Generalized Nonlinear Balance Criteria and Inertial Stability

JOHN A. KNOX

Department of Atmospheric and Oceanic Sciences, University of Wisconsin—Madison, Madison, Wisconsin

(Manuscript received 4 December 1995, in final form 11 September 1996)

ABSTRACT

The connections between the concept of nonlinear balance and the classical criterion of inertial stability are explored in the context of historical work on this subject. New analytic results are derived establishing that ellipticity and inertial stability are, in general, separate and distinct measures of balanced flows, even in the case of gradient flow. In particular, nonlinear balance is violated more for weaker anticyclonic flows than is inertial balance. These conclusions are supported by analysis of observational data.

A hierarchy of nonlinear balance criteria is constructed, which ranges from ellipticity to ''realizability'' conditions first obtained by Petterssen and Kasahara. Expressions interrelating all the nonlinear balance criteria and inertial stability are derived, clarifying the relationship between the Petterssen criterion, Kasahara's realizability, and inertial stability. The balance-criteria hierarchy is tested for cyclonic and anticyclonic conditions using a nonlinear inviscid *f*-plane trajectory model. The modeling results confirm the analytical ellipticity–inertial stability relationship. In addition, an intercomparison of balance criteria reveals that Petterssen's realizability (in the form derived here) is the most general and most physically interpretable balance criterion. The implications of this work for generalizations of inertial instability theory are briefly explored.

1. Introduction

The relationship between inertial stability and nonlinearly balanced flows has not been completely clarified to date. Inertial stability corresponds to a profile of angular momentum that increases outward from a vortex (Rayleigh 1916; Solberg 1933), or increases equatorward in the case of large-scale zonal-mean dynamics. This is equivalent to positive (negative) absolute vorticity in the Northern (Southern) Hemisphere. It is perhaps best defined by its opposite: inertial *in*stability, a physically observable hydrodynamic instability caused by an imbalance of pressure gradient and total centrifugal forces. This instability is often described as the horizontal analog of convective instability and is characterized by horizontal ''pancake'' circulations that cause lateral mixing until the imbalance is eliminated (Andrews et al. 1987; Hitchman et al. 1987). Balance, by contrast, is usually defined by a mathematical solvability condition governing simplified versions of the primitive equations of motion, a well-known balance criterion being ''ellipticity.'' Nonelliptic regions in a dataset violate this criterion and transform the solution of the nonlinear balance equation into a problematic attempt to solve a partial differential equation in a mixed hyperbolic–elliptic domain (Ghil et al. 1977).

For decades a connection of some sort has been noted between the concepts of nonellipticity and inertial instability in the context of understanding nonellipticity from the viewpoint of numerical weather prediction. However, the relationship of the mathematical requirement of solvability and its violations to any observable atmospheric flow regimes ''has remained a nagging question'' (Kasahara 1982). For example, Bolin (1956) observed on this subject that the ''physical meaning of the [ellipticity] criterion is not clear but it is interesting to notice that in the linearized case or in the case of a circular vortex it is similar to the criterion of inertia (or dynamic) stability.'' Miyakoda (1956) speculated that strongly nonelliptic regions would be accompanied by large changes in divergence, referring to Syō no (1948), who explicitly implicated inertial instability in the ''avalanche'' of cold air from the southern flank of the Siberian high. Charney (1962) drew a distinction between the ellipticity of the nonlinear balance equation and its numerical solvability, implying that inertial stability was a more relevant measure of convergence of the iterative solution: ''When the potential vorticity became negative, the solution immediately blew up. When [the ellipticity condition] became negative nothing happened.'' More recently, Paegle and Paegle (1974) derived a version of the balance equation for which the ellipticity condition apparently is identical to inertial stability. Tribbia (1981) also equated inertial stability with ellipticity in the case of gradient flow and moreover

Corresponding author address: Dr. John A. Knox, Center for Climate Systems Research, Columbia University and NASA/Goddard Institute for Space Studies, Armstrong Hall, 2880 Broadway, New York, NY 10025.
E-mail: jknox@giss.nasa.gov

contended that nonelliptic regions are ''potentially natural regions in which the transfer of energy between rotational and gravitational modes occurs, either through inertial instability or rotational mode breaking.'' However, research on inertial instability has rarely touched upon the topic of nonellipticity.

Árnason (1958) proposed that ''a thorough investigation of the dynamic stability of a nondivergent flow may provide a physical interpretation of the ellipticity criteria.'' Conversely, a generalized examination of balance conditions may also inform our understanding of inertial instability as a balance mechanism. The purpose of this work is to help clarify the relationship between the mathematics and the physical reality of this branch of atmospheric dynamics. Below, a hierarchy of balance criteria that includes quantities valid for the full range of the primitive equations is described, derived, and tested numerically. Several new analytical results are obtained that interrelate the balance criteria, most notably ellipticity and inertial stability. The development of this hierarchy, in turn, sheds light on the need for a more generalized approach to inertial instability as well, the latter being the subject of a separate investigation (see Knox 1996). In section 2, the relationship between ellipticity and inertial stability in idealized balanced flows is clarified. In section 3, violations of the ellipticity and inertial stability criteria in observed flows are discussed from the perspective of observational datasets. In sections 4 and 5, balance criteria more general than ellipticity, first described by Kasahara (1982) and Petterssen (1953), are derived and discussed. In section 6, relationships between the various balance criteria are examined. In section 7, an intercomparison of various generalized balance criteria is performed using a simple nonlinear model, including a discussion of how these modeling results compare with the idealized theory of section 2. The implications of this work for inertial instability research are explored in section 8, and the main results of this study and future directions of research are summarized in section 9.

2. Ellipticity and inertial stability in simple nonlinear balanced flows

a. Classical theory

1) ELLIPTICITY IN GRADIENT FLOW

The nonlinear balance equation is obtained from the horizontal divergence equation of the primitive equations of momentum on a midlatitude f- or β-plane by neglecting all frictional, diabatic, and divergent terms [Holton 1992, Eq. (11.15)]. Ellipticity arises as a mathematical constraint describing when the nonlinear balance equation can be solved as a boundary value problem. Although the resulting ellipticity criterion can be applied to meteorological data or models without reference to the underlying mathematical theory, for completeness both perspectives are provided below.

From a mathematical perspective, the constraint of ellipticity is derived using Monge–Ampère theory as follows. The nonlinear balance equation is expressed in terms of a streamfunction, and the resulting eigenvalue problem is solved as an expression involving the coefficients of the terms (Sirovich 1988, 294–299). Classical theory focuses on the second-order partial derivatives of the streamfunction; by perturbation theory these terms determine the behavior of the high-frequency modes, which in turn determines the elliptic, hyperbolic, or parabolic character of the solution (J. C. Strikwerda 1995, personal communication).

For meteorological applications on the sphere, the ellipticity condition of the nonlinear balance equation is most generally expressed as (Houghton 1968)

$$E \equiv \frac{f^2}{2} + \nabla^2\Phi + \beta u_\psi + \frac{u_\psi^2 + v_\psi^2}{a^2} > 0, \qquad (1)$$

where $f = 2\Omega \sin\phi$ is the Coriolis parameter, Φ is the geopotential, $\beta = 2\Omega \cos\phi/a$, and the subscript ψ indicates the nondivergent component of the flow. The ''metric'' terms in (1) arising from the sphericity of the earth are usually negligible; moreover, the β term is sometimes ignored, even in applications on the sphere (e.g., Randel 1987). Note that first-order derivative terms are incorporated into the ellipticity condition in addition to the second-order terms that are the focus of mathematical theory. For clarity, it is emphasized that all terms involving the horizontal divergence, friction, and vertical motion are absent from the nonlinear balance equation and therefore are ignored in deriving the ellipticity condition.

As noted, the ellipticity condition is a mathematical construct; that is, the choice of which terms appear in (1) is dictated by the application of Monge–Ampère theory to the nonlinear balance equation, rather than a scale analysis of the equation. Therefore, to obtain physical insight into (1), it is common (e.g., Kasahara 1982; Daley 1991) to examine one of the simplest cases of nonlinear balance, namely steady flow on the axisymmetric midlatitude f-plane. In this highly simplified case, the momentum equations reduce to the familiar gradient wind equation [Holton 1992, Eq. (3.10)], with regular (non-anomalous) solution in anticyclonic flow equal to

$$v_{gr} = -\frac{f_0 R}{2} - \left(\frac{f_0^2 R^2}{4} - R\frac{\partial\Phi}{\partial n}\right)^{1/2}, \qquad (2)$$

where R is the radius of curvature of trajectories and n is the direction to the left of the flow. In this special case, Daley (1991) has shown that (1) is equivalent to

$$-\frac{1}{R}\frac{\partial}{\partial n}\left(\frac{f_0^2 R^2}{4} - R\frac{\partial\Phi}{\partial n}\right) > 0, \qquad (3)$$

which is simply the radial derivative of the condition for real solutions of (2). Thus, ''The ellipticity criterion is related to the existence of a real solution to the gra-

dient wind equation. The gradient wind has no solution for the windfield in regions of intense high pressure, and neither does the nonlinear balance equation'' (Daley 1991, 233).

Investigating regular anticyclonic gradient wind solutions more closely, it is easily found that real solutions of (2) are bounded by

$$v_{gr} \leq -\frac{f_0 R}{2}. \tag{4}$$

In terms of the absolute vorticity η, this translates into

$$\eta \equiv f_0 + \zeta = f_0 + \zeta_{curv} + \zeta_{shear} \geq 0, \tag{5}$$

where ζ is the relative vorticity, $\zeta_{curv} = v/R$ is the curvature component of the relative vorticity in natural coordinates, and $\zeta_{shear} = -\partial v/\partial n$ is the shear component of the relative vorticity. Since regular cyclonic solutions of the gradient wind obviously have $\eta \gg 0$, this implies that all real regular solutions of the gradient wind, and therefore from (3), all elliptic regions, are confined to conditions of nonnegative absolute vorticity.

2) Inertial stability in regular gradient flow

The classical inertial stability criterion for a zonal geostrophic flow on the inviscid *f*-plane is (Holton 1992)

$$I \equiv f_0(f_0 + \zeta_g) > 0, \tag{6}$$

where ζ_g is the geostrophic relative vorticity. For gradient flow, Alaka (1961) derived the corresponding inertial stability criterion on an isentropic surface:

$$I \equiv \left(f_0 + 2\frac{v}{R}\right)(f_0 + \zeta) > 0. \tag{7}$$

Alaka's form is employed by Hoskins et al. [1985, Eq. (30)]. Violation of (6), or (7) for regular flow, requires negative absolute vorticity in the Northern Hemisphere or, more generally, negative potential vorticity (Hoskins 1974). Therefore, occurrences of negative absolute vorticity in statically stable flow are identified with inertial instability, although strictly speaking this connection is limited to idealized cases (Panchev 1985). Air masses are expected to maintain $I \geq 0$ through meridional or radial mixing (Andrews et al. 1987; Holton 1992). As such, (6) and (7) constitute measures of inertial balance, just as (1) is a measure of nonlinear balance.

At this point an *erroneous* connection is sometimes made between ellipticity and inertial stability. From (3), (4), and (5), it might be argued, albeit incorrectly, that since (i) the ellipticity criterion is violated where the gradient wind has no real solution and (ii) real regular solutions of the gradient wind require nonnegative absolute vorticity, then (iii) ellipticity requires nonnegative absolute vorticity, which, given (7), is equivalent to stating that for regular flow (iv) ellipticity is identical to inertial stability. However, this specious syllogism overstates the kinship between ellipticity and inertial stability. Although ellipticity requires nonnegative η, this is not equivalent to the assertion that elliptic conditions have the same lower bound of $\eta = 0$ as does classical inertial stability. As will now be shown, ellipticity and inertial stability are *not* equivalent measures of balance, even in the case of regular gradient flow.

b. New results

1) Nonlinear balance versus inertial balance

Árnason (1958) derived the following criteria for the case of a stationary circular vortex on a Northern Hemisphere *f*-plane. For ellipticity,

$$E^* \equiv (f_0 + 2\zeta_{curv})(f_0 + 2\zeta_{shear}) > 0, \tag{8}$$

and for inertial stability,

$$I^* \equiv (f_0 + 2\zeta_{curv}) + (f_0 + 2\zeta_{shear}) > 0. \tag{9}$$

Note that for the *f*-plane, $E^* = 2E$ and $f_0 I^* = 2I$ in the case of regular flow [cf. (1) and (7)]. These multiplying factors make the criteria more clearly comparable, but do not affect interpretation.

