University of Reading

Rossby number expansions, slaving principles, and balance dynamics

Article

Published Version

Warn, T., Bokhove, O., Shepherd, T. G. ORCID: https://orcid.org/0000-0002-6631-9968 and Vallis, G. K. (1995) Rossby number expansions, slaving principles, and balance dynamics. Quarterly Journal of the Royal Meteorological Society, 121 (523). pp. 723-739. ISSN 1477-870X doi: 10.1002/qj.49712152313 Available at https://centaur.reading.ac.uk/32891/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.
Published version at: http://dx.doi.org/10.1002/qj.49712152313
To link to this article DOI: http://dx.doi.org/10.1002/qj.49712152313

Publisher: Royal Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Q. J. R. Meteorol. Soc. (1995), **121**, pp. 723–739 551.513.1:551.509.313.12

Rossby number expansions, slaving principles, and balance dynamics

By T. WARN[1], O. BOKHOVE[2], T. G. SHEPHERD[2]* and G. K. VALLIS[3]

[1] *McGill University, Canada*
[2] *University of Toronto, Canada*
[3] *University of California, Santa Cruz, USA*

(Received 5 January 1994; revised 8 July 1994)

SUMMARY

We consider the problem of constructing balance dynamics for rapidly rotating fluid systems. It is argued that the conventional Rossby number expansion—namely expanding all variables in a series in Rossby number—is secular for all but the simplest flows. In particular, the higher-order terms in the expansion grow exponentially on average, and for moderate values of the Rossby number the expansion is, at best, useful only for times of the order of the doubling times of the instabilities of the underlying quasi-geostrophic dynamics. Similar arguments apply in a wide class of problems involving a small parameter and sufficiently complex zeroth-order dynamics.

A modified procedure is proposed which involves expanding only the fast modes of the system; this is equivalent to an asymptotic approximation of the slaving relation that relates the fast modes to the slow modes. The procedure is systematic and thus capable, at least in principle, of being carried to any order—unlike procedures based on truncations.

We apply the procedure to construct higher-order balance approximations of the shallow-water equations. At the lowest order quasi-geostrophy emerges. At the next order the system incorporates gradient-wind balance, although the balance relations themselves involve only linear inversions and hence are easily applied. There is a large class of reduced systems associated with various choices for the slow variables, but the simplest ones appear to be those based on potential vorticity.

1. INTRODUCTION

The large-scale circulation of the mid-latitude atmosphere and oceans can be partially characterized by the Rossby number

$$\varepsilon = V/fL, \tag{1}$$

where V is a typical velocity magnitude, L is a typical length scale, and f is a characteristic value of the Coriolis parameter. For the atmosphere one usually takes $V \approx 10\,\mathrm{m\,s^{-1}}$, $L \approx 10^6\,\mathrm{m}$, and $f \approx 10^{-4}\mathrm{s}^{-1}$, in which case $\varepsilon \approx 0.1$. A still smaller value typifies the large-scale ocean circulation. The fact that the Rossby number is small and that the evolution is slow compared with the frequencies of gravity-inertial waves suggests the application of a formal asymptotic analysis of the (mid-latitude) equations of motion. Specifically one expands all dependent variables in a series in Rossby number and solves to leading order to obtain simplified, quasi-geostrophic dynamics (Pedlosky 1987). A similar procedure is widely used to simplify the dynamics in other branches of fluid mechanics, a notable example being the equations for flow in incompressible fluids at small Mach number.

While the conventional Rossby-number expansion referred to above has been widely accepted as providing a justification for quasi-geostrophic theory, it is purely formal and not always valid. It will, not surprisingly, break down if the scaling is either internally inconsistent (as is the case for ultralong waves or equatorial motion on the sphere) or does not remain uniformly valid in time (as appears to be the case near developing fronts or cyclones, for example). Furthermore the expansion only appears to have been carried

* Corresponding author: Department of Physics, University of Toronto, 60 St George Street, Toronto, Ontario, Canada M5S 1A7.

out beyond leading order when the underlying flows are relatively simple. We argue below that the conventional expansion is non-uniformly valid for realistic flows with chaotic quasi-geostrