The Next-Order Corrections to Quasigeostrophic Theory

DAVID J. MURAKI

Courant Institute, New York University, New York, New York

CHRIS SNYDER AND RICHARD ROTUNNO

National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 13 January 1998, in final form 12 June 1998)

ABSTRACT

Quasigeostrophic theory is an approximation of the primitive equations in which the dynamics of geostrophically balanced motions are described by the advection of potential vorticity. Quasigeostrophy also represents a leading-order theory in the sense that it is derivable from the full primitive equations in the asymptotic limit of zero Rossby number. Building upon quasigeostrophy, and the centrality of potential vorticity, a systematic asymptotic framework is developed from which balanced, next-order corrections in Rossby number are obtained. The simplicity of the approach is illustrated by explicit construction of the next-order corrections to a finite-amplitude Eady edge wave.

1. Introduction

Much of current understanding of midlatitude dynamics in both the atmosphere and the ocean arises from the quasigeostrophic theory first proposed by Charney (1948). Quasigeostrophy (QG) can be derived as the leading-order theory in an expansion of the primitive equations (PEs) for small Rossby number (Pedlosky 1987), which is defined as $\epsilon = V/fL$, where V and L are characteristic scales for horizontal velocity and length, and f is the Coriolis parameter. Quasigeostrophy is conceptually simple and is also qualitatively successful when applied to atmospheric and oceanic flows, particularly in diagnostic applications. [Interpretations of numerous atmospheric phenomena in terms of QG and further references are given in Bluestein (1992).] There are, however, known deficiencies, both quantitative and qualitative, of QG solutions (see, e.g., Mudrick 1974).

Since QG is the leading-order theory in an asymptotic expansion, it is natural to ask whether that expansion could be extended through additional orders in ϵ, and whether the aforementioned deficiencies of QG would thereby be reduced. Thus, our goal is to extend systematically, and to produce solutions of, the asymptotic expansion of the PE through additional orders in ϵ beyond QG. We are motivated in large part by a desire to refine the dynamical insight already available from QG, but we also seek a framework in which to address quantitatively more fundamental questions relating to the nature of balanced flows and their coupling to inertia-gravity waves.

Allen (1993) presents one approach to systematically extending QG. With pressure chosen as the distinguished variable in the theory (i.e., the variable required for initialization and from which all other variables may be diagnosed), an iterative solution technique is developed that yields an additional order of accuracy in ϵ per iteration. Vallis [1996; see also Warn et al. (1995) and Bokhove (1997)] proposes a different strategy in which potential vorticity is distinguished and dynamics is derived from its material conservation.

Our approach, which is similar to that of Vallis, builds on two basic elements of QG that lead to its exceptional simplicity and wide applicability: the material conservation of an analog of Ertel potential vorticity, and the simple relation between the flow variables and a single scalar potential (the geostrophic streamfunction or geopotential height). The key step in our approach (given in section 3) is to reformulate the PE by first including the conservation of potential vorticity as a prognostic equation, or equivalently, choosing potential vorticity as a distinguished variable; and second, introducing the change of dependent variables,

* The National Center for Atmospheric Research is supported by the National Science Foundation.

Corresponding author address: Dr. David J. Muraki, Courant Institute, New York University, New York, NY 10003.
E-mail: muraki@cims.nyu.edu

$$\begin{pmatrix} v \\ -u \\ \theta \end{pmatrix} = \nabla\Phi + \nabla \times \begin{pmatrix} F \\ G \\ 0 \end{pmatrix} = \begin{pmatrix} \Phi_x - G_z \\ \Phi_y + F_z \\ \Phi_z + G_x - F_y \end{pmatrix}, \quad (1)$$

where (u, v) is the horizontal velocity and θ is potential temperature. (Nondimensionalizations implicit in the above are given in section 2.) The equations that result from this substitution into the PE will be referred to as the QG$^+$ reformulation.

As will be shown in sections 3 and 4a, the QG$^+$ equations have the appealing characteristics that, for slow motions whose timescale is $O(\epsilon^{-1}f^{-1})$, (i) only a single time derivative appears at leading order, that of Φ or potential vorticity; (ii) QG follows directly from QG$^+$ by setting $\epsilon = 0$; and (iii) subsequent asymptotic corrections derive from a straightforward expansion of Φ, F, and G in powers of ϵ. In the opposite limit of fast motions whose timescale is $O(f^{-1})$, it is the time derivatives of F and G that appear at leading order of QG$^+$, and the inertia-gravity wave dispersion relation is recovered. Boundary conditions also require careful attention because of the change of dependent variables; we specify a well-posed initial boundary value problem for the expansion to next order beyond QG.

While this asymptotic framework is an important result of this paper, it is also only a first step, since solving equation sets that are accurate beyond QG is often more difficult than deriving them. Our purpose is thus not merely to put forward (yet another) approximation to the PE, but to find equation sets for which solutions are accessible both analytically and computationally and to evaluate those solutions relative to the PE. To that end, we present in section 5a an analytic solution for a finite-amplitude surface edge wave that is accurate through next order in ϵ. This solution illustrates the well-posedness of boundary conditions and the simplicity of the solution procedure for the resulting equations. Moreover, a subsequent paper will discuss the numerical implementation of this asymptotic framework and will present solutions for the development of a baroclinic wave on a jet, following Hoskins and West (1979). For $\epsilon = 0.3$ (based on the jet speed), the numerical solutions improve the qualitative deficiencies of QG solutions and have significantly increased quantitative accuracy at nondimensional times that are roughly $O(\epsilon^{-1})$.

The approach presented here has certain desirable properties. First, it builds upon the foundation of QG through a natural extension of its mathematical structure. To the extent that one is interested in improving the dynamical insight available from QG (which is arguably most of midlatitude dynamics), this is the obvious step. Second, the scheme is computationally and analytically straightforward; in its simplest form, only advection of PV and inversion of Laplacians are required. Because it builds upon QG, the scheme is easily implemented given an existing QG numerical model. Finally, the present approach suggests a clean framework for examining how balanced flows interact with gravity waves.

In section 2, the nondimensionalized PEs are presented and QG is discussed from the perspective of a reduced model for balanced dynamics. Section 3 then presents a natural reformulation of the PE that is based on the change of variables given in (1), which produces QG$^+$. Unlike the original PE, this reformulation reflects the structure of QG in the sense that QG can be obtained directly by setting $\epsilon = 0$ in the QG$^+$ equations. The equations for the next-order corrections beyond QG, referred to as QG^{+1}, are derived in section 4. The well-posedness of the QG^{+1} equation set is illustrated in section 5 by the analytical construction of the corrections to a finite-amplitude Eady edge wave. In the concluding section, we summarize and discuss further the relation of the present approach to existing balanced models.

2. Primitive equations and the QG approximation

a. Primitive equations

We take as our starting point equations for an f-plane geophysical fluid system that is adiabatic, inviscid, Boussinesq, and hydrostatic (Gent and McWilliams 1983b):

$$u_x + v_y + w_z = 0$$

$$\frac{D}{Dt}\begin{pmatrix} u \\ v \end{pmatrix} + f\begin{pmatrix} -v \\ u \end{pmatrix} = -\begin{pmatrix} \phi^T_x \\ \phi^T_y \end{pmatrix}$$

$$\phi^T_z = \frac{g}{\theta_0}\theta^T$$

$$\frac{D\theta^T}{Dt} = 0, \quad (2)$$

where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z} \quad (3)$$

denotes the usual advective derivative. In the atmosphere u, v, and w represent the wind velocities, and θ^T represents potential temperature. Although the quantities ϕ^T and z will be referred to as pressure and vertical height, in a strict atmospheric sense, they represent geopotential height and a modified pressure coordinate (Hoskins and Bretherton 1972). The coefficients f, g, and θ_0 are the Coriolis parameter, the gravitational constant, and a reference potential temperature.

Reference profiles of θ^T and ϕ^T are defined by a constant, horizontally uniform stratification,

$$\frac{g}{\theta_0}\theta^M \equiv N^2 z, \qquad \phi^M \equiv \frac{1}{2}N^2 z^2, \quad (4)$$

where N is the (constant) Brunt–Väisälä frequency. The primitive equations as written below will model disturbance flows to this continuously stratified atmosphere.