Extending Árnason's discussion, we expand (8):

$$E^* \equiv f_0[(f_0 + 2\zeta_{curv}) + (f_0 + 2\zeta_{shear})] - f_0^2 + 4\zeta_{curv}\zeta_{shear} > 0, \tag{10}$$

which can be expressed in terms of (9) as

$$E^* > 0: f_0 I^* > f_0^2 - 4\zeta_{curv}\zeta_{shear}. \tag{11}$$

Using (11), we can now directly compare nonlinear balance and inertial balance. Since from (9), $f_0 I^* > 0$ in the case of inertial stability in the Northern Hemisphere, the rhs of (11) is a measure of the difference between nonlinear and inertial balance. As a function of η, this expression traces out three solutions, two parabolic and one linear, depending on the relative signs and magnitudes of ζ_{curv} and ζ_{shear}. The generic flow configurations corresponding to these three cases are depicted graphically in Fig. 1; Fig. 2 illustrates the rhs of (11) at 30°N, in which it is assumed that $|\zeta_{shear}| \geq |\zeta_{curv}|$. Each case is now discussed separately.

Case A ($\zeta_{curv}/\zeta_{shear} > 0$) is commonly observed in large-scale flow that shear and curvature vorticities are of the same sign (Newton and Palmén 1963; Bell and Keyser 1993). An idealized depiction of this case is shown for anticyclonic conditions in Fig. 1a. Given this circumstance, a lower limit for solutions of (11) can be established:

$$f_0 I^* > f_0^2 - \zeta^2. \tag{12}$$

The derivation of (12) is given in the appendix. The lower bound of solutions of (12) is indicated by the solid parabola in Fig. 2; note that since case A solutions

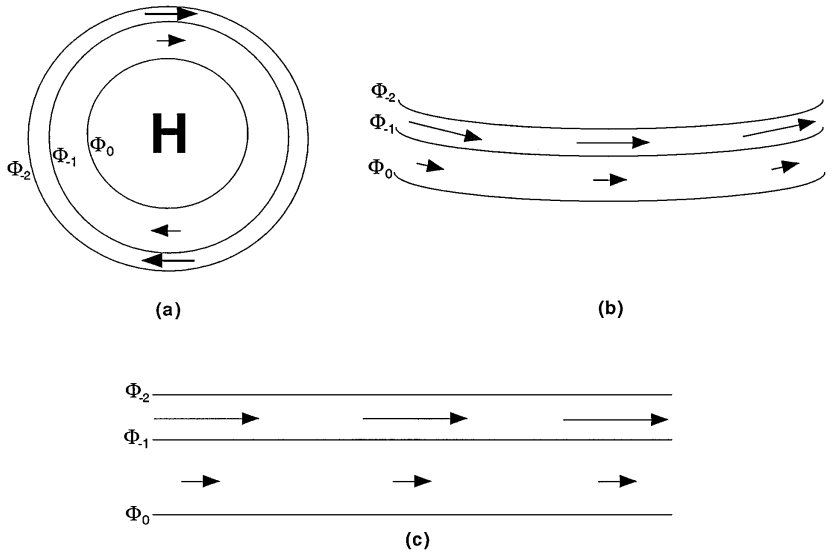

FIG. 1. Anticyclonic flow configurations corresponding to solutions of (11) in the Northern Hemisphere: (a) $\zeta_{curv}/\zeta_{shear} > 0$, (b) $\zeta_{curv}/\zeta_{shear} < 0$, and (c) $\zeta_{curv}/\zeta_{shear} = 0$. Arrows denote wind vectors; Φ denotes geopotential height contours, with negative subscripts indicating lower heights.

are always the lower limit for the rhs of (11), (12) represents the absolute lower limit for *all* solutions of (11). As such, the quantity $f_0^2 - \zeta^2$ might be useful as an easily calculated estimate of the difference between nonlinear balance and inertial balance for flows exhibiting both shear and curvature. This hypothesis is tested using a numerical model in section 7.

Since ζ is real, it follows from (12) that in case A

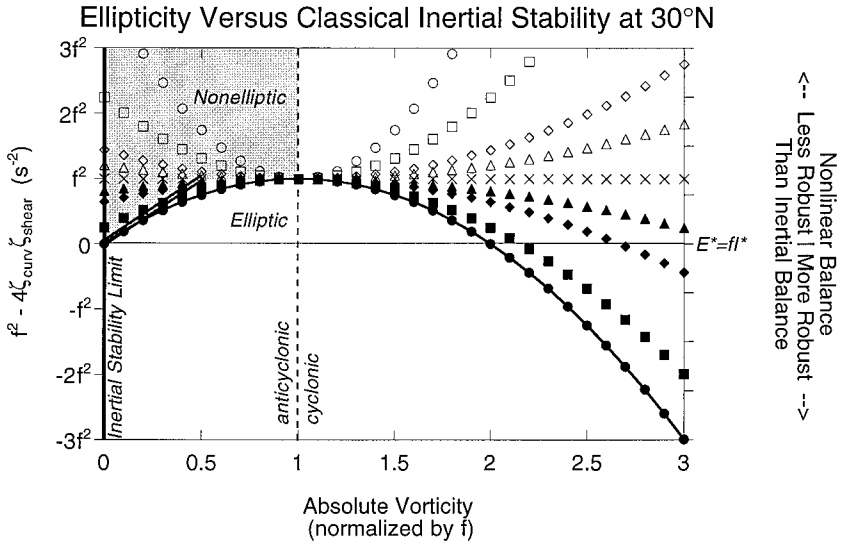

FIG. 2. Comparison of ellipticity and inertial stability as a function of absolute vorticity, based on (11) for *f*-plane gradient flow at 30°N. The thick vertical line at the left-hand side of the graph is the limit of classical inertial stability for regular flows [(7)]. The diagonal double line denotes the boundary between ellipticity and nonellipticity as defined by (17). Nonelliptic regions are shaded. The dashed vertical line separates anticyclonic from cyclonic flows (relative to the earth). The solid horizontal line is the locus of points where nonlinear balance and inertial balance are equivalent, denoted by $E^* = f_0 I^*$. The solid parabola is the lower limit of (12). Solid symbols correspond to case A solutions (see Fig. 1a); open symbols correspond to case B solutions (see Fig. 1b). The symbol code is as follows. Circle corresponds to $|\zeta_{curv}|/|\zeta| = 0.5$, square to $|\zeta_{curv}|/|\zeta| = 0.25$, diamond to $|\zeta_{curv}|/|\zeta| = 0.10$, and triangle to $|\zeta_{curv}|/|\zeta| = 0.05$. The *X* symbols correspond to case C (pure shear) solutions (see Fig. 1c).

situations the greatest possible minimum value of $f_0 I^*$ is f_0^2. For ellipticity to be equivalent to inertial stability, the rhs of (12) would need to be zero; Fig. 2 shows that this occurs at no more than two points for a given ratio of ζ_{curv} versus ζ_{shear}. More specifically, for flows with relatively small η, nonlinear balance is less robust than inertial balance; here, ''less robust'' means that the ellipticity condition is closer to being violated than is the inertial stability criterion, that is, $f_0 I^* > E^*$. This nonequivalence of ellipticity and inertial stability in anticyclonic flow is more pronounced for unequal ζ_{curv} and ζ_{shear}. The reverse is true for strongly cyclonic flows with large η. Although such flows are strongly elliptic and inertially very stable, ellipticity and nonlinear balance are further from violation than inertial stability in these cases, the nonequivalence being greatest when $\zeta_{curv} = \zeta_{shear}$.

Case B ($\zeta_{curv}/\zeta_{shear} < 0$) is for situations in which ζ_{curv} and ζ_{shear} oppose each other as in Fig. 1b [examples include the Rankine vortex (Kundu 1990); see also Newton and Palmén (1963), Fig. 15]; Fig. 2 reveals that nonlinear balance is always closer to violation than inertial balance. Here, f_0^2 is the smallest minimum value of $f_0 I^*$; therefore, the rhs of (11) is greater than zero, unlike (9) for the inertial stability criterion.

For case C ($\zeta_{curv}/\zeta_{shear} = 0$), the dividing line between cases A and B occurs when ζ is all shear or all curvature, for example, linear shear flow as in Fig. 1c, in which case (11) becomes

$$f_0 I^* > f_0^2. \tag{13}$$

Once again, nonlinear balance is less robust than inertial balance, since the rhs of (13) is always larger than zero (recalling that the analysis is for the midlatitude *f*-plane).

In the special case of pure shear flow where $\zeta = \zeta_g$, the difference between nonlinear and inertial balance can also be determined by writing the two criteria in terms of the geostrophic absolute vorticity $\eta_g = f_0 + \zeta_g$. For ellipticity,

$$\eta_g = \frac{f_0}{2} + \frac{E}{f_0} \geq \frac{f_0}{2}, \tag{14}$$

whereas from (6), for inertial stability,

$$\eta_g > 0. \tag{15}$$

2) NONELLIPTICITY VERSUS INERTIAL INSTABILITY

It is also possible to relate the *limits* of inertial and nonlinear balance, which are of significance since violations of inertial and nonlinear balance are of prime interest dynamically. We assume, as in Fig. 2, that large-scale flows are dominated by shear rather than curvature, that is, $|\zeta_{shear}| > |\zeta_{curv}|$. Then from (8) it is obvious that, in inertially stable or neutral flow at the limit of ellipticity,

$$E^*_{crit} : \zeta_{shear} = -f_0/2. \tag{16}$$

Using (11), (16) simplifies to

$$E^*_{crit} : f_0 I^* = f_0^2 + 2 f_0 \zeta_{curv}. \tag{17}$$

Values of $f_0^2 - 4\zeta_{curv}\zeta_{shear} > f_0^2 + 2 f_0 \zeta_{curv}$ are therefore nonelliptic; they are shaded in Fig. 2. The line of demarcation between ellipticity and nonellipticity defined by (17) is the diagonal double line in Fig. 2. (Note that all anticyclonic case B solutions are nonelliptic, since $\zeta_{shear} < -f_0/2$ in all such instances.) The limit of inertial balance, by contrast, is the solid line at the left-hand edge of the graph. From Fig. 2, it is seen that the limit of nonlinear balance and the limit of inertial balance are separate and distinct, although they are identical when the two criteria are identical at $\eta = 0$, $\zeta_{curv} = \zeta_{shear}$. Since both the ellipticity and inertial stability criteria are typically of most interest when they are near zero, that is, the point of nonellipticity or inertial neutrality, this special case—in which the two criteria are the same and are equally on the brink of violation—has perhaps fueled the incorrect assumption that ellipticity and inertial stability are identical in all cases. The results here show that in more general cases the ellipticity condition is usually violated for weaker anticyclonicity than is the inertial stability criterion.

3) COMPARISON WITH EXISTING THEORY

A remaining task is to reconcile the results presented above with seemingly contradictory claims in the extant literature. Since gradient flow is a subset of nonlinear balanced flow, how can (11) and Fig. 2 be reconciled with Paegle and Paegle's (1974) claim, apparently confirmed by Tribbia (1981), that the ellipticity condition of the nonlinear balance equation ''corresponds *exactly* to the condition of non-negative absolute vorticity'' (emphasis added)? A reexamination of Paegle and Paegle's work suggests that their ellipticity condition is not equivalent to the $\eta \geq 0$ condition for real solutions of the nonlinear balance equation [their Eq. (24)]; instead, their ellipticity condition should be identical to (1) on the *f*-plane. In contrast, their condition for real solutions is (1) plus the sum of the squares of the divergence and deformation terms, which are nonzero in general. [Miyakoda (1956) and Kasahara (1982), addressing this issue, both explicitly note this distinction between ellipticity and real solutions of the nonlinear balance equation; see Miyakoda's Fig. 1 and Kasahara's Eq. (11). Ghil et al. (1977) also derive an ellipticity condition for the nonlinear shallow-water equations that is not equivalent to inertial stability. Iversen and Nordeng (1982) observe that ellipticity and inertial stability coincide only when deformation and divergence are both zero, for example, the solid curve in Fig. 2.] Therefore, Paegle and Paegle's correspondence of ellipticity and inertial stability may not be exact after all, which is to be expected given the results of this section. Finally, Tribbia's (1981) own form of the ellipticity condition [his Eq.

(8b)] can easily be shown to lead to (8) and therefore (11).

