.

——, K. H. Bergman and C. A. Paulson, 1962: Radioisotopes, temperature and potential vorticity. University of Washington, Department of Meteorology and Climatology, Seattle, p. 54.

——, R. S. Hipskind, S. E. Gaines, G. W. Sachse, G. L. Gregory and G. F. Hill, 1987: Three-dimensional analysis of potential vorticity associated with tropopause folds and observed variations of ozone and carbon monoxide. *J. Geophys. Res.,* **92,** 2105–2111.

Haynes, P. H., and M. E. McIntyre, 1987: On the evolution of vorticity and potential vorticity in the presence of diabatic heating and frictional or other forces. *J. Atmos. Sci.,* **44,** 828–841.

Kiehl, J. T., and S. Solomon, 1986: On the radiative balance of the stratosphere. *J. Atmos. Sci.,* **43,** 1525–1534.

Reed, R. J., 1955: A study of a characteristic type of upper-level frontogensis. *J. Meteor.,* **12,** 226–237.

——, and E. F. Danielsen, 1959: Fronts in the vicinity of the tropopause. *Arch. Meteor. Geophys. Bioklim.,* **11,** 1–17.

Staley, D. O., 1960: Evaluation of potential-vorticity changes near the tropopause and the related vertical motions, vertical advection of vorticity, and transfer of radioactive debris from stratosphere to troposphere. *J. Meteor.,* **17,** 591–620.