Nondimensionalization of the disturbance equations is based upon typical midlatitude synoptic scales where the Burger number (B) is taken to be unity, and the Rossby number (ϵ) is assumed small where

$$B \equiv \left(\frac{NH}{fL}\right)^2 = 1, \qquad \epsilon \equiv \frac{V}{fL} \ll 1 \tag{5}$$

relates the scales for the height H, horizontal distance L, and horizontal velocity V. The primitive variables are scaled by

$$\begin{aligned} x, y &\sim L \qquad z \sim H \\ t &\sim (\epsilon f)^{-1} \\ u, v &\sim V \qquad w \sim \epsilon^2 fH \\ \phi = \phi^T - \phi^M &\sim \epsilon N^2 H^2 \\ \frac{g}{\theta_0}\theta = \frac{g}{\theta_0}(\theta^T - \theta^M) &\sim \epsilon N^2 H. \end{aligned} \tag{6}$$

These scaling assumptions give the nondimensional PE for continuity,

$$u_x + v_y + \epsilon w_z = 0; \tag{7}$$

horizontal momentum and hydrostatic balance,

$$\begin{aligned} \epsilon\left(\frac{Du}{Dt}\right) - v &= -\phi_x \\ \epsilon\left(\frac{Dv}{Dt}\right) + u &= -\phi_y \\ 0 &= -\phi_z + \theta; \end{aligned} \tag{8}$$

temperature advection,

$$\left(\frac{D\theta}{Dt}\right) + w = 0; \tag{9}$$

and material derivative,

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + \epsilon w\frac{\partial}{\partial z}. \tag{10}$$

Because of the scaling for w in (6), which was chosen in anticipation of QG, the vertical velocity does not enter at leading order in either (7) or (10).

Boundary conditions are also crucial; they must, for example, be implemented correctly for asymptotically consistent balance models (Gent and McWilliams 1983a). The simplest geophysically relevant geometry is a horizontally periodic cell bounded by rigid horizontal surfaces at $z^s = 0, 1$. The boundary conditions are zero normal flow ($w = 0$) on the bounding surfaces and horizontal periodicity. This simple geometry will be assumed in what follows; the additional complexities of horizontally bounded domains (ocean basins) and weak surface topography are deferred to the appendixes.

The primitive equations contain two exact conservation laws,

$$\frac{D\theta^T}{Dt} = 0, \qquad \frac{DQ}{Dt} = 0, \tag{11}$$

where the first is total (nondimensionalized) potential temperature $\theta^T = z + \epsilon\theta$, and the second is the Ertel potential vorticity Q. In the hydrostatic limit, the definition of the Ertel potential vorticity contains only the horizontal winds

$$\mathbf{u}_H \equiv \begin{pmatrix} u \\ v \\ 0 \end{pmatrix} \tag{12}$$

so that

$$\begin{aligned} Q &\equiv (\mathbf{k} + \epsilon\boldsymbol{\nabla} \times \mathbf{u}_H) \cdot \boldsymbol{\nabla}\theta^T \\ &= \boldsymbol{\nabla} \cdot [\theta^T(\mathbf{k} + \epsilon\boldsymbol{\nabla} \times \mathbf{u}_H)]. \end{aligned} \tag{13}$$

In the midlatitude scaling, the uniform stratification in θ^T imparts a nominal unity value to potential vorticity, while the dynamically active contribution is of smaller order in Rossby number. This motivates the introduction of a disturbance potential vorticity

$$\begin{aligned} q \equiv (v_x - u_y + \theta_z) \\ + \epsilon[(v_x - u_y)\theta_z - v_z\theta_x + u_z\theta_y] \end{aligned} \tag{14}$$

so that $Q = 1 + \epsilon q$. Unless specified otherwise, the term potential vorticity (PV) will hereafter be used in conjunction with this disturbance value.

An important consequence of the exact advection of PV is that its volume integral (in finite, or periodic, domains) is a conserved quantity of the PE dynamics. Since Q is a perfect divergence derived from primitive variables, a consistency requirement can be derived from the application of a Greens identity to the volume integral of (13). For the periodic-cell geometry,

$$\begin{aligned} \int\int\int q\, dx\, dy\, dz \\ = \int\int [\theta + \epsilon\theta(v_x - u_y)]\Big|_{z=0}^{z=1} dx\, dy. \end{aligned} \tag{15}$$

Within the context of the PE model, this integral is a consistency property that q inherits from the flow; in PV-based balance models, this integral becomes a solvability requirement in the determination of a balanced state.

b. QG revisited

We next present the QG approximation to the PE (7)–(9) from a perspective that identifies the basic mathematical structure underlying the QG balance and dynamics. This structure will suggest a natural reformulation of the PE, referred to as QG$^+$, which then serves as the starting point for the systematic expansion in Rossby number of section 3.

Because the notion of ''balance'' will appear frequently in what follows, it is worth fixing terminology before proceeding. As a system of partial differential equations, the

PEs are third order in time so that three initial conditions (for instance, u, v, and θ) are required to determine the dynamics. The necessity for three initial conditions is consistent with the fact that, when linearized about a resting basic state, the PEs support three independent modes—a Rossby wave and two inertia-gravity waves. When QG is formally valid ($\epsilon \ll 1$ and $t \ll 1/\epsilon$), there is a separation of timescales between the slow Rossby wave and the fast gravity waves. In the following, ''balanced flows'' are simply taken as asymptotic solutions of the PE that are devoid of gravity waves up to the order of the expansion; and ''balanced models'' represent reductions of the PE that describe only these asymptotic solutions and have only a single independent time derivative. Balanced models consist (implicitly, at least) of two elements (Warn et al. 1995): (a) ''balance relations'' defining conditions free of (fast) gravity waves and (b) ''balanced dynamics'' describing the (slow) evolution in time of the balanced flow. We will adhere to the above distinction in our exposition of QG, QG$^+$, and QG^{+1}. Further discussion of the notion of balance can be found in Warn et al. (1995) (M. E. McIntyre and W. A. Norton 1994, personal communication).

Quasigeostrophy is the leading-order (in ϵ) approximation to the PE; its balance relations are obtained by setting $\epsilon = 0$ in the definition of PV (14) and the momentum equations (8):

(a1) PV and surface temperature,

$$\nabla^2\phi^0 = q, \qquad (\phi^0_z)^s = \theta^s; \quad \text{and} \tag{16}$$

(a2) primitive variables,

$$\begin{aligned} v^0 &= \phi^0_x \\ -u^0 &= \phi^0_y \\ \theta^0 &= \phi^0_z, \end{aligned} \tag{17}$$

where ∇^2 denotes a three-dimensional Laplacian. The superscript 0 indicates the leading-order QG values, and the superscript s in (16) denotes evaluation at either the top or bottom horizontal surfaces ($z^s = 0, 1$).

The QG dynamics follows from the material conservation of PV [(11)] and θ^s, and from the truncation of the material derivative to advection only by the QG horizontal winds (u^0, v^0):

(b1) QG advection,

$$\mathcal{D}^0 \equiv \frac{\partial}{\partial t} + u^0\frac{\partial}{\partial x} + v^0\frac{\partial}{\partial y}; \quad \text{and} \tag{18}$$

(b2) advection dynamics,

$$\mathcal{D}^0 q = 0, \qquad (\mathcal{D}^0\theta)^s = 0. \tag{19}$$

The conservation of θ^s by the horizontal flow follows from the requirement that $w^s = 0$.

Thus, QG has the following interrelated properties. First, knowledge of the PV and surface potential temperature determine, through (a1), the pressure ϕ^0 and, through (a2), the other primitive variables. Of particular significance to our development of QG$^+$ is the identification of (v^0, $-u^0$, θ^0) with the three-dimensional gradient ϕ^0. Second, the fundamental dynamical variables are PV and surface potential temperature, in that their conservation determines the evolution of the flow. Third, QG possesses only a single interior time derivative (q) and, more importantly, only a single unconstrained initial condition (ϕ^0). The importance of the first two properties has been argued eloquently by Hoskins et al. (1985), and that of the third property by Warn et al. (1995). All of these aspects of the QG scheme will be incorporated into the subsequent reformulation and asymptotic expansion of the PE.

Quasigeostrophy also inherits a version of the PE integral property [(15)]. In (a1), a linear Poisson problem must be solved for ϕ^0 given q and θ^s; a solution exists only when

$$\int\int\int q\, dx\, dy\, dz = \int\int \theta\Big|_{z=0}^{z=1} dx\, dy. \tag{20}$$

Thus, while every pressure ϕ^0 state determines a unique q and θ^s, the reverse is not necessarily true. It is important to understand that this solvability condition is not specifically a consequence of the QG approximation, nor of the assumption of balance; but rather that, as a leading-order solution to the PE model, the QG solution must conform to the leading-order part of (15).