The primary conclusions to be grasped here are that (a) even for simple flows, ellipticity and inertial stability are not synonymous, but rather are separate and distinct measures of balance, and that (b) for situations of weak inertial stability, ellipticity should be closer to violation than the classical inertial stability criterion. (However, due to the limitations of gradient flow, the discussion so far has not been able to address the relationship between the ellipticity and inertial stability criteria in regions of inertial instability.) In the following sections, we confirm and extend these results by examining more complex observed and modeled flows.

3. Ellipticity and inertial stability criteria in observed flows

A recurring enigma in numerical weather prediction and atmospheric dynamics has been the detection of large-scale regions of nonellipticity in observational datasets where (1) cannot be satisfied. Petterssen (1953) described a region of $E^* < 0$ at 500 hPa over the southeastern United States for 1 day in November 1952. This region was colocated with an area of small ($< 1 \times 10^{-5}$ s^{-1}) absolute geostrophic vorticity, as would be expected from (14) and (15). Bolin (1956) reported that 3%–5% of mid- and high-latitude 500-hPa data were nonelliptic during fall 1954. Using 1969–70 National Meteorological Center (now the National Centers for Environmental Prediction) analyses, Paegle and Paegle (1976a) found subtropical oceanic regions at 200 hPa where nonellipticity occurred as much as 80% of the time in winter; even the data-rich southern United States exhibited nonellipticity as much as 20% of the time at 200 hPa in the summertime. MacDonald (1977) also located regions of nonellipticity in data-rich regions of the troposphere over North America. Kasahara (1982) detected large regions of nonellipticity in 13 January 1979 European Centre for Medium-Range Weather Forecasting (ECMWF) analyses for the troposphere and stratosphere. Randel (1987), employing the *f*-plane ellipticity condition, identified tropical and midlatitude regions of nonellipticity in several days of 1979 and 1983 stratospheric data. However, questions have long been raised concerning the physical validity of the data from which these conclusions have been derived, since the nonelliptic regions tend to occur where data is sparsest or least reliable (Cressman 1959; Paegle and Paegle 1976a).

Similarly, large-scale regions that violate the classical inertial stability criteria (6) or (7) have been observed repeatedly. Since the 1940s, these regions have been detected on the equatorward flank of the jet streams (e.g., University of Chicago Staff Members 1947; Reiter 1961; Angell 1962; Ciesielski et al. 1989) and also on occasion in the midlatitude lower troposphere (Thorpe et al. 1993). Young (1981) noted negative absolute vorticity over the Indian Ocean at 900 hPa, 0°–8°N, 50°–70°E during July 1979. Hitchman et al. (1987) described persistent negative potential vorticity in the equatorial lower mesosphere, between roughly 0.1 and 1.0 hPa, and 0° and 10°N during the period November 1978–January 1979. Anomalous absolute vorticity (where $f\eta < 0$) has also been inferred from directly sensed winds in the zonal-mean equatorial stratosphere from 30 to 40 km at 2.5°N and 25°S around the 1992–95 winter solstices (Ortland et al. 1996). These regions, like regions of nonellipticity, also frequently occur in zones of sparse and questionable data, and arguments against their existence and persistence have surfaced from time to time (e.g., Blumen and Washington 1969; Leary 1974; Holton 1983).

It is rare, however, to see ellipticity and inertial stability examined simultaneously using the same dataset. To illustrate and relate these findings more compactly, we turn to a stratospheric–mesospheric dataset from the Limb Infrared Monitor of the Stratosphere (LIMS). Figure 3 shows a 6-day average from 12 to 17 December 1978 of zonal-mean ellipticity and zonal-mean inertial stability derived from LIMS Version 4 three-dimensional geopotential height data, which have been smoothed 1–2–1 in the meridional (see Hitchman and Leovy 1986 for more details concerning the dataset). In Fig. 3a, ellipticity is calculated from LIMS heights by retaining the β and metric terms in (1) and approximating $\overline{u}_\psi$ by $\overline{u}_g$. In Fig. 3b, inertial stability is calculated from (7) for $R \to \infty$; meridional wind shears near the equator are calculated using sixth-order finite difference schemes to minimize errors. In both figures, zonal winds are approximated by the zonal-mean zonal gradient wind [Randel 1987, Eq. (7)].

Figure 3 reveals that ellipticity and inertial stability are not equivalent in observed situations, as argued in section 2 for simple flows. Large, vertically coherent regions of nonellipticity exist between 30°S and 30°N, with the greatest occurrence being in the winter hemisphere; this is fully in accord with Kasahara (1982). Comparing Figs. 3a and 3b, nonellipticity is more prevalent than classical inertial instability, as expected from the discussion in section 2. Nevertheless, significant regions of classical inertial instability do exist in the equatorial winter mesosphere, which have been correlated with observed pancake structures by Hitchman et al. (1987). Violations of inertial stability are generally colocated with nonellipticity, but many nonelliptic regions are weakly inertially stable.

In Fig. 4, the difference between inertial and nonlinear balance is quantified for the same time period as in Fig. 3. All values in the figure are multiplied by 10^9, and positive values indicate that $I > E$; that is, inertial balance is more robust than nonlinear balance. From the figure it is obvious that, except for a small region near the equator, inertial balance is nearly always greater than

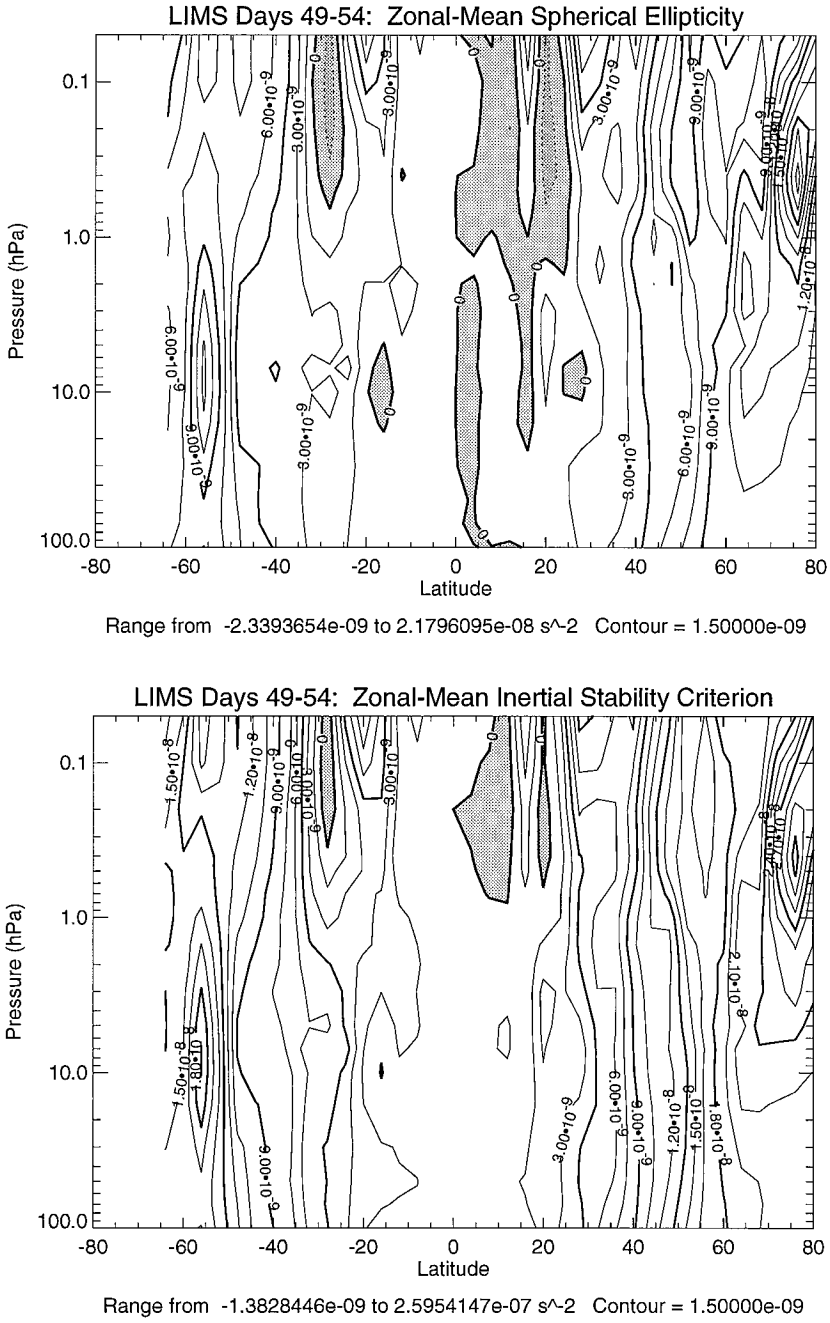

FIG. 3. Six-day averaged zonal means of quantities derived from LIMS Version 4 geopotential height data for 12–17 December 1978. Shown are (a) the spherical form of the ellipticity criterion (1) and (b) the zonal-mean inertial stability criterion (7). The southern extent of the LIMS dataset is 64°S. Negative values are shaded.

nonlinear balance, even in regions of classical inertial instability (cf. Fig. 3b).

Clearly, nonellipticity and inertial instability are not synonymous in theory or observations. But what physical mechanisms lead to each? As will now be shown, ellipticity is a simplified balance criterion that ignores some physical processes and can be generalized to account for these processes.

4. Beyond ellipticity I: Kasahara's realizability

The ubiquity of observed violations of the ellipticity criterion has prompted study of the physical premises upon which this mathematical criterion is based. Kasahara (1982) placed the nonellipticity problem in perspective by examining the condition of ellipticity for the divergence equation, which retains all terms ignored

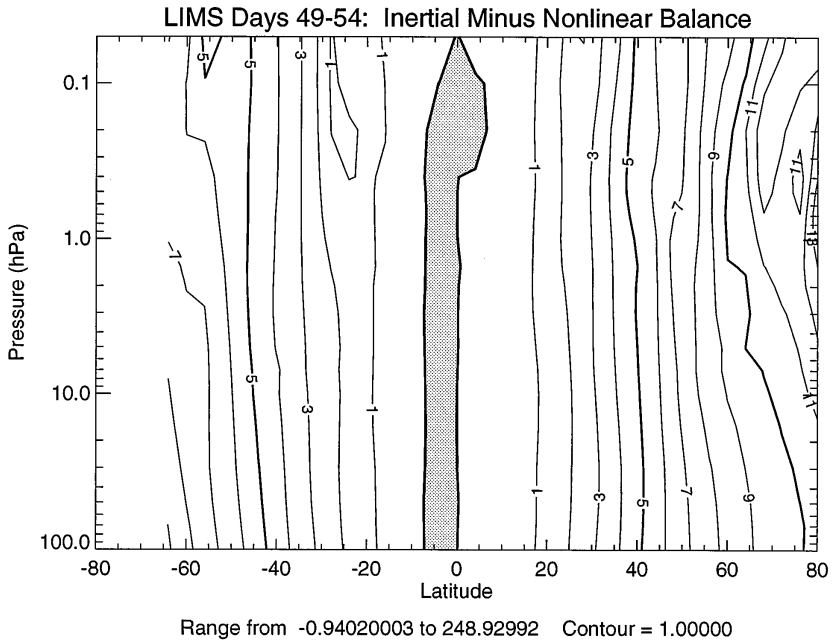

FIG. 4. The arithmetic difference of the inertial stability criterion (7) and the spherical ellipticity criterion (1) for the same dataset and time period as in Fig. 3. All values have been multiplied by 10^9. Negative values indicate that $I < E$ and are shaded; positive values indicate that $I > E$.

in the nonlinear balance equation from which the ellipticity criterion (1) is derived. The result of applying Monge–Ampère theory to the divergence equation is Kasahara's ''realizability,'' defined in spherical coordinates as

$$K \equiv E + F > 0, \tag{18}$$

where F is defined as

$$F = \frac{\partial D}{\partial t} + \mathbf{v}\cdot\nabla D + \omega\frac{\partial D}{\partial p} + D^2 + \nabla\omega\frac{\partial \mathbf{v}}{\partial p} + \beta u_\chi - \nabla\cdot\mathbf{X} - 2[J(u_\psi, v_\chi) + J(u_\chi, v_\psi) + J(u_\chi, v_\chi)] + \frac{u_\chi^2 + v_\chi^2 + 2(u_\psi u_\chi + v_\psi v_\chi)}{a^2} + \tan\phi\frac{\partial}{\partial\phi}\left(\frac{u^2 + v^2}{a^2}\right). \tag{19}$$

In (19), D is the horizontal velocity divergence in spherical coordinates, or $(a\cos\phi)^{-1}\,\partial u/\partial\lambda + \partial(v\cos\phi)/\partial\phi$; $\mathbf{v}$ indicates the horizontal velocity; the gradient operator ∇ is $(a\cos\phi)^{-1}\,\partial/\partial\lambda + a^{-1}\,\partial/\partial\phi$; ω is the vertical velocity in isobaric coordinates; the subscript χ indicates the divergent component of the circulation; $\mathbf{X}$ is the contribution from subgrid-scale horizontal motions; and $J(u,v)$ is the Jacobian operator $(a^2\cos\phi)^{-1}\,(\partial u/\partial\lambda\,\partial v/\partial\phi - \partial v/\partial\lambda\,\partial u/\partial\phi)$. It should be noted with reference to F that the βu_χ term was apparently omitted by Kasahara (cf. Haltiner and Williams 1980, 69); in addition, the *divergence* of the subgrid-scale terms was not explicitly indicated by Kasahara, nor was the form of the terms in the second, third, and fourth lines of (19).

Using this more general approach to the problem, Kasahara showed that in the zonal average, nonelliptic regions in the Tropics, where (1) is not satisfied, do indeed satisfy the more general solvability condition (18); that is, they are realizable. In short, ''nonelliptic regions exist in the atmosphere because the terms neglected in deriving the nonlinear balance equation are not always negligible'' (Daley 1991). Hence, ellipticity is a mathematical constraint on an equation that does not always describe the atmosphere accurately, whereas Kasahara's realizability is a mathematical constraint on a much more general equation. The implication is that $K = 0$ is the appropriate demarcator between balanced and unbalanced flows for regions that satisfy the primitive equations but not the nonlinear balance equation.

Which term or terms in F cause regions of the atmosphere to deviate from nonlinear balance? This question is identical to asking which physical processes distinguish Kasahara's realizability from ellipticity. Kasahara rejected the terms D^2 and $\mathbf{v}\cdot\nabla D$ on the basis of quantitative arguments, although it should be pointed out that late 1970s ECMWF analyses likely suffered from a deficiency in divergence (Hoskins et al. 1989). Instead, Kasahara speculated that ''(i)t is conceivable that none of the terms in [19], except for the contribution . . . due to subgrid-scale motions, is responsible for producing a positive imbalance F in the nonelliptic regions.''

A similar conclusion was reached by Randel (1987) for the stratosphere, who did not acknowledge Kasahara's work. Large regions of nonellipticity were noted at 10 hPa in the Tropics and the midlatitudes during a sudden stratospheric warming. Furthermore, a strong correlation existed between the residual term and the nonelliptic regions. Randel suggested that ''the true balance in these areas must contain important contributions from either vertical advective terms or scales that are not resolved here.''

Since Kasahara's study, it does not appear that his criterion has been applied to observed or modeled flows. In the recent literature, Daley (1991) remarked approvingly on Kasahara's extension of ellipticity but did not extend the discussion further.

5. Beyond ellipticity II: Petterssen's realizability

Although Kasahara's realizability criterion reincorporates all terms that are ignored in the derivation of the ellipticity condition, this is *not* equivalent to stating that Kasahara's criterion is the most general balance criterion. On the contrary, it is merely the most general solvability condition that can be derived from the divergence equation and Monge–Ampère theory. However, another realizability criterion exists that is based only on primitive equation assumptions and algebra, with no direct reference to the theory of partial differential equations. This criterion was first obtained in a simpler context by Petterssen (1953); it is derived below for the full primitive equations, and the relationships between Petterssen's criterion, Kasahara's realizability, and the inertial stability criterion are clarified.

Beginning with the horizontal frictional primitive equations, the divergence equation can be expressed as

$$\frac{\partial D}{\partial t} + \mathbf{v}\cdot\nabla D + \omega\frac{\partial D}{\partial p} + \nabla\omega\cdot\frac{\partial \mathbf{v}}{\partial p} + \frac{1}{2}(A^2 + B^2 + D^2 - \zeta^2) = -\nabla^2\Phi + f\zeta - \beta u + \nabla\cdot\mathbf{X}, \tag{20}$$

where A is the shearing deformation, defined by Petterssen in Cartesian coordinates as $\partial u/\partial x - \partial v/\partial y$, and B is the stretching deformation $\partial v/\partial x + \partial u/\partial y$. (Note that the velocity retains both the divergence and rotational components, in contrast to the nonlinear balance equation.) By multiplying (20) by 2 and adding f^2 to each side, the following exact expression is obtained:

$$(f + \zeta)^2 = P, \tag{21}$$

where

$$P \equiv 2\left(E + \frac{\partial D}{\partial t} + \mathbf{v}\cdot\nabla D + \omega\frac{\partial D}{\partial p} + \nabla\omega\cdot\frac{\partial \mathbf{v}}{\partial p} + \beta u_\chi - \nabla\cdot\mathbf{X}\right) + A^2 + B^2 + D^2 \geq 0. \tag{22}$$

Petterssen (1953) obtained a version of (21) and (22) that did not include frictional or tilting terms and was expressed as $P^{1/2}$. Interestingly, the ellipticity E arises in (22) despite the fact that no appeal has been made to Monge–Ampère theory.

Equation (21), which is solely a consequence of the divergence equation and algebraic rearrangement, can be interpreted in at least two complementary ways. It has been customary (e.g., Petterssen 1953; Miyakoda 1956; Reiter 1963; Paegle and Paegle 1974; Kasahara 1982) to take the square root of (21) to obtain an expression for the absolute vorticity, the usual inference being that in the Northern Hemisphere $\eta \geq 0$ as with regular solutions of the gradient wind equation. This inference is not completely persuasive, given the previous discussion concerning the many observations of anomalous η and also the existence of primitive-equation models that easily develop anomalous potential vorticity (e.g., O'Sullivan and Hitchman 1992). An equally valid interpretation focuses on the rhs of (21): from this perspective, (21) defines a new measure of nonlinear balance P and requires P to be nonnegative *even if the absolute vorticity is anomalous.*

Given this balance-criterion interpretation of (21), P is Petterssen's realizability, a physically derived measure of balance that retains the full physics of the primitive equations with *no* other assumptions. (Here, ''physically derived'' means that the balance measure is obtained directly from $\mathbf{F} = m\mathbf{a}$, not from solvability conditions.) For $P \geq 0$ and large radius of curvature R, the relationship between this balance criterion and inertial stability criterion (7) is particularly simple:

$$P = \frac{I^2}{f^2}. \tag{23}$$

For small-R flows that are still solutions of the primitive equations, the denominator in (23) is $(f + 2V/R)^2$; but in such cases $I^2 = (f + \zeta)^2(f + 2V/R)^2$, and so the modification divides out.

Petterssen's realizability also can be related to other balance criteria. Using the algebraic rearrangement

$$A^2 + B^2 + D^2 = \zeta^2 + 2D^2 - 4J(u, v), \tag{24}$$

a new expression for P in terms of Kasahara's realizability criterion (18) can be derived:[1]

$$P = 2K + \zeta^2 - 4J(u_\psi, v_\psi). \tag{25}$$

The particular importance of P as expressed above is that on the one hand, it can be related to I through (23); on the other hand, it is related to K and therefore all the nonlinear balance criteria through (25). Therefore, P can be used to establish generalized relationships between balance criteria such as Kasahara's realizability and ellipticity and the classical inertial stability crite-

[1] A reviewer notes that Eq. (10) in Kasahara (1982) is not the same as nor is directly comparable to (25) since the source term F in section 2 of Kasahara does not include second-order derivatives of the streamfunction.

TABLE 1. Summary of nonlinear balance criteria discussed in the text, including the approximations inherent in each criterion.

Criterion	Symbol	Derived from	*f*-plane?	Steady?	Nondivergent?	Inviscid?
Classical inertial stability	I	Linearized equations of motion	Yes	Yes	Yes	Yes
Ellipticity	E	PDE theory and balance equation	No	Yes	Yes	Yes
Motion parameter	M	Modified divergence equation	Yes	Yes	No	Yes
Kasahara's realizability	K	PDE theory and divergence equation	No	No	No	No
Petterssen's realizability	P	Divergence equation	No	No	No	No

rion. This approach is explored in detail in the next section.

Another useful aspect of P is the physical significance of violations of $P \geq 0$. Any violations of (22) dissolve the relationship between Petterssen's realizability and inertial stability in (23) and are directly interpretable as violations of the dynamical assumptions that undergird the traditional primitive equations: hydrostatic balance and the smallness of the horizontal rotational component of the Coriolis force (Andrews et al. 1987). As such, violations of (22) should correspond to more physically tangible flow regimes than violations of the other criteria, perhaps being manifested as nonhydrostatic effects associated with vigorous divergent circulations. (Although it is true that the primitive equations themselves are but a mathematical artifice, violations of the assumptions that lead to them are more comprehensible physically than violations of solvability conditions.)

Petterssen's criterion has been noted periodically in the literature. However, its status as a physically based generalized balance criterion has been obscured by other interests. Miyakoda (1956) derived an inviscid *f*-plane version of (22) and noted that ''it is somewhat curious'' that the quantity P should be compelled to be nonnegative. Miyakoda's goal was to solve the balance equation, however, and so any negative values of his version of (22) were automatically set to zero. Reiter (1963) discussed Petterssen's version of (21), but approximated it in order to quantify the differences between geostrophic and observed relative vorticities. Later, Paegle and Paegle (1974) and MacDonald (1977), following Petterssen, derived a less general ''motion parameter,'' M, valid for the *f*-plane:

$$M \equiv 2E + A^2 + B^2, \tag{26}$$

and they used negative values of M to indicate regions of high-divergence steady-state flows. More pertinent to this discussion, Kasahara (1982) carefully noted that E, K, and P are distinct: ''The ellipticity measure [E] . . . or the realizability measure [K] . . . can be negative *locally*, as long as [P] remains positive . . . 'realizability' used here refers to a mathematically required situation . . . [whereas $P \geq 0$ is] a physically necessary situation in order to yield physically realizable flows.'' There do not seem to be any subsequent discussions of Petterssen's criterion, nor any applications of this criterion to observed or modeled flows.

As a summary, Table 1 defines the various balance criteria discussed in sections 2 through 5 and lists the assumptions that undergird each criterion.

6. Balance criteria interrelationships

Using the results of the previous section, we can relate nonlinear balance criteria to the inertial stability criterion I as was done in section 2. In section 2, the analysis was limited to a comparison of ellipticity and inertial stability in regular gradient flow. Given the results of sections 4 and 5, and in particular (23) and (25), the treatment can now be expanded to include a suite of balance criteria governing the full range of primitive-equation flows.

The relationships between the various balance criteria, which are predicated on the primitive-equation assumptions (and large R for relations with I), are presented as a matrix in Table 2. These relationships have

TABLE 2. Matrix of conversions between all nonlinear balance criteria discussed in the text, given that primitive-equation assumptions are satisfied. All symbols are defined in the text.

I (*f*-plane, large R)	E	M (*f*-plane)	K	P
I —	$\{f^2[2E + 2F + \zeta^2 - 4J(u_\psi, v_\psi)]\}^{1/2}$	$\langle f^2\{M + 2F - D^2 + 4 \cdot[J(u, v) - J(u_\psi, v_\psi)]\}\rangle^{1/2}$	$\{f^2[2K + \zeta^2 - 4J(u_\psi, v_\psi)]\}^{1/2}$	$(f^2P)^{1/2}$
E $[I^2/f^2 - \zeta^2 + 4J(u_\psi, v_\psi)]/2 - F$	—	$(M - A^2 - B^2)/2$	$K - F$	$[P - \zeta^2 + 4J(u_\psi, v_\psi)]/2 - F$
M $I^2/f^2 + D^2 - 4 \cdot[J(u, v) - J(u_\psi, v_\psi)] - 2F$	$2E + A^2 + B^2$	—	$2K - 2F + A^2 + B^2$	$P + D^2 - 4[J(u,v) - J(u_\psi, v_\psi)] - 2F$
K $[I^2/f^2 - \zeta^2 + 4J(u_\psi, v_\psi)]/2$	$E + F$	$(M + 2F - A^2 - B^2)/2$	—	$[P - \zeta^2 + 4J(u_\psi, v_\psi)]/2$
P I^2/f^2	$2E + 2F + \zeta^2 - 4J(u_\psi, v_\psi)$	$M + 2F - D^2 + 4 \cdot[J(u, v) - J(u_\psi, v_\psi)]$	$2K + \zeta^2$-$4J(u_\psi, v_\psi)$	—

been determined from (1), from (7) for large R, and from (18), (21), (23), (24), (25), and (26).