Although the vertical velocity w^0 completely decouples from the QG balance model, it can be included among the balance relations (a2). Note that w can be obtained either by diagnosing the time-derivative of temperature [(9)] or by inverting the ''omega equation'' (Gill 1982)

$$\nabla^2 w_0 = 2[J(\phi^0_z, \phi^0_x)]_x + 2[J(\phi^0_z, \phi^0_y)]_y, \tag{21}$$

where the $J(f, g) \equiv (f_x g_y - f_y g_x)$ denotes the Jacobian derivative. Boundary conditions of $w^s = 0$ and lateral periodicity are sufficient to permit inversion for w from (21).

As noted in Pedlosky (1987, section 6.3), the derivation of QG from PE is somewhat subtle. That is, directly setting $\epsilon = 0$ in PE (7)–(9) results in a loss of prognostic time derivatives, the familiar geostrophic degeneracy. The above derivation of QG avoids the subtleties of the degeneracy by directly imposing conservation of PV (19) as a resolution of the dynamics. Our systematic derivation of higher-order corrections is also greatly simplified by recognizing from the outset the central role of PV conservation in the dynamics. Thus, the following treatment of a next-order theory begins from a reformulation of the PE, which reflects the structure of the QG equations and retains PV as the key dynamical variable.

3. The QG$^+$ system

We wish to develop a complete and more accurate theory of PE dynamics, one which preserves and builds

on the simple mathematical structure of QG. The development proceeds in two steps. In this section, the PEs are rewritten in a form, which we will call QG$^+$, that reflects the structure of QG. The foundation of QG$^+$ is the representation of the primitive variables (v, $-u$, θ) in terms of a set of potentials (section 3a) that exploits the underlying three-dimensional gradient structure of the QG balance relations (17). Since QG$^+$ involves no approximations, its solutions also include inertia-gravity waves, which can be extracted by an appropriate rescaling (section 3c). The second step in the development, to be presented in section 4, is to refine QG by extending the balanced dynamics to higher orders in ϵ. The required asymptotic expansion in ϵ is straightforward and systematic because of the form of QG$^+$.

a. The potential representation

The natural three-dimensional gradient structure of QG balance relations (17) emerges from a rewriting of the PE into the gradient-organized form

$$\begin{aligned} v &= \phi_x + \epsilon\left(\frac{Du}{Dt}\right) \\ -u &= \phi_y + \epsilon\left(\frac{Dv}{Dt}\right) \\ \theta &= \phi_z. \end{aligned} \tag{22}$$

It is clear that the $O(\epsilon)$ terms in (22) are unlikely to preserve the purely gradient structure of the QG representation (17) at next order. The QG$^+$ reformulation thus begins from the most general representation of the primitive quantities as a *three-dimensional* Helmholtz decomposition,

$$\begin{pmatrix} v \\ -u \\ \theta \end{pmatrix} = \nabla\Phi + \nabla \times \Psi, \tag{23}$$

where Φ is a gradient potential and and Ψ is a three-component curl potential. Since only three independent potential functions are necessary, the z component of Ψ can be taken to be zero, thus defining the QG$^+$ near-gradient representation:

$$\begin{pmatrix} v \\ -u \\ \theta \end{pmatrix} = \nabla\Phi + \nabla \times \begin{pmatrix} F \\ G \\ 0 \end{pmatrix} = \begin{pmatrix} \Phi_x - G_z \\ \Phi_y + F_z \\ \Phi_z + G_x - F_y \end{pmatrix}. \tag{24}$$

Asymptotic consistency with QG requires that Φ be the QG pressure ϕ^0 at leading order, and that F and G be $O(\epsilon)$.

It is important to note that the potentials in (24) need not be unique. This is also the case in QG, where ϕ^0 is determined only up to an arbitrary constant. The nonuniqueness is a more subtle issue for QG$^+$, since the zero vector has the nontrivial potential representation

$$\mathbf{0} = \nabla H_z + \nabla \times \begin{pmatrix} -H_y \\ H_x \\ 0 \end{pmatrix} \tag{25}$$

whenever $H(x, y, z)$ is a harmonic function ($\nabla^2 H = 0$). In the QG$^+$ scheme, this "harmonic ambiguity" allows flexibility in applying the physical boundary conditions to the three potentials.

Because the continuity equation (7), which relates vertical velocity to (u, v), is diagnostic, it is also easy to write w in terms of the new potentials. Substituting the form (24) for u and v into (7), a vertical integration yields

$$\epsilon w = F_x + G_y. \tag{26}$$

Horizontal divergence ($-w_z$) therefore is associated only with F and G.

As a final note on the near-gradient representation, the potentials of (24) can also be reconfigured into a three-dimensional vector velocity potential,

$$\begin{pmatrix} u \\ v \\ \epsilon w \end{pmatrix} = -\nabla \times \begin{pmatrix} G \\ -F \\ \Phi \end{pmatrix}, \qquad \theta = \nabla \cdot \begin{pmatrix} G \\ -F \\ \Phi \end{pmatrix}, \tag{27}$$

that satisfies continuity equation (7) and incorporates the QG structure through the divergence relation for potential temperature.

b. QG$^+$ reformulation of PE

We next write the PE in a form that reflects the structure of the QG balance relations (a1) and (a2), and balanced dynamics (b1) and (b2) using the near-gradient representation [(24)]. Although it is an unapproximated version of the PE, this reformulation will be termed QG$^+$ to suggest its relation to QG.

In QG$^+$ the three potentials are determined from three equations that reduce to Poisson problems when $\epsilon = 0$. The equations for the curl potentials are obtained by equating the curls of the near-gradient representation (24) and the gradient-organized PE (22) and making substitutions involving (26) and (9), which yields:
(a0$^+$) curl potentials,

$$\left.\begin{aligned} \nabla^2 F &= \epsilon\left[-\left(\frac{D\theta}{Dt}\right)_x + \left(\frac{Dv}{Dt}\right)_z\right] \\ \nabla^2 G &= \epsilon\left[-\left(\frac{D\theta}{Dt}\right)_y - \left(\frac{Du}{Dt}\right)_z\right] \end{aligned}\right\}, \qquad (F_x + G_y)^s = 0. \tag{28}$$

These equations are labeled (a0$^+$), rather than being incorporated into (a1$^+$) below, as they have no counterpart in the QG relations (a1) and (a2); they are, however, intimately related to the QG diagnostics for w as will be shown later in this section.

The surface boundary condition in (28) follows from

(26) and the requirement that $w^s = 0$. This single condition is insufficient to uniquely determine both curl potentials, but a convenient consequence of the harmonic ambiguity (25) is that one may choose any boundary conditions that yield unique F and G and that are consistent with $(F_x + G_y)^s = 0$.

In QG$^+$ the gradient potential Φ plays a role analogous to the QG pressure ϕ^0. Its determining equation is the definition of potential vorticity equation (14):

(a1$^+$) gradient potential,

$$\nabla^2\Phi = q - \epsilon\boldsymbol{\nabla}\cdot\{\theta(\boldsymbol{\nabla}\times\mathbf{u}_H)\},$$
$$(\Phi_z + G_x - F_y)^s = \theta^s. \qquad (29)$$

Note that the boundary condition for Φ depends on the boundary conditions chosen for F and G at $z^s = 0, 1$. In terms of the three potentials, the physical quantities are given by

(a2$^+$) primitive variables,

$$\begin{aligned} v &= \Phi_x - G_z \\ -u &= \Phi_y + F_z \\ \theta &= \Phi_z + G_x - F_y \\ \epsilon w &= F_x + G_y, \end{aligned} \qquad (30)$$

where these relations are simply restatements of the near-gradient representation (24) and continuity (26).

The QG$^+$ equation set is completed by the full three-dimensional advection of PV and its associated temperature boundary condition:

(b1$^+$) 3D advection,

$$\mathcal{D} \equiv \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + \epsilon w\frac{\partial}{\partial z}; \quad \text{and} \qquad (31)$$

(b2$^+$) advection dynamics,

$$\mathcal{D}q = 0, \qquad (\mathcal{D}\theta)^s = 0. \qquad (32)$$

It is again emphasized that the QG$^+$ equation set (28)–(32) is an unapproximated reformulation of the full PE (7)–(9) in terms of the potentials Φ, F, and G.

The pressure ϕ in QG$^+$ completely decouples from the dynamics. If desired, it can be recovered by a Poisson inversion as the deviation $(\phi - \Phi)$ from the gradient potential:

$$\nabla^2(\phi - \Phi) = \epsilon\left[\left(\frac{Du}{Dt}\right)_x + \left(\frac{Dv}{Dt}\right)_y\right],$$
$$[(\phi - \Phi)_z - G_x + F_y]^s = 0, \qquad (33)$$

which results from equating the divergences of (24) and (22).

What have we achieved by these manipulations? No approximations have been made in reformulating the PE into QG$^+$, but the structure of the equations now mimics QG. Indeed, QG can be derived simply by setting $\epsilon = 0$ in the QG$^+$ equations: (28) shows that $F = G = 0$, (29) and (30) reduce to the QG balance relations with the addition of $w = 0$, (32) becomes the QG dynamics, and (33) shows that $\Phi = \phi^0$. The intimate relation of the curl potentials F and G to the usual QG diagnostics for the ageostrophic flow is also apparent in the limit of small ϵ, as the right side of (28) becomes the components of the vector $\mathbf{Q}$, which in QG forces the ageostrophic circulation (Gill 1982; section 12.10) and whose divergence also appears as the forcing term of the QG omega equation (21).

c. Gravity waves in QG$^+$

Although the scaling and organization of the QG$^+$ equation set (28)–(32) are designed around the structure of QG balance, QG$^+$ remains an unapproximated reformulation of the original PE (7)–(9). The balanced motions are then characterized by solutions with $\Phi = O(1)$, $F, G = O(\epsilon)$ and which evolve on an $O(1)$ timescale implied by the nondimensionalization (6).

A different dynamics results from the assumption that solutions evolve on a fast timescale of $\tau = t/\epsilon$ and that $\Phi = O(\epsilon)$ and $F, G = O(1)$. Since τ derivatives scale $O(1/\epsilon)$ larger than t derivatives, the τ derivatives in the QG$^+$ equations for the curl potentials (28) appear at leading order and must be promoted to the left side:

$$\begin{aligned} \nabla^2 F + [(G_x - F_y)_x + G_{zz}]_\tau &= O(\epsilon) \\ \nabla^2 G + [(G_x - F_y)_y - F_{zz}]_\tau &= O(\epsilon), \end{aligned} \qquad (34)$$

where only nonlinear spatial derivatives remain on the $O(\epsilon)$.

Homogeneous $O(1)$ solutions of (34) for F, G are pure inertia-gravity wave modes. Fourier wave solutions of the form $\exp[i(kx + ly + mz - \omega\tau)]$ possess the linear dispersion relation

$$\omega^2 = 1 + \frac{k^2 + l^2}{m^2}, \qquad (35)$$

which is as expected for hydrostatic gravity waves in a rotating fluid (Gill 1982, section 8.4). The assumed absence of a balanced part to the flow (Φ, $q \ll F$, G) implies that only $O(\epsilon)$ variations of q are induced by these wave motions [(29)].

Together with the QG solutions that emerge from QG$^+$ by retaining slow time dependence and setting $\epsilon = 0$, these gravity wave solutions demonstrate that the near-gradient representation effects a natural separation of the solutions into predominantly balanced and gravity wave components through the decomposition into the gradient and curl potentials. That this separation is only strictly true for leading order inversions is evident by the coupling at next order in the QG$^+$ equations. Thus, in the expansion of QG$^+$ presented in the next section, ''filtering'' of gravity waves is not inherent in the near-gradient representation but rather is enforced through

the restriction that the time derivative (or, more generally, material derivative) of the solution be $O(1)$ under nondimensionalization. This choice naturally excludes the fast timescale homogeneous modes of the curl potentials and ensures that F, G remain $O(\epsilon)$ on timescales $t \ll O(1/\epsilon)$.

4. Next-order corrections to QG

a. Expansion of QG^+

A systematic perturbation theory for QG^+ results by representing the potentials as power series expansions in Rossby number:

$$\begin{aligned} \Phi(x, y, z, t) &\sim \Phi^0 + \epsilon\Phi^1 + \epsilon^2\Phi^2 + \ldots \\ F(x, y, z, t) &\sim \epsilon F^1 + \epsilon^2 F^2 + \ldots \\ G(x, y, z, t) &\sim \epsilon G^1 + \epsilon^2 G^2 + \ldots . \end{aligned} \tag{36}$$

Since we know that the leading-order theory for balance is QG and has $F^0 = G^0 = 0$, the form of the series is chosen accordingly.

Derivation of equations for (F^n, G^n, Φ^n) is now straightforward. The resulting equations through one order beyond QG are given in section 4b. In the iterative calculation of these series, the primitive variables in correction terms in $(a0^+)$ and $(a1^+)$ are replaced using the potential representations $(a2^+)$. The dynamical variables (q, θ^s), however, have not been expressed explicitly as a perturbation series. Allen (1993) and Warn et al. (1995) also advocate leaving the dynamical variable unexpanded in similar asymptotic schemes. Such an approach is both natural and convenient; in computational implementations, for example, it is simplest and asymptotically correct to advect the full (q, θ^s) by the winds correct to the desired order. That is, the asymptotic correctness of balanced models resulting from (36) is just the order to which balanced winds are obtained for the advection of (q, θ^s). Expansions for q and θ^s can be introduced when necessary, as in the analytic solution for a surface edge wave in section 5a or the application of solvability conditions below.

The restriction of the dynamics to balanced motions requires not only the form of the series expansion in (36), but also the assumption that material derivatives in $(a0^+)$ respect an $O(1)$ scaling. Thus, the material derivatives appearing in the balance relation $(a0^+)$ (and possibly in lateral boundary conditions) can be diagnosed from lower-order terms. If initial conditions or the subsequent evolution of the flow cause these material derivatives to become large, diagnosing them is no longer asymptotically valid. This is typically the case after long times $t \sim O(\epsilon)$. This condition signals the breakdown of ordering of the perturbation series (36) and the potential involvement of gravity waves.

It remains to verify that the linear elliptic problems resulting from the iterated solution of $(a0^+)$ and $(a1^+)$ are well posed at each order. This is an important technical issue that involves boundary and solvability conditions. We consider here only the horizontally periodic case, while appendix B shows through a generalization of the arguments of McWilliams (1977) that the case of lateral walls, though more complicated, also yields well-posed problems.

In a horizontally periodic cell, the only requirement on the bounding surface for F and G is $w^s = (F_x + G_y)^s = 0$. As noted earlier, conditions at $z^s = 0, 1$ for F or G individually are arbitrary and can be chosen for convenience, as long as they are consistent with $w^s = 0$. A particularly good choice for computations is

$$F^s = 0, \qquad G^s = 0, \tag{37}$$

which avoids solvability issues arising from purely derivative boundary conditions and insures that the Poisson problems for F and G are well posed at each order.

The PV inversion for Φ in (29) is assured provided that the PE integral constraint of (15) is satisfied to the order of the expansion. Two issues arise in relation to this solvability condition. The first is the initialization issue, which demands that the initial conditions be consistent with the above integral constraint, as well as balanced, up to the desired order of accuracy of the perturbation scheme. The second is the inversion issue, which arises at subsequent times when the inversion for Φ is done at the level of a perturbation expansion, where solvable Poisson problems must be posed at each order (up to the initialized accuracy).

A construction for initializing a higher-order balanced flow is to specify $\Phi^0(t = 0)$, which thus provides consistent q^0 and θ^{0s}. Further corrections (F^n, G^n, Φ^n) can be generated from the initial Φ^0 provided the the potential vorticity q is adjusted by an $O(\epsilon^n)$ constant to ensure the existence of each $\Phi^n(t = 0)$ as iterated solutions of (29). This adjusted q is then a consistent value for PV to higher order in ϵ.

The solvability condition on the PV inversion becomes more subtle at subsequent times. In the full PE, the integral constraint of (15) is satisfied exactly at all times. For a perturbation solution truncated to $O(\epsilon^N)$, the integral constraint will only be maintained up to the order of truncation, and small $O(\epsilon^{N+1})$ discrepancies are to be expected.

Inversion for $\Phi = \Phi^0 + \ldots + \epsilon^N \Phi^N$ thus requires modifications of the given q and θ^s by at most $O(\epsilon^{N+1})$, consistent with the asymptotic correctness of the solution. In practice, however, inversions for each Φ^n are done order by order and invertibility for Φ^n will require a temporary partitioning of the solvability condition that resembles ϵ expansions of q and θ^s. For example, when the full q and θ^s is used in the inversion for Φ^0 using (29) at leading order, it is unlikely that the leading-order part of the solvability integral (15) will be exactly satisfied. Fortunately, this error will be at most $O(\epsilon)$ and the integral can be satisfied by an $O(\epsilon)$-close choice of q^0 and θ^{0s}, with the $O(\epsilon)$ adjustments $q - q^0$, and $\theta^s - \theta^{0s}$ being deferred to the Φ^1 inversion. The deferring

of successively smaller violations of solvability can then be continued up to the initialized order of accuracy, since solving for the full $\Phi = \Phi^0 + \ldots + \epsilon^N \Phi^N$ requires net adjustments no larger than $O(\epsilon^{N+1})$ to q and θ^s.