As in section 2, it is clear from Table 2 that ellipticity and inertial stability are not equivalent criteria. Using (7), (18), (23), and (25), for significant curvature the exact relationship is

$$E = \frac{1}{2}\left[\frac{I^2}{\left(f + \frac{2v}{R}\right)^2} - \zeta^2 + 4J(u_\psi, v_\psi)\right] - F. \quad (27)$$

At first glance, (27) might seem to have little in common with the relationship (11) between E and I derived for gradient flow in section 2. However, by using relation (24), (27) is easily simplified to

$$E = \frac{f^2}{2} + f\zeta + 2J(u_\psi, v_\psi) - F. \quad (28)$$

Since $J(u_\psi, v_\psi) = \zeta_{curv}\zeta_{shear}$ and $F = 0$ for a stationary circular vortex, (28) is identical to $E^*/2$ as defined in (8), from which (11) and the E^*–I^* relationships in Fig. 2 are derived. Furthermore, for $F = 0$ (28) reduces to the familiar f-plane version of $E = f^2/2 + \nabla^2\Phi$. This is an example of how the relations in Table 2 are the primitive-equation generalizations of the results in section 2.

Table 2 also highlights how the generalized balance criteria relate to each other and to the classical inertial stability criterion. For example, by reading down the second column of Table 2, it is found that the differences between E and the more general balance criteria are usually some combination of divergence or frictional effects implied by F and flow deformation in A and B. In contrast to the rather complex relationship between E and I found in section 2, Table 2 shows that I and Petterssen's criterion are intimately related for all primitive-equation flows, regardless of the amount of divergence, friction, or deformation. Note also that P may remain nonnegative even in primitive-equation flows where divergence or frictional effects render K negative, since $P \approx 2K + \zeta^2$ when the Jacobian term is negligible.

In the following section, the balance criteria listed in Table 1 and interrelated in Table 2 are compared and contrasted using a simple numerical model. In section 8, we will return to these tables and explore the ramifications for inertial instability research.

7. Generalized balance criteria in a simple nonlinear model

a. Trajectory model description

We now analyze the evolution of the balance parameters in Table 1, using a simple trajectory model first derived by Paegle and Paegle (1976b). The intent of this section is to demonstrate the utility of these criteria in a situation uncluttered by the myriad complications of application to "real" data or model output. As will be demonstrated, different measures of nonlinear balance evolve quite differently depending on the properties of the flow, confirming and extending the results of section 2.

The model is based upon the divergence, absolute vorticity, shearing deformation, and stretching deformation equations on an inviscid f-plane:

$$\frac{dD}{dt} = \frac{1}{2}(\eta^2 - A^2 - B^2 - D^2) - E, \quad (29)$$

$$\frac{dA}{dt} = -DA + f(B - B_g), \quad (30)$$

$$\frac{dB}{dt} = -DB - f(A - A_g), \quad (31)$$

$$\frac{d\eta}{dt} = -\eta D, \quad (32)$$

where d/dt is the material horizontal derivative and the subscript g denotes the geostrophic component. Ignoring the tilting terms, assuming a circular anticyclonic pressure field and a constant value of the ellipticity, the system of equations closes to

$$\frac{dD}{dt} = \frac{1}{2}(b\eta^2 - D^2) - E, \quad (33)$$

$$\frac{d\eta}{dt} = -\eta D, \quad (34)$$

where $b = 1 - (A_0^2 + B_0^2)/\eta_0^2$, in which the subscript $_0$ denotes the initial value of a variable. The model results below are based on the system of (33) and (34), and are interpretable (when $\eta > 0$) as perturbations on a gradient-balance flow.

The background state in the model is assumed to be unvarying and to have uniform η and geostrophic deformations. Thus, the larger the total parcel excursion is, the less reliable are the assumptions employed in the model since the parcel would eventually leave its original environment. The solutions are also quantitatively valid only to the extent that E, A_g, and B_g are relatively uniform throughout the trajectory. However, since the purpose of this experiment is to compare balance criteria qualitatively rather than to describe the flow quantitatively, this limitation should not affect the conclusions significantly. Paegle and Paegle (1976b) favorably compared results from this model to observed cases of nonelliptic flow. Despite the simple form of the model, it (a) retains strong nonlinearity, (b) is solvable in regions of nonellipticity, unlike the nonlinear balance equation, and (c) permits easy and rapid calculation of the generalized balance criteria defined above, *with respect to the parcel*. In particular, this model permits us to test the impact of irrotational processes on the measure of balance on an f-plane.

Paegle and Paegle (1976b) solved their trajectory

TABLE 3. Parameters used in the model calculations shown in Figs. 5–10. Values are based on Reiter (1963), observations from Kasahara (1982) and Hitchman et al. (1987), and values employed by Paegle and Paegle (1976b). A latitude of 30°N is assumed.

Regime	ζ_g/f	E ($\times 10^{-10}$ s^{-2})	η_0 ($\times 10^{-5}$ s^{-1})	D_0 ($\times 10^{-5}$ s^{-1})	$(A^2 + B^2)/\eta^2$
Strongly cyclonic (Fig. 5)	3.26	200	20	0.25	0.25
Weakly cyclonic (Fig. 6)	0.44	50	10	0.25	0.25
Weakly anticyclonic (Fig. 7)	−0.12	20	6	0.25	0.25
Strongly anticyclonic (Fig. 8)	−0.48	1	2	0.50	0.75
Inertially unstable (Fig. 9)	−0.69	−10	−0.5	1.00	1.75
Strongly inertially unstable (Fig. 10)	−1.06	−30	−5	1.00	1.75

model for D and η analytically but did not examine the evolution of balance parameters within the model, the behavior of elliptic solutions, or inertially unstable flows for synoptic-scale conditions. The model exhibits sinusoidal, steady, and asymptotic solutions; the character of the solution changes with the sign of the ellipticity, not the sign of the absolute vorticity. The reasons for this will be explored in a future article.

The model was run with a variety of realistic, although spatially uniform, background states; these and the initial conditions imposed are listed in Table 3. These conditions were obtained primarily from values employed by Paegle and Paegle (1976b), and data in Kasahara (1982) and Hitchman et al. (1987). Values of η_0 were estimated from geostrophic relative vorticity using a relation found in Reiter [1963, Eq. 1.239 (11)]; however, Reiter's equation fails for inertially unstable flow. The last two cases in Table 2 are based largely on Fig. 3, which shows regions of nonellipticity colocated with regions of weak and strong classical inertial instability. A latitude of 30°N is assumed throughout to facilitate calculation of parameters and comparison with Fig. 2.

In all model runs, a fourth-order Runge–Kutta method with a time step of 100 s was employed (Vreugdenhil 1994), with calculations ending at $t = 5 \times 10^5$ s (about 5.8 days). For each run, the following quantities were calculated with respect to the parcel: D, η, and the five balance criteria $2E$, I, $2K$, P, and M, which are defined

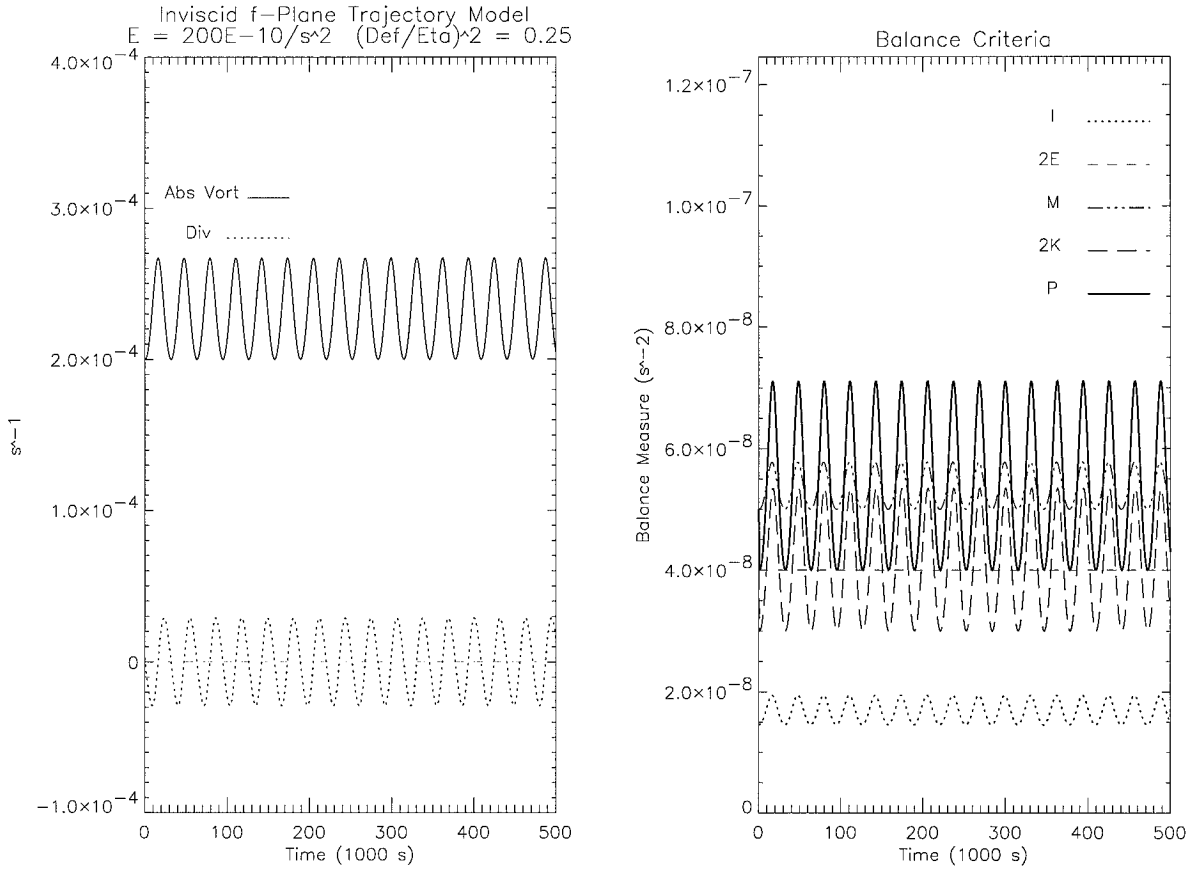

FIG. 5. Time evolution of the inviscid f-plane trajectory model described in section 7, for strongly cyclonic flow, as described in Table 3. Variables plotted are (a) absolute vorticity and horizontal divergence, and (b) generalized balance criteria, as defined in Tables 1 and 2.

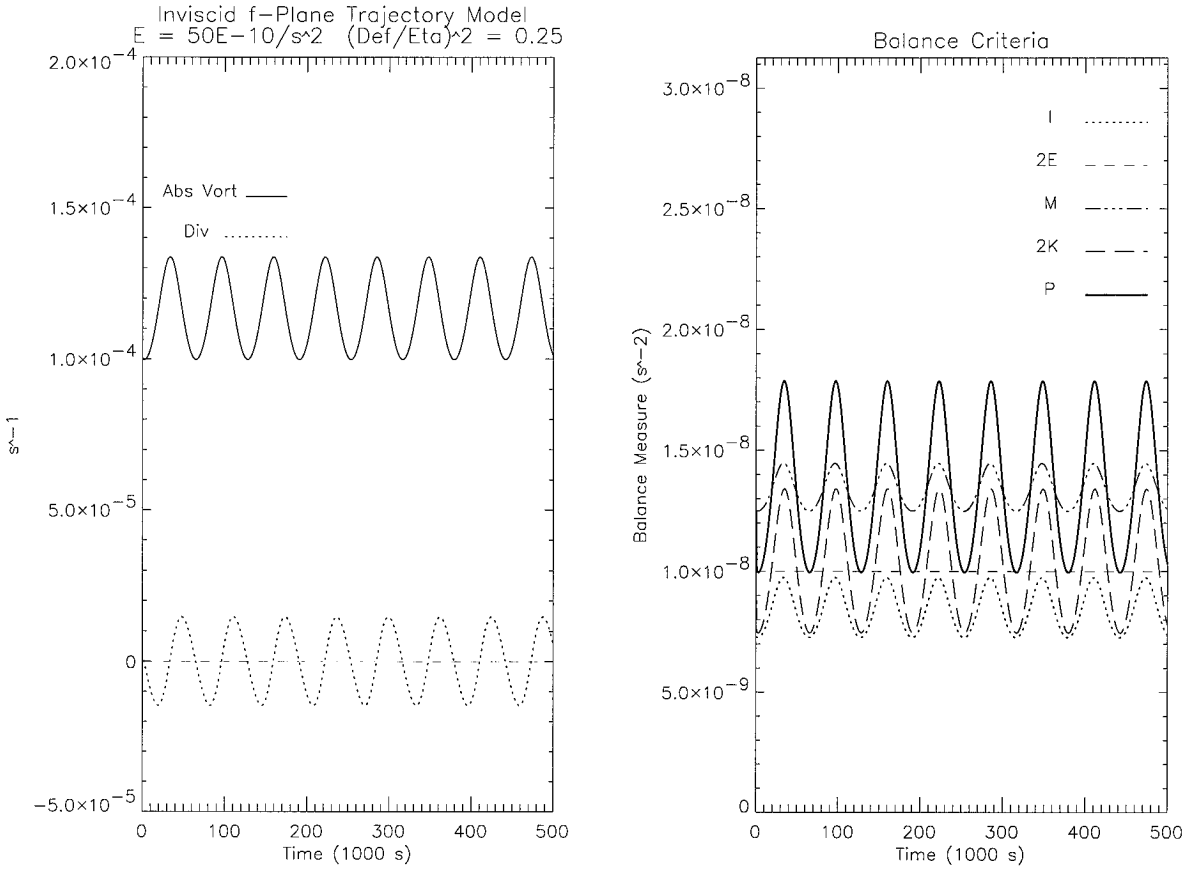

FIG. 6. As in Fig. 5 but for weakly cyclonic flow, as described in Table 3.

in (1), (7) with $R \to \infty$, (18), (25), and (26), respectively. As in section 2, a factor of 2 is included in some cases to make all criteria directly comparable. In this simple model, only D, dD/dt, and the deformation terms A and B are retained in the more general criteria. The ageostrophic Jacobian terms in K and P have been omitted; calculations of K with and without these terms reveals that their omission changes the value of the balance criteria by less than 1%.

b. Model results

Figures 5 through 10 illustrate the evolution of the balance criteria listed in Table 1 for the flow regimes listed in Table 3.

For strongly cyclonic flow (Fig. 5), an oscillating solution is obtained for D and η. The balance criteria in all elliptic runs mirror the evolution of η, which corresponds to our understanding of the close relationship between nonlinear balance and inertial balance. However, the criteria are not identical. The lower bound for P is $2E$ [see (25)], suggesting that ellipticity underestimates balance due to its inherent approximations as Kasahara (1982) posited. All criteria indicate that this situation is balanced, in accord with expectations. Note that the inertial stability criterion I suggests a weaker balance than all other criteria, including $2E$ or even E. Comparing this result to the simple calculations in section 2, we find in Fig. 2 that inertial stability is less robust than ellipticity for $\eta > 2$–$3f$ in case A situations, the exact value depending on the curvature of the flow. In the model, $\eta \geq 2.74f$ for this case; hence the relationship between E and I in this run is in good agreement with Fig. 2.

Weakly cyclonic flow (Fig. 6): Again, inertial stability gives the least balanced assessment of the flow. However, even though $I < 2E$, note that $I > E$. In this simulation, $1.4f \leq \eta \leq 1.8f$, and Fig. 2 suggests that in this range nonlinear balance should be less robust than inertial balance, in agreement with the model. Kasahara's criterion is bounded below by inertial stability and above by MacDonald's motion parameter. The Petterssen criterion suggests greater balance than the other criteria for periods of large η.

Weakly anticyclonic flow (Fig. 7): In this case, I suggests much more robust balance than E, particularly compared to the weakly cyclonic case. This matches expectations based on Fig. 2: by the theory of section 2, the greatest difference between inertial and nonlinear balance should occur for flows with weakest relative

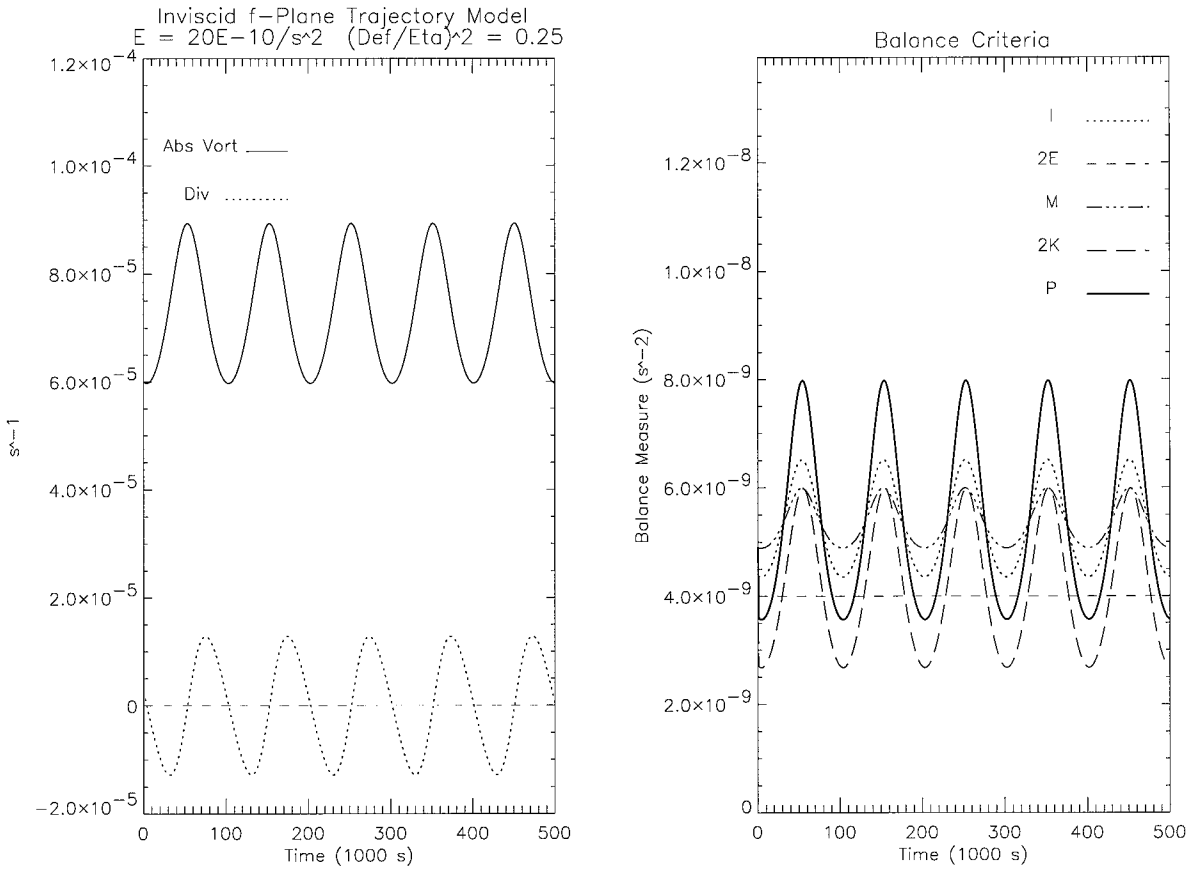

FIG. 7. As in Fig. 5 but for weakly anticyclonic flow, as described in Table 3.

vorticity, and from Table 3 it is clear that this case has weaker relative rotation than the weakly cyclonic case in Fig. 6. Of all the generalized criteria, Kasahara's realizability diagnoses the weakest balance.

Strongly anticyclonic flow (Fig. 8): A much slower oscillation is obtained in this case, qualitatively in agreement with classical theory. As before, the inertial stability criterion now *over*estimates balance versus E and $2E$; here, $0.2f \leq \eta \leq 0.7f$, and from Fig. 2 it is seen that ellipticity is indeed closer to violation than inertial stability in this range. The arithmetic difference $I - E$ is less than that in Fig. 7, however, which also is in accord with Fig. 2.

Classically inertially unstable flow (Fig. 9): The model behavior changes markedly, corresponding to the steady ''vergic'' balance described by Paegle and Paegle (1976b) and MacDonald (1977). The flow evolution is interpretable as an inertially unstable parcel attempting to achieve inertial neutrality through strong divergence. The linkage between observed regions of nonellipticity and strongly divergent flows has been noted by, among others, MacDonald (1977). Inertial instability and strongly divergent flow patterns have been colocated in a number of primitive-equation model simulations (e.g., O'Sullivan and Hitchman 1992). Here, E is negative and M asymptotes to E; $2K$, on the other hand, asymptotes to $-2E$. In this case, the asymptotic behavior is linked to the relationship between the deformation terms and η assumed in the model. Also, I converges to zero on the timescale of 2 days, and P asymptotes to zero much more quickly. This is of course to be expected, since $P \propto I^2$. Note that P is much smaller than $2K$, although both are nonnegative. As discussed in section 5, a value of P of zero would imply the least balance achievable for a primitive-equation model, and so this suggests that the vergic state is on the precipice of imbalance, as is normally assumed for high-divergence, near-zero absolute vorticity states.

Strongly inertially unstable flow (Fig. 10): For this regime, the differences between the various criteria become evident. As the nonellipticity is increased, Kasahara's realizability increases, since it asymptotes to $-2E$. This behavior does not seem to be especially useful for a balance criterion; one might logically prefer instead a measure that monotonically decreases with decreasing balance. Both I and P converge to zero as in Fig. 9. Nonlinear balance, as measured by E, becomes much less robust than inertial balance (i.e., $E < I$) as $\eta \rightarrow 0$ from large negative absolute vorticities. If the parabolic relationships shown in Fig. 2 hold approxi-

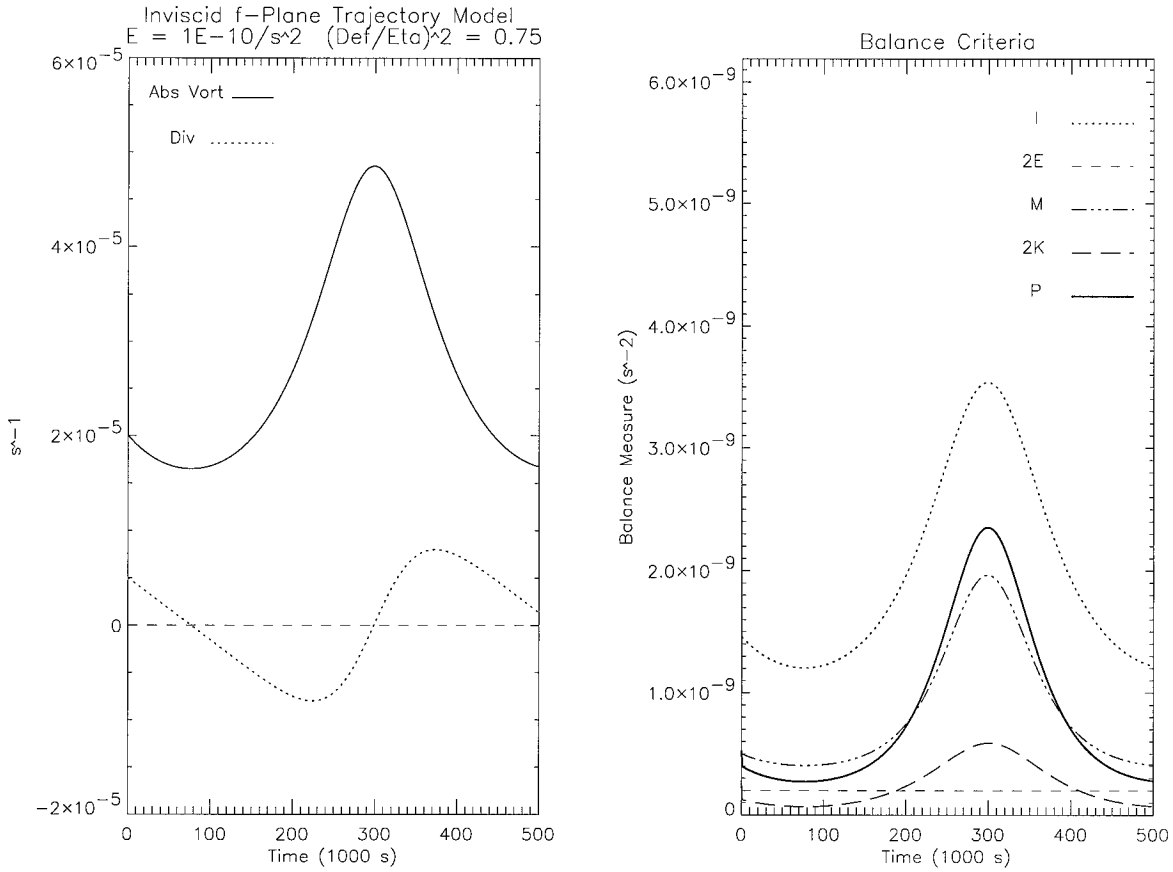

FIG. 8. As in Fig. 5 but for strongly anticyclonic flow, as described in Table 3.

mately for the inertially unstable case, then this result is in complete agreement with the case A theory in section 2; the result also appears to be consistent with the observational data shown in Figs. 3 and 4.

c. Interpretation of balance criteria intercomparison

The model results above show that the balance criteria described in Table 1 have very different Lagrangian evolution in a nonlinear model, and this behavior is at least qualitatively consistent with the simple theory derived in section 2. This consistency is not limited to the points discussed above. For example, the conjecture made in section 2 that the difference between nonlinear and inertial balance can be estimated by the quantity $f^2 - \zeta^2$ is borne out in Fig. 11, which shows the ratio $(f^2 - \zeta^2)/(I - E)$ for strongly anticyclonic flow.