In short, both the initialization and inversion issues in an $O(\epsilon^N)$-correct expansion can be resolved, *at any specific time*, through an asymptotically ordered partitioning of the PV,

$$q \sim q^0 + \epsilon q^1 + \epsilon^2 q^2 + \ldots + \epsilon^N q^N + O(\epsilon^{N+1}), \quad (38)$$

which permits inversion for Φ^n at each level ($0 \le n \le N$). For convenience, the q^n partitions for $n > 0$ can be taken to be constants. If desired, θ^s can be similarly partitioned for convenience in satisfying the solvability constraint at each order. We emphasize that the full, unpartitioned q is used in the advection equation (b2$^+$); the partition [(38)] is merely a temporary expedient required for solvability of the order-by-order inversions for Φ^n.

These ideas are illustrated by the analytical construction of the next-order correction for a surface edge wave in section 5a. Results of a numerical implementation of the QG$^+$ approach will be presented in a separate article.

b. QG^{+1} equations

It is now a simple matter to derive equations that extend QG to an additional order of accuracy in ϵ; the resulting balanced model will be referred to as QG^{+1}. Substituting the series expansion (36) into the QG$^+$ equations (28)–(32) and omitting $O(\epsilon^2)$ terms yields

(a1^{+0}) QG pressure,

$$\nabla^2 \Phi^0 = q^0, \qquad \Phi^{0s}_z = \theta^{0s}; \quad (39)$$

(a0^{+1}) curl potentials,

$$\left.\begin{aligned} \nabla^2 F^1 &= 2J(\Phi^0_z, \Phi^0_x) \\ \nabla^2 G^1 &= 2J(\Phi^0_z, \Phi^0_y) \end{aligned}\right\}, \qquad (F^1_x + G^1_y)^s = 0; \quad (40)$$

(a1^{+1}) gradient potential,

$$\nabla^2 \Phi^1 = q^1 - [(\nabla^2 \Phi^0)\Phi^0_{zz} - |\nabla \Phi^0_z|^2], \qquad (\Phi^1_z + G^1_x - F^1_y)^s = \theta^{1s}; \quad (41)$$

(a2^{+1}) primitive variables,

$$\begin{aligned} v &\sim \Phi^0_x + \epsilon(\Phi^1_x - G^1_z) \\ -u &\sim \Phi^0_y + \epsilon(\Phi^1_y + F^1_z) \\ \theta &\sim \Phi^0_z + \epsilon(\Phi^1_z + G^1_x - F^1_y) \\ w &\sim F^1_x + G^1_y + O(\epsilon); \end{aligned} \quad (42)$$

(b1$^+$) 3D advection,

$$\mathcal{D} \equiv \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + \epsilon w \frac{\partial}{\partial z}; \quad \text{and} \quad (43)$$

(b2$^+$) advection dynamics,

$$\mathcal{D} q = 0, \qquad (\mathcal{D}\theta)^s = 0. \quad (44)$$

As discussed in the previous section, if properly initialized at $t = 0$, then at all later times a temporary partition of the the dynamical variables q, θ^s can be constructed

$$\begin{aligned} q &\sim q^0 + \epsilon q^1 + O(\epsilon^2) \\ \theta^s &\sim \theta^{0s} + \epsilon \theta^{1s} + O(\epsilon^2), \end{aligned} \quad (45)$$

which satisfy the first two orders of the solvability constraint (15)

$$\int\int\int q^0 \, dV = \int\int (\theta^{0s})\Big|_{z=0}^{z=1} dx\, dy \quad (46)$$

$$\int\int\int q^1 \, dV = \int\int [\theta^{1s} + \Phi^0_z(\Phi^0_{xx} + \Phi^0_{yy})]\Big|_{z=0}^{z=1} dx\, dy, \quad (47)$$

where the first integral guarantees the invertibility of (39) and the second (41). In computational implementations, the small $O(\epsilon^2)$ discrepancies in the partitioning of (45) are to be expected and can be neglected. We emphasize again that this partition is required only for the inversion of q for Φ_0 and Φ_1; in particular, the advection dynamics of (44) needs not incorporate (or produce an evolution of q that respects) the partition.

Only the leading order w is required in the advection of potential vorticity. In a horizontally periodic cell, it may be obtained without diagnosing a time derivative. With other boundary conditions, however, this is not generally possible (see appendix B).

5. The finite-amplitude Eady edge wave in QG^{+1}

a. Perturbation solution

Initial understanding of the QG^{+1} equations can be obtained through the direct construction of a next-order solution. Here, we present the corrections for a finite-amplitude surface edge wave, which serves as a demonstration of the iterative solution procedure and an illustration that the QG^{+1} equation set is analytically tractable.

An exact, nonlinear solution to the QG equations is the Eady edge wave (Gill 1982, section 13.2), which represents a traveling wave trapped against the surface $z = 0$. The solution superimposes a horizontally periodic traveling wave Φ^e onto y-independent geostrophic zonal flow with constant vertical shear:

$$\begin{aligned} \Phi^0 &\equiv -yz + \Phi^e \\ &= -yz + A \cos k(x - ct) \cos ly \, e^{-\lambda z}. \end{aligned} \quad (48)$$

The upper boundary is replaced by a decay condition as $z \to \infty$.

The simplest case has zero interior potential vorticity, $q^0 = 0$. The Laplacian of Φ^0 then vanishes, which yields the wavenumber condition

$$k^2 + l^2 = \lambda^2. \quad (49)$$

Furthermore, the edge wave trivially satisfies the portion of the QG dynamics [(19)] relating to interior PV. The requirement in (19) that $(\mathcal{D}^0\theta)^s = 0$ then determines the phase speed c of the wave; replacing the time derivative $(\partial/\partial t)$ by a horizontal derivative $(-c\partial/\partial x)$ results in

$$J(cy + \Phi^e, -y + \Phi^e_z) = 0 \quad \text{for} \quad z = 0, \tag{50}$$

which, by the exponential decay of Φ^e in the vertical, selects

$$c = 1/\lambda. \tag{51}$$

Although it is typical for parameter relations to require perturbative corrections, this proves unnecessary for the QG^{+1} edge wave so that both (49) and (51) are exact through next order.

Construction of a QG^{+1} edge wave solution requires that next-order corrections be found that also propagate with speed c. Substituting Φ_0 into (40) and making liberal use of $\nabla^2\Phi^e = 0$ and $\Phi^e_z = -\lambda\Phi^e$, the solutions for the curl potentials are

$$F^1 = -\frac{z}{\lambda}\Phi^e_{xx} - \frac{\lambda}{2}(\Phi^{e2})_y,$$

$$G^1 = -\frac{z}{\lambda}\Phi^e_{xy} + \frac{\lambda}{2}(\Phi^{e2})_x. \tag{52}$$

These imply

$$w^0 = F^1_x + G^1_y = \lambda z\Phi^e_x, \tag{53}$$

thus satisfying the boundary condition $w^s = 0$.

Substitution of the leading-order edge wave Φ^0 into the QG^{+1} equation for the scalar potential (41) gives

$$\nabla^2\Phi^1 = q^1 + [(\Phi^e_{xz})^2 + (\Phi^e_{yz} - 1)^2 + (\Phi^e_{zz})^2]. \tag{54}$$

Solvability of the Poisson equation requires that an $O(\epsilon)$ correction to the PV be introduced ($q^1 = -1$), so that the solution for Φ^1 is

$$\Phi^1 = -z\Phi^e_y + \frac{\lambda^2}{2}(\Phi^e)^2 + \tilde{\Phi}^1, \tag{55}$$

where $\tilde{\Phi}^1$ is a homogeneous (harmonic) solution to be determined by the requirement that the $O(\epsilon)$ dynamics is consistent with the assumed traveling wave solution. Since q is still spatially constant, its advection remains trivially satisfied. After some manipulation, the QG^{+1} equation (44) for θ^s can be written as

$$J\left(cy + \Phi^e, \left[\theta^1 + \lambda\Phi^1 + \Phi^e_y + \frac{\lambda^3}{2}(\Phi^e)^2\right]\right) = O(\epsilon), \quad \text{on } z = 0, \tag{56}$$

which certainly is satisfied if the second argument of the Jacobian is zero.

Replacing the temperature by the potential representation $\theta^1 = \Phi^1_z + G^1_x - F^1_y$ then gives the boundary condition for the homogeneous part of the scalar potential

$$\tilde{\Phi}^1_z + \lambda\tilde{\Phi}^1 = -\frac{\lambda}{2}\nabla^2_H(\Phi^{e2}) \quad \text{on} \quad z = 0, \tag{57}$$

where ∇^2_H is the horizontal Laplacian. Solving $\nabla^2\tilde{\Phi}^1 = 0$ subject to (57) shows that $\tilde{\Phi}^1$ consists of harmonics generated by the QG^{+1} perturbations

$$\tilde{\Phi}^1 = \frac{A^2}{2}\frac{k^2\lambda}{\lambda - 2k}\cos 2k(x - ct)\, e^{-2kz} + \frac{A^2}{2}\frac{l^2\lambda}{\lambda - 2l}\cos 2ly\, e^{-2lz} - \frac{A^2}{2}\lambda^2\cos 2k(x - ct)\cos 2ly\, e^{-2\lambda z} \tag{58}$$

and completes the specification of all three QG^{+1} potentials. The pressure correction can be inverted from (33) and is given by the expression

$$\phi^1 - \Phi^1 = -\frac{1}{4}\nabla^2_H(\Phi^{e2}) \tag{59}$$

so that the QG^{+1} edge wave pressure and temperatures have been determined.

b. Flow patterns

In the QG$^+$ representation, the potentials Φ, F, and G are not uniquely specified because of the harmonic ambiguity, so only the derived quantities (winds, pressure, etc.) are physically significant. Figures 1a and 2a show fields of pressure and potential temperature at the surface for the square QG edge wave (48) with ($k = l = 1$, $\lambda = \sqrt{2}$). Since $w = 0$ on $z = 0$, the isentropes are also streamlines for the wave-relative flow. In the wave-relative frame of reference, typical trajectories move from right to left.

The QG solutions have two distinct flow patterns, which are associated with a transition that occurs at a critical amplitude $A_c = 1/\lambda$. At small amplitudes, the streamlines are wavy perturbations to uniform flow, as shown in Fig. 1a for $A = 0.5\, A_c$. But above the critical amplitude (Fig. 2a, $A = 2.0A_c$), isolated extrema appear in potential temperature, indicating regions of fluid trapped within closed streamlines. These regions consist of cyclone–anticyclone pairs embedded within the wave. In either case, however, the QG solutions have symmetry between relatively warm and relatively cold air and between low and high pressure.

Figures 1b and 2b show the symmetry breaking by the QG^{+1} corrections, using the same parameters as the QG solutions and $\epsilon = 0.1$. Relative to the symmetric QG wave, extrema are now biased toward warm anomalies and low pressures, and regions of warm temperatures or low pressures (relative to a horizontal average) have shrunk in area. The asymmetry in the large-amplitude case is also manifest in the cyclone–anticyclone pairs, which are biased toward the cyclones. Not apparent from these surface plots is a northward tilt with

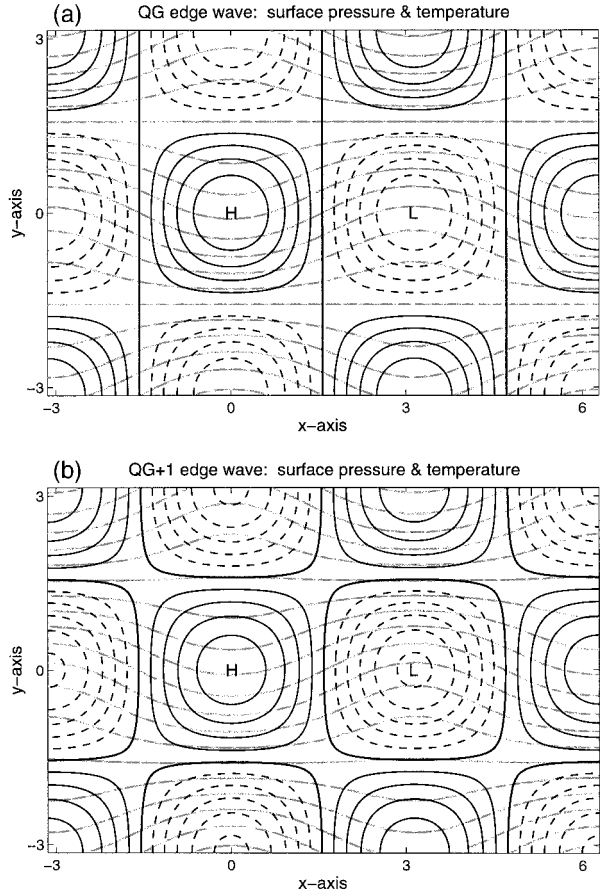

FIG. 1. Edge wave solutions for (a) QG and (b) QG^{+1}. The amplitude $A = A_c/2$ is below the critical amplitude at which closed streamlines and vortices appear. Surface pressure contours are shown in black at an interval of 0.2 with positive/zero/negative values indicated by solid/thick/dashed contours. Surface temperature contours are shown in gray at intervals of $\pi/2$. For these traveling wave solutions, the isotherms are also material streamlines with flow from right to left (in the eastward moving, wave-relative frame of reference).

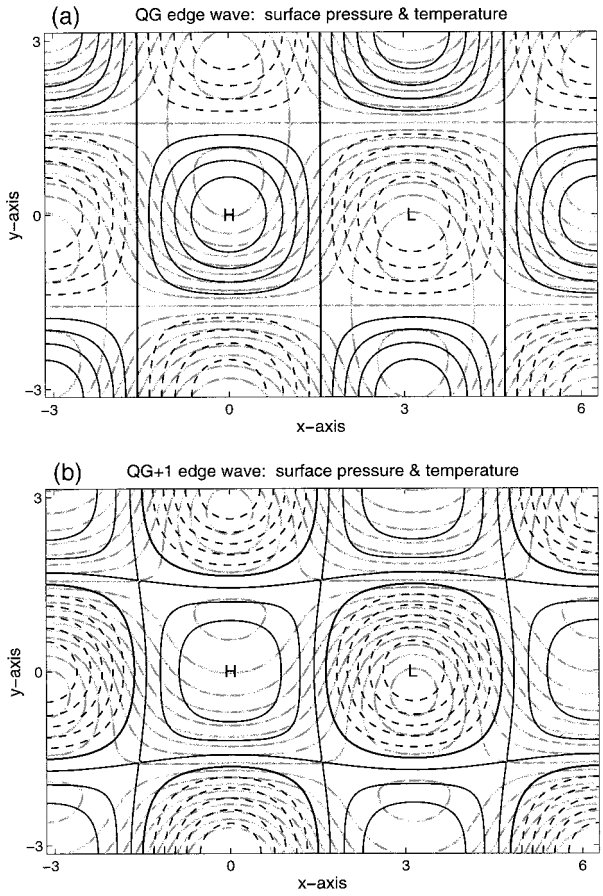

FIG. 2. As in Fig. 1 but for an amplitude $A = 2A_c$ sufficient to produce closed streamlines. Note the substantial asymmetries in (b) between regions of cyclonic and anticyclonic flow.

height of the QG^{+1} solution, which results from the first term in (55).

There is also an intermediate regime very close to the critical amplitude where the QG^{+1} waves possess only cyclonic vortices. An example is shown in Fig. 3 ($A = A_c + 0.005$).

Unfortunately, we have been unsuccessful in extracting a clear physical mechanism for the cyclonic asymmetry. Within the QG^{+1} edge wave solution, all of the asymmetries are associated with the harmonics generated by quadratic nonlinearities that appear in all three of the potentials. For pressure, there are three instances of nonlinear corrections: the particular solution part of Φ^1 (55), the homogeneous solution part of Φ^1 (58), and the pressure deviation (59). One indication of the complex interaction of these nonlinearities in the square edge wave is that only the homogeneous contribution (58) enhances the low pressure, while the others enhance the high.

Further evidence of the subtlety of the QG^{+1} effects is that certain elements of the symmetry breaking are reversed for elongated waves with either k/l or l/k larger than $\sqrt{3}$. In particular, cold anomalies and high pressures are enhanced, and closed streamlines form first within anticyclones as A is increased. This altered symmetry breaking is associated with a sign change in either of the denominators $\lambda - 2k$ or $\lambda - 2l$ in $\tilde{\Phi}^1$ (58). It is as yet unclear whether the resonance phenomenon that occurs when $\lambda = 2k$ or $\lambda = 2l$ reflects a realistic feature of atmospheric dynamics, or merely results from an accidental consequence of the traveling wave assumption.