Of all the criteria, P appears most useful for diagnosing balance in a generalized sense, since the distance P is from zero is a reliable measure of the amount of balance and $P \geq 0$ in all model runs, which implies satisfaction of primitive-equation assumptions. By contrast, K varies nonmonotonically as balance is reduced, increasing without bound as $E \ll 0$. This is evidence for the utility of P as a diagnostic of generalized balance.

It could be pointed out that, from (23) and (25), P is merely a proxy for I, and so in fact I is the most useful balance measure. This is true for the model simulations here, since violations of the primitive equations are not permitted in the model. However, in more general situations (e.g., a nonhydrostatic model), P should be able to provide information not obtainable from I concerning violations of primitive-equation assumptions. As an illustration, despite several numerical studies of anomalous potential vorticity in inertial instability, the relationship between inertial instability and nonhydrostatic conditions is still a matter of speculation (e.g., McWilliams 1991; O'Sullivan and Hitchman 1992). From (23) and the studies themselves, it is clear that $I < 0$ can exist in primitive-equation flows; but are *all* occurrences of $I < 0$ governed by the primitive equations? Merely examining I provides no answer to this question. However, an accurate calculation of P could provide more insight into the relationship, since regions of $P < 0$ are unequivocal evidence of violations of the assumptions implicit in the divergence equation, which in turn is derived from the primitive equations.

To explore violations of the Kasahara and Petterssen criteria and the physical causes for such violations, experiments must be run with a much more complex model

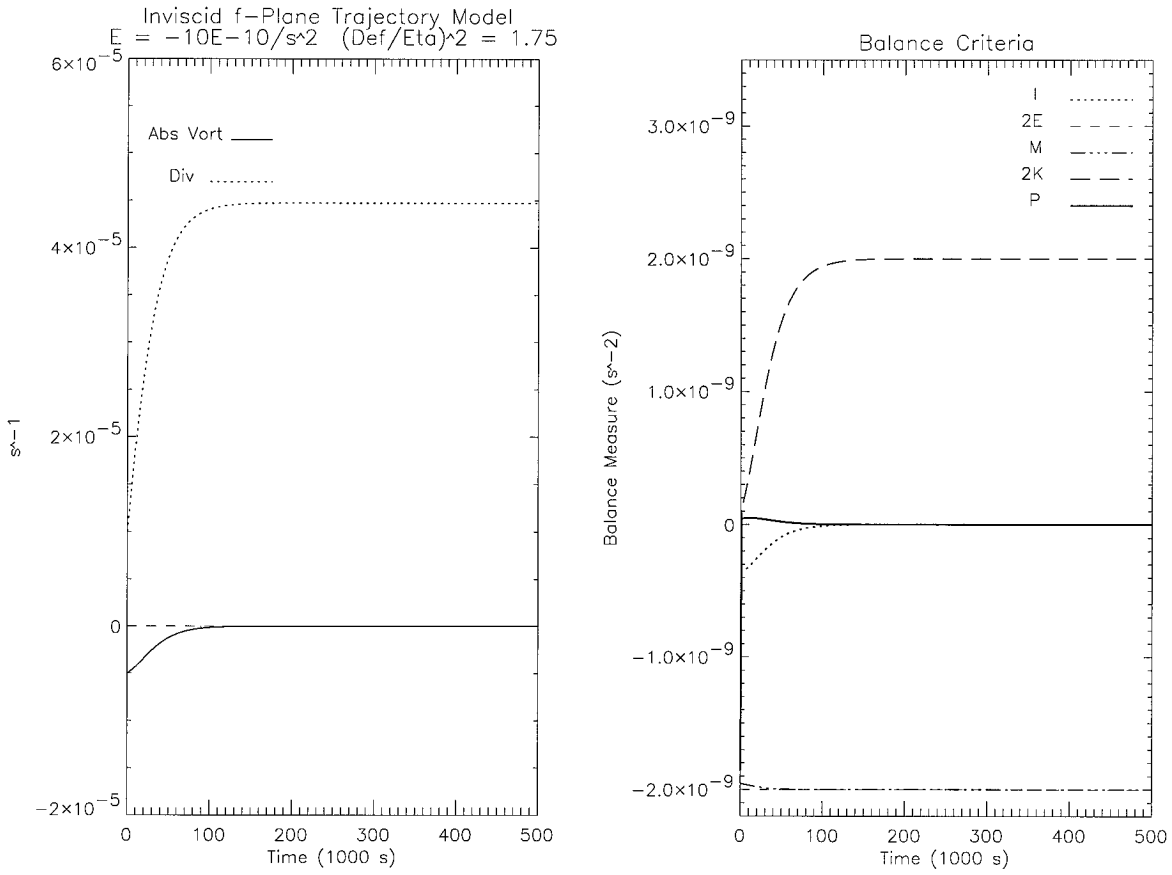

FIG. 9. As in Fig. 5 but for nonelliptic, inertially unstable flow, as described in Table 3.

or dataset. This work is in progress; results will be discussed in future articles.

8. Implications for inertial instability research

The previous sections have examined generalized nonlinear balance theory in connection with inertial stability. Conversely, one may also consider generalizations of inertial instability theory in light of nonlinear balance and the approach taken in the previous sections. Certainly the comparison is warranted: inertial instability and nonellipticity both describe situations of strongly anticyclonic flow in or near the Tropics, and the traditional response to the presence of either in datasets has been to remove the offending region artificially (Cressman 1959; Holton 1983; Tan and Curry 1993). The proof in section 2 that inertial stability and ellipticity are not identical notwithstanding, the similarities between violations of the two criteria invite further comparison.

From Table 1, it is clear that both the ellipticity criterion and the classical inertial stability criterion are derived from inviscid theory that ignores much of the nonlinearity inherent in the primitive equations (see also Stevens and Crum 1987). In the case of ellipticity, these assumptions have been shown by Kasahara (1982) to be unwarranted for observed flows, and the model results in section 7 also suggest that ellipticity is not the most general balance criterion available. Table 2 illustrates the generalization of ellipticity beyond its restrictive assumptions. It seems clear that some mechanisms, probably frictional and/or divergent in nature, frequently disrupt nonlinear balance while the overall situation is still a (possibly steady) solution of the primitive equations, that is, ''realizable'' according to the Kasahara or Petterssen criteria.

Turning now to inertial instability theory, a situation seems to exist that is similar to that described above for nonlinear balance: the most common flow diagnostic is neither universally appropriate nor the most general criterion available. Inertial instability, whether occurring in the equatorial middle atmosphere or in connection with the Asian monsoon, takes place in regions where frictional and/or divergent processes may not be ignorable. Yet the theoretical criterion for inertial instability that is commonly applied to observational datasets, usually (6), is derived for idealized conditions that do ignore these processes.

Therefore, the implication for research in inertial instability theory is to follow the method outlined in this

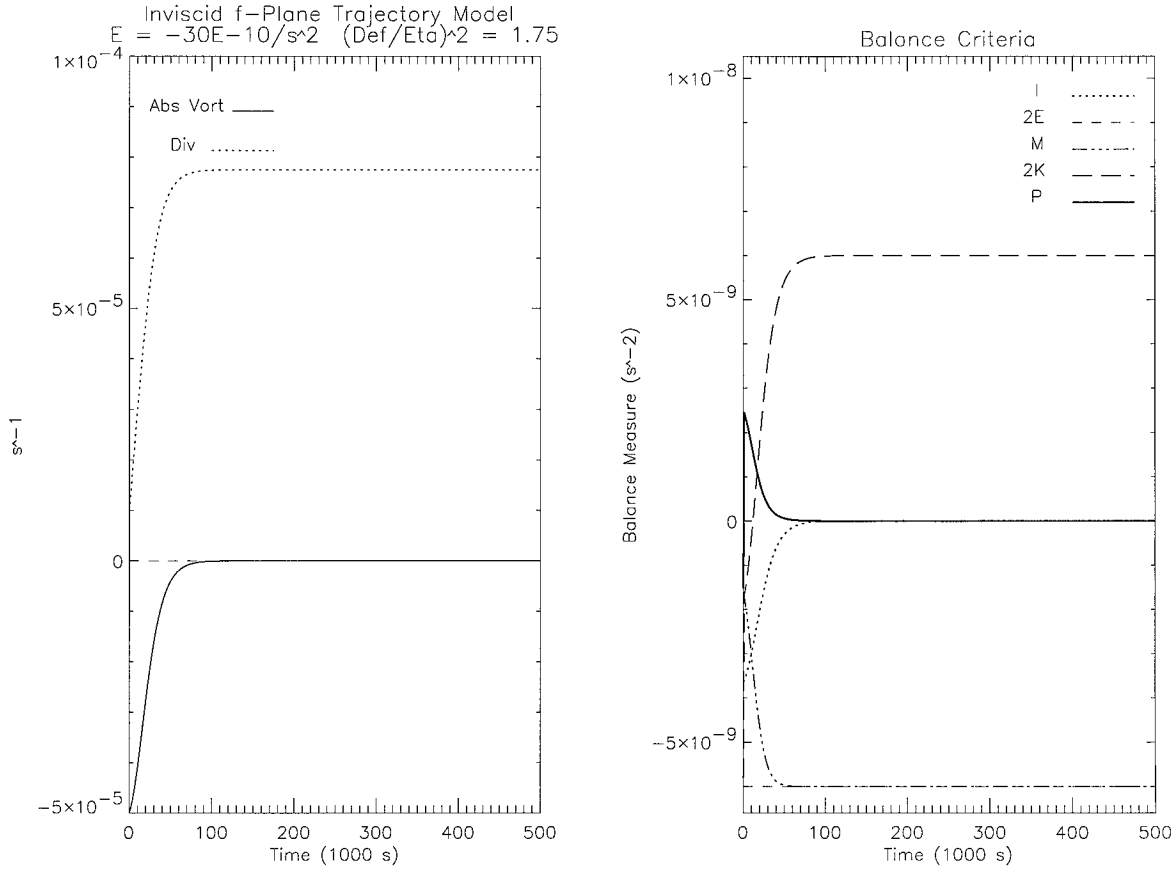

FIG. 10. As in Fig. 5 but for nonelliptic and strongly inertially unstable flow, as described in Table 3.

paper: to begin with the primitive equations and derive generalized criteria that are more widely applicable to observations than the idealized theory. This approach to inertial instability theory is described in detail in Knox (1996). Although the two approaches are not identical—it is more appropriate for inertial instability theory to derive an equation for the second derivative of the velocity, rather than the material derivative of the divergence as in nonlinear balance theory—similar terms involving the frictional and divergent terms appear in each approach. It is found that the frictional term representing gravity wave drag is a significant part of the dynamics of inertial instability in the equatorial lower mesosphere. This result echoes Kasahara's (1982) and Randel's (1987) speculations concerning the role of unresolved motions in creating regions of nonellipticity (see section 4). The reader is directed to Knox (1996) for further details.