Finally, the QG^{+1} solution can be checked in the limit case of $l = 0$ against the semigeostrophic solution, which is also accurate through $O(\epsilon)$ when $l = 0$ (Gent et al. 1994). With uniform PV, the semigeostrophic equations are equivalent to QG, but with x replaced by the geostrophic coordinate $X = x + \epsilon v = x + \epsilon\phi_x$ (Hoskins and Bretherton 1972). Thus, for the edge wave

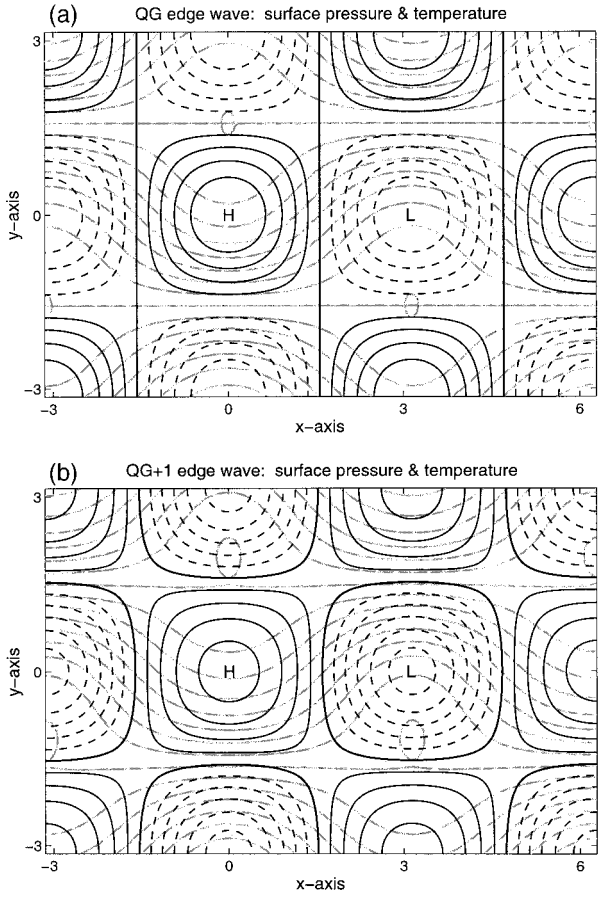

FIG. 3. The QG and QG^{+1} edge wave solutions for an amplitude ($A = A_c + 0.005$) close to the QG critical value. In the QG^{+1} case, closed streamlines exist only within cyclones.

with $l = 0$, the semigeostrophic solution gives the following implicit expression for the pressure gradient ϕ_x:

$$\phi_x = -kAe^{-kz}\sin k(X - ct)$$
$$= -kAe^{-kz}\sin k(x + \epsilon\phi_x - ct), \qquad (60)$$

which can be solved to $O(\epsilon)$ and integrated to

$$\phi = A\cos k(x - ct) + \frac{1}{4}\epsilon k^2A^2e^{-2kz}[1 - \cos 2k(x - ct)]. \qquad (61)$$

This is identical to the QG^{+1} result, using (58) and (59).

6. Summary and discussion

Quasigeostrophic theory can be derived as the leading order in an asymptotic expansion of the primitive equations for small Rossby number ϵ (Pedlosky 1987). Here we have presented a systematic scheme for extending such an expansion to arbitrary order in ϵ, with particular emphasis (section 4b) on the next-order theory beyond QG, which we term QG^{+1}. Our approach is to rewrite the PE in a form that anticipates the mathematical structure of QG. To this end, we replace [as in (1)] the primitive variables u, v, and θ with scalar potentials Φ, F, and G, where to leading order Φ is the geostrophic streamfunction and F and G determine the ageostrophic secondary circulation. Within this formulation, asymptotic expansion of the PE is straightforward.

Of course, ours is not the only approach to extending QG—numerous other balanced models provide an additional order of asymptotic accuracy in ϵ beyond QG. These include models based on decomposing the horizontal flow into nondivergent and irrotational contributions; approximate equation sets are then obtained by truncating the horizontal divergence equation and either the vertical vorticity or PV equations [see Lorenz (1960) and Gent and McWilliams (1983b) for the former, Charney (1955) and Raymond (1992) for the latter], or by approximating the horizontal momentum equations directly (Allen 1991; Xu 1994). Other approaches are based on neglecting time derivatives of the horizontal divergence (McIntyre and Norton 1994, personal communication), or attempt to preserve Hamiltonian structure in the approximate equations (Holm 1996), or involve, as here, an expansion in Rossby number (Snyder et al. 1991; Allen 1993; Vallis 1996a,b). All of these approaches agree by definition at $O(\epsilon)$ but differ at $O(\epsilon^2)$ owing to differences in their methods or in the form of the PE from which they begin.

The work of Allen (1993) and Allen and Newberger (1993) merits special mention. Together these constitute the first implementation of a systematic extension of QG to higher order in ϵ. In contrast to the formulation presented here, which is based on PV and its advection, Allen chooses pressure as the distinguished, unexpanded variable whose evolution governs the flow. Furthermore, this pressure inversion is formulated as an iterative perturbation procedure that offers a very compact computational implementation, as well as a clear recipe for stably diagnosing time derivatives.

Our approach has the virtue that it builds directly upon the foundation of QG. Because the mathematical structure has a direct relation to that of QG, we expect that, conceptually, the dynamics are also closely related to QG. Moreover, our approach results in equations that are computationally and analytically simple. Numerical solution is straightforward, as only inversions of Poisson equations are required, and existing code for a QG model can be easily modified to compute QG^{+1} solutions. Nontrivial solutions are also analytically accessible, as is illustrated here (section 5a) with a solution for a finite-amplitude surface edge wave that is accurate through an order in ϵ beyond QG. In comparison, obtaining solutions to many of the aforementioned balanced models is often difficult.

The QG^{+1} equations, however, lack other properties that are potentially attractive. First, the conservation relations of the PE, other than those deriving from material conservation of PV, are reproduced only through

the order of accuracy of the expansion. Many other balanced models, in contrast, are specifically tailored to preserve analogs of as many such conservation relations as possible. Second, QG^{+1} is not formally valid outside the regime of validity of QG given in (5), that is, outside of $\epsilon \ll 1$ and $B = O(1)$; yet models based on the smallness of the horizontal irrotational flow, such as Lorenz (1960) and Charney (1955), can retain accuracy outside that regime. If the flow is strongly stratified and the length scale is small compared to the radius of deformation, the Froude number, $F = V/NH$, is small and these models make errors that are $O(F)$ even when $\epsilon = O(1)$ (McWilliams 1985). For frontal flows, the appropriate small parameter is the ratio of cross-front to alongfront velocities (Hoskins and Bretherton 1972) and the same models have leading-order accuracy in an expansion in that ratio (Gent et al. 1994). Although we believe that the asymptotic principles presented here have broader application, it remains for future work to apply our approach to other geophysical regimes.

Acknowledgments. The authors thank Jim McWilliams and Jeff Whittaker for their insights into the subtleties of balance. DJM wishes to acknowledge the meteorological and social hospitality of the MMM group at NCAR Boulder over the past three summers. Additional thanks from DJM to the 1997 Woods Hole GFD Summer School for encouragement and many discussions related to the presentation of this work, especially Ed Spiegel, Rupert Ford, Joe Pedlosky, Rick Salmon, and Jack Whitehead. In this research, DJM has been supported by NSF DMR-9404374, DOE DE-FG02-88ER25053, and the Alfred P. Sloan Foundation.

APPENDIX A

Nonconstant Stratification

Inherent in our choice of the primitive equations (7)–(9) is a rest state atmosphere of (4) with constant stratification. This assumption can be relaxed by a nonconstant vertical profile, $s(z)$, of the rest state PV, and that also appears in the mean stratification

$$\frac{g}{\theta_0}\theta_z^M \equiv N^2 s(z), \tag{A1}$$

where $s(z) = 1$ recovers constant stratification. This modifies the PE model in the temperature equation

$$\left(\frac{D\theta}{Dt}\right) + ws(z) = 0, \tag{A2}$$

which in turn introduces factors of $s(z)$ throughout the QG^{+} formulation. The most significant change in the fluid mechanics is that $O(1)$ PV advection effects occur when $O(\epsilon)$ weak vertical flows advect the $O(\epsilon^{-1})$ strong gradient of stratification. The immediate impact to the QG analysis is that w is needed for the leading-order PV advection. This difficulty can be circumvented either by including vertical advection into the leading order dynamics (requires diagnosing w), or through a modified dynamical equation described by the advection of a ''pseudo-PV,'' which does not require w at leading order, and whose equation is presented here without derivation,

$$\frac{D}{Dt}\left[\frac{q}{s} + \left(\frac{1}{s}\right)_z \theta\right] - \epsilon w\left[\left(\frac{1}{s}\right)_z q + \left(\frac{1}{s}\right)_{zz} \theta\right] = 0, \tag{A3}$$

but essentially follows that in Pedlosky (1987, section 6.5).

APPENDIX B

Basin Boundary Conditions on Vertical Walls

For the sake of clarity, the discussion of lateral boundary conditions in the QG^{+1} model was limited to the simplest case of horizontal periodicity. The next simplest boundary conditions impose no normal flow at vertical walls, as they are typically used to enclose ocean basins. The many subtle issues encountered in implementing impermeable wall conditions for balance flow has been well addressed for QG by McWilliams (1977). This appendix summarizes the main ideas and the generalizations required when finding an implementable scheme for boundary conditions in the QG^{+} balance model.

By continuity, the curl potentials satisfy a divergence relation that becomes the boundary condition along a vertical wall:

$$(F_x + G_y)^w = \epsilon w = -\epsilon\left(\frac{D\theta}{Dt}\right)^w, \tag{B1}$$

where a superscript w denotes evaluation on the wall. Thus, for all correction orders, this boundary condition requires the diagnosing of a time derivative of the temperature from a previous order. Note that the same is true in a formulation obtaining w from the omega equation [(21)], so that the diagnosing of time derivatives is a general feature in the application of consistent boundary conditions. Another computational issue is that uncoupled boundary conditions between F and G are preferable as this avoids iteration in the Poisson inversions for F and G. This can sometimes be circumvented in simple geometries, one example being a rectangular basin where

$$F_x = \epsilon w, \qquad G = 0 \quad \text{on sides with constant } x, \quad \text{or}$$
$$G_y = \epsilon w, \qquad F = 0 \quad \text{on sides with constant } y \tag{B2}$$

satisfy the condition [(B1)].

The physical conditions of no normal flow are applied in the inversion for the gradient potential Φ. For strictly vertical walls, the orthogonality of the horizontal winds with the (unit, outward) normal to the boundary $\hat{\mathbf{n}}$ specifies the tangential derivative of Φ. At a fixed height,

z, boundary values of Φ are given by an arclength integral,

$$(\Phi)^w = -\int^w \begin{pmatrix} F_z \\ G_z \end{pmatrix} \cdot \hat{\mathbf{n}}\, ds + K^w(z, t), \qquad \text{(B3)}$$

where the path orientation is counterclockwise around exterior walls and clockwise around interior islands. The important result (Phillips 1954; McWilliams 1977) is that the constants of integration $K^w(z, t)$ at each height z are determined through the inclusion of an additional constraint—one derived from another exact PE integral property, the circulation Γ^w of the horizontal winds along the wall boundary (Pedlosky 1987, section 2.2)—

$$\Gamma^w(z, t) \equiv \int^w (u\, dx + v\, dy). \qquad \text{(B4)}$$

In QG this wall circulation is a constant in time, but QG$^+$ requires an additional dynamical equation:

(b3$^+$) wall circulation,

$$\Gamma^w_t = -\epsilon \int^w w(u_z\, dx + v_z\, dy). \qquad \text{(B5)}$$

In summary, the inversion for Φ with impermeable vertical walls introduces an additional unknown function $K^w(z, t)$ for each wall boundary as in (B3); the additional equation, which completes the well-posedness, is the circulation constraint of (B4).

The final consideration arising with vertical wall conditions is that the inversion for Φ still requires a solvability condition. It is again supplied by a PE consistency requirement that is obtained from the volume integral of potential vorticity (14):

$$\begin{aligned} &\iiint q\, dx\, dy\, dz \\ &\quad = \iint [\theta + \epsilon\theta(v_x - u_y)]\Big|_{z^s=0}^{z^s=1} dx\, dy \\ &\qquad + \sum_w \int_0^1 \left[\Gamma^w + \epsilon \int^w (\theta u_z\, dx + \theta v_z\, dy)\right] dz, \end{aligned} \qquad \text{(B6)}$$

where the sum denotes integral contributions from each wall boundary. In particular, for a straight, periodic channel there are the two sidewall contributions; for a closed, rectangular basin there is just one.

APPENDIX C

Weak Topographic Surface Boundary

With the assumed small Rossby number scaling in the PE (7)–(9), the effects of surface topography can be introduced through a roughening of the surface location to $z = \epsilon h(x, y)$. The boundary condition of no normal flow gives the modified surface condition

$$w = h_x u + h_y v \quad \text{at} \quad z = \epsilon h(x, y). \qquad \text{(C1)}$$

For small Rossby number, it is more convenient to maintain a flat surface and replace the topographic boundary condition by its Taylor expansion at $z = 0$:

$$\begin{aligned} (w + \epsilon h w_z + \ldots) \sim h_x(u + \epsilon h u_z + \ldots) \\ + h_y(v + \epsilon h v_z + \ldots). \end{aligned} \qquad \text{(C2)}$$

The continuity condition $F_x + G_y = \epsilon w$ is likewise applied at $z = 0$ using the above expansion for w.

REFERENCES

Allen, J. S., 1991: Balance equations based on momentum equations with global invariants of potential enstrophy and energy. *J. Phys. Oceanogr.,* **21,** 265–276.

——, 1993: Iterated geostrophic intermediate models. *J. Phys. Oceanogr.,* **23,** 2447–2461.

——, and P. A. Newberger, 1993: On intermediate models for stratified flow. *J. Phys. Oceanogr.,* **23,** 2462–2486.

Bluestein, H. B., 1992: *Synoptic-Dynamic Meteorology in Midlatitudes.* Vol. II. Oxford University Press, 594 pp.

Bokhove, O., 1997: Slaving principles, balanced dynamics, and the hydrostatic Boussinesq equations. *J. Atmos. Sci.,* **54,** 1662–1674.

Charney, J. G., 1948: On the scale of atmospheric motions. *Geophys. Publ. Oslo,* **17** (2), 1–17.

——, 1955: The use of primitive equations of motion in numerical prediction. *Tellus,* **7,** 22–26.

Gent, P. R., and J. C. McWilliams, 1983a: Consistent balanced models in bounded and periodic domains. *Dyn. Atmos. Oceans,* **7,** 67–93.

——, and ——, 1983b: Regimes of validity of balanced models. *Dyn. Atmos. Oceans,* **7,** 167–183.

——, ——, and C. Snyder, 1994: Scaling analysis of curved fronts: Validity of the balance equations and semigeostrophy. *J. Atmos. Sci.,* **51,** 160–163.

Gill, A. E., 1982: *Atmosphere–Ocean Dynamics.* Academic Press, 662 pp.

Holm, D. D., 1996: Hamiltonian balance equations. *Physica D,* **98,** 379–414.

Hoskins, B. J., and F. P. Bretherton, 1972: Atmospheric frontogenesis models: Mathematical formulation and solution. *J. Atmos. Sci.,* **29,** 11–37.

——, and N. V. West, 1979: Baroclinic waves and frontogenesis. Part II: Uniform potential vorticity jet flows—Cold and warm fronts. *J. Atmos. Sci.,* **36,** 1663–1680.

——, 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.

Lorenz, E. N., 1960: Energy and numerical weather prediction. *Tellus,* **12,** 364–373.

McWilliams, J. C., 1977: A note on a consistent quasi-geostrophic model in a multiply connected domain. *Dyn. Atmos. Oceans,* **1,** 427–441.

——, 1985: A uniformly valid model spanning the regimes of geostrophic and isotropic, stratified turbulence: Balanced turbulence. *J. Atmos. Sci.,* **42,** 1773–1774.

Mudrick, S. E., 1974: A numerical study of frontogenesis. *J. Atmos. Sci.,* **31,** 869–892.

Pedlosky, J., 1987: *Geophysical Fluid Dynamics.* Springer-Verlag, 710 pp.

Phillips, N. A., 1954: Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level, quasi-geostrophic model. *Tellus,* **6,** 273–286.

Raymond, D. J., 1992: Nonlinear balance and potential-vorticity

thinking at large Rossby number. *Quart. J. Roy. Meteor. Soc.,* **118,** 987–1015.

Snyder, C., W. C. Skamarock, and R. Rotunno, 1991: A comparison of primitive equation and semigeostrophic simulations of baroclinic waves. *J. Atmos. Sci.,* **48,** 2179–2194.

Vallis, G. K., 1996a: Potential vorticity inversion and balanced equations of motion for rotating and stratified flows. *Quart. J. Roy. Meteor. Soc.,* **122,** 291–322.

——, 1996b: Approximate geostrophic models for large-scale flow in the ocean and atmosphere. *Physica D,* **98,** 647–651.

Warn, T., O. Bokhove, T. G. Shepherd, and G. K. Vallis, 1995: Rossby number expansions, slaving principles and balance dynamics. *Quart. J. Roy. Meteor. Soc.,* **121,** 723–729.

Xu, Q., 1994: Semibalance model—Connection between geostrophic-type and balanced-type intermediate models. *J. Atmos. Sci.,* **51,** 953–970.