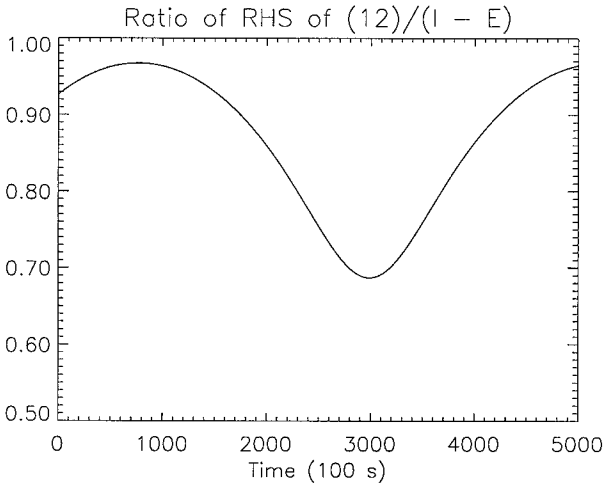

FIG. 11. Ratio of $(f^2 - \zeta^2)/(I - E)$ for the strongly anticyclonic flow model simulation in Fig. 8.

9. Discussion

Theory, observations, and nonlinear modeling results have been used to clarify and expand our understanding of nonlinear and inertial balance. A hierarchy of measures of balance have been derived that range from inviscid *f*-plane quantities to criteria appropriate for the full primitive equations. New analytic results of note are relationships between the inertial stability criterion and the ellipticity criterion for gradient flow, a relationship between Petterssen's (1953) and Kasahara's

(1982) realizability criteria, and Table 2, which relates various generalized nonlinear balance criteria to the inertial stability criterion. The numerical modeling approach extends the analytical work of Paegle and Paegle (1976b) and also confirms and extends the theoretical results presented herein.

This work suggests several new avenues for research. The most obvious future direction is to establish the usefulness of these generalized balance criteria in observational data and model analyses. Work is in progress to demonstrate that the criteria can be used as diagnostics that provide a detailed quantitative interpretation of the ''balancedness'' of complex flows. In addition, such work should also shed light on the physical meaning of violations of the various balance criteria. From the perspective of inertial instability research, numerical simulations could reveal the unexplored connections between the instability and violations of various definitions of balanced flow. In summary, it is hoped that this work provides future researchers with a hierarchy of readily accessible tools for use in the diagnosis of complex large-scale flows.

Acknowledgments. This work comprised a portion of the author's Ph.D. dissertation at the University of Wisconsin—Madison and was supported by NASA Grant NAG5-2806. I thank Akira Kasahara for his insightful review. Thanks also go to Matt Hitchman and David Ortland for reading early drafts of this paper, and to John Strikwerda for insights into the mathematical theory of ellipticity. Stephen Jascourt, Jonathan Martin, and Steve Silberberg provided many helpful comments on the subtleties of nonlinear balanced flows and inertial oscillations. Keiko Yumi expertly translated the Syō no article.

APPENDIX

Derivation of the Ellipticity–Inertial Stability Relation (12)

From (11), we know that

$$f_0 I^* > f_0^2 - 4\zeta_{curv}\zeta_{shear} \tag{A1}$$

is equivalent to the ellipticity condition (8) for gradient flow conditions. Notice that the rhs of (A1) is minimized when ζ_{curv} and ζ_{shear} are the same sign, for example, case A in section 2. Specifically, the rhs of (A1) is at its nadir whenever

$$\zeta_{curv} = \zeta_{shear}. \tag{A2}$$

Assuming (A2), (A1) becomes

$$f_0 I^* > f_0^2 - 4\zeta_{curv}^2. \tag{A3}$$

Equation (A3) can then be rewritten as

$$f_0 I^* > f_0^2 - (\zeta_{curv}^2 + 2\zeta_{curv}^2 + \zeta_{curv}^2). \tag{A4}$$

Invoking (A2) once again, (A4) can be expressed as

$$f_0 I^* > f_0^2 - (\zeta_{curv}^2 + 2\zeta_{curv}\zeta_{shear} + \zeta_{shear}^2), \tag{A5}$$

which is exactly equivalent to

$$f_0 I^* > f_0^2 - (\zeta_{curv} + \zeta_{shear})^2. \tag{A6}$$

Since $\zeta = \zeta_{curv} + \zeta_{shear}$ for gradient flow, (A6) is simply

$$f_0 I^* > f_0^2 - \zeta^2, \tag{A7}$$

which completes the derivation.

REFERENCES

Alaka, M., 1961: The occurrence of anomalous winds and their significance. *Mon. Wea. Rev.,* **89,** 482–494.

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

Angell, J. K., 1962: The influence of inertial instability upon transosonde trajectories and some forecast implications. *Mon. Wea. Rev.,* **90,** 245–251.

Árnason, G., 1958: A convergent method for solving the balance equation. *J. Meteor.,* **15,** 220–225.

Bell, G. D., and D. Keyser, 1993: Shear and curvature vorticity and potential-vorticity interchanges: Interpretation and application to a cutoff cyclone event. *Mon. Wea. Rev.,* **121,** 76–102.

Blumen, W., and W. M. Washington, 1969: The effect of horizontal shear flow on geostrophic adjustment in a barotropic fluid. *Tellus,* **21,** 167–176.

Bolin, B., 1956: An improved barotropic model and some aspects of using the balance equation for three-dimensional flow. *Tellus,* **8,** 61–75.

Charney, J. G., 1962: Integration of the primitive and balance equations. *Proc. Int. Symp. Numerical Weather Prediction,* Tokyo, Japan, Meteor. Soc. Japan, 131–152.

Ciesielski, P. E., D. E. Stevens, R. H. Johnson, and K. R. Dean, 1989: Observational evidence for asymmetric inertial instability. *J. Atmos. Sci.,* **46,** 817–831.

Cressman, G., 1959: An operational objective analysis system. *Mon. Wea. Rev.,* **87,** 367–374.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

Ghil, M., B. Shkoller, and V. Yangarber, 1977: A balanced diagnostic system compatible with a barotropic prognostic model. *Mon. Wea. Rev.,* **105,** 1223–1237.

Haltiner, G. J., and R. T. Williams, 1980: *Numerical Prediction and Dynamic Meteorology.* John Wiley and Sons, 477 pp.

Hitchman, M. H., and C. B. Leovy, 1986: Evolution of the zonal mean state in the equatorial middle atmosphere during October 1978–May 1979. *J. Atmos. Sci.,* **43,** 3159–3176.

——, ——, J. C. Gille, and P. L. Bailey, 1987: Quasi-stationary zonally asymmetric circulations in the equatorial lower mesosphere. *J. Atmos. Sci.,* **44,** 2219–2236.

Holton, J. R., 1983: The influence of gravity wave breaking on the general circulation of the middle atmosphere. *J. Atmos. Sci.,* **40,** 2497–2507.

——, 1992: *An Introduction to Dynamic Meteorology.* Academic Press, 511 pp.

Hoskins, B. J., 1974: The role of potential vorticity in symmetric stability and instability. *Quart. J. Roy. Meteor. Soc.,* **100,** 480–482.

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

——, H. H. Hsu, I. N. James, M. Masutani, P. D. Sardeshmukh, and G. H. White, 1989: *Diagnostics of the Global Atmospheric Circulation Based on ECMWF Analyses 1979–1989.* World Meteor. Org., 217 pp.

Houghton, D. D., 1968: Derivation of the elliptic condition for the balance equation in spherical coordinates. *J. Atmos. Sci.,* **25,** 927–928.

Iversen, T., and T. E. Nordeng, 1982: A convergent method for solving the balance equation. *Mon. Wea. Rev.,* **110,** 1347–1353.

Kasahara, A., 1982: Significance of non-elliptic regions in balanced flows of the tropical atmosphere. *Mon. Wea. Rev.,* **110,** 1956–1967.

Knox, J. A., 1996: A theoretical and observational study of inertial instability and nonlinear balance. Ph.D. dissertation, University of Wisconsin—Madison, 351 pp. [Available from University Microfilms, Inc., P.O. Box 1346, Ann Arbor, MI 48106.]

Kundu, P. K., 1990: *Fluid Mechanics.* Academic Press, 638 pp.

Leary, C., 1974: Comment on ''Anomalous gradient winds: Existence and implications.'' *Mon. Wea. Rev.,* **102,** 257–258.

MacDonald, A. E., 1977: On a type of strongly divergent steady state. *Mon. Wea. Rev.,* **105,** 771–785.

McWilliams, J., 1991: Geostrophic vortices. *Proceedings of the International School of Physics ''Enrico Fermi'' Course CIX,* A. Osborne, Ed., North-Holland, 7–50.

Miyakoda, K., 1956: On a method of solving the balance equation. *J. Meteor. Soc. Japan,* **34,** 364–367.

Newton, C. W., and E. Palmén, 1963: Kinematic and thermal properties of a large-amplitude wave in the westerlies. *Tellus,* **15,** 99–119.

Ortland, D. A., W. R. Skinner, P. B. Hays, M. D. Burrage, R. S. Lieberman, A. R. Marshall, and D. A. Gell, 1996: Measurements of stratospheric winds by the High Resolution Doppler Imager. *J. Geophys. Res.,* **101** (D), 10 351–10 363.

O'Sullivan, D. J., and M. H. Hitchman, 1992: Inertial instability and Rossby wave breaking in a numerical model. *J. Atmos. Sci.,* **49,** 991–1002.

Paegle, J., and J. N. Paegle, 1974: An efficient and accurate approximation of the balance wind with application to non-elliptic data. *Mon. Wea. Rev.,* **102,** 838–846.

——, and ——, 1976a: On geopotential data and ellipticity of the balance equation: A data study. *Mon. Wea. Rev.,* **104,** 1279–1288.

——, and ——, 1976b: On the realizability of strongly divergent supergradient flows. *J. Atmos. Sci.,* **33,** 2300–2307.

Panchev, S., 1985: *Dynamic Meteorology.* D. Reidel, 360 pp.

Petterssen, S., 1953: On the relation between vorticity, deformation and divergence and the configuration of the pressure field. *Tellus,* **5,** 231–237.

Randel, W. J., 1987: The evaluation of winds from geopotential height data in the stratosphere. *J. Atmos. Sci.,* **44,** 3097–3120.

Rayleigh, Lord, 1916: On the dynamics of revolving fluids. *Proc. Roy. Soc. London, Ser. A,* **93,** 447–453.

Reiter, E. R., 1961: The detailed structure of the wind field near the jet stream. *J. Meteor.,* **18,** 9–30.

——, 1963: *Jet-Stream Meteorology.* University of Chicago Press, 515 pp.

Sirovich, L., 1988: *Introduction to Applied Mathematics.* Springer-Verlag, 370 pp.

Solberg, H., 1936: Le mouvement d'inertie de l'atmosphere stable et son role dans la theorie des cyclones. *Sixth Assembly,* Edinburgh, United Kingdom, Union Geodesique et Geophysique Internationale, 66–82.

Stevens, D., and F. Crum, 1987: Dynamic meteorology. *Encyclopedia of Physical Science and Technology,* R. Meyers, Ed., Vol. 8, Academic Press, 227–260.

Syōno, S., 1948: On the mechanism of generation of cold wave (in Japanese). *J. Meteor. Soc. Japan,* **26,** 1–14.

Tan, Y.-C., and J. Curry, 1993: A diagnostic study of the evolution of an intense North American anticyclone during winter 1989. *Mon. Wea. Rev.,* **121,** 961–975.

Thorpe, A. J., H. Volkert, and D. Heimann, 1993: Potential vorticity of flow along the Alps. *J. Atmos. Sci.,* **50,** 1573–1590.

Tribbia, J., 1981: Nonlinear normal-mode balancing and the ellipticity condition. *Mon. Wea. Rev.,* **109,** 1751–1761.

University of Chicago Staff Members, 1947: On the general circulation of the atmosphere in middle latitudes. *Bull. Amer. Meteor. Soc.,* **28,** 255–280.

Vreugdenhil, C. B., 1994: *Numerical Methods for Shallow-Water Flow.* Kluwer, 261 pp.

Young, J. A., 1981: Low-level monsoon circulations. *Proc. Int. Conf. on Early Results of FGGE and Large-Scale Aspects of its Monsoon Experiments,* Tallahassee, FL. World Meteor. Org., 5-4–5-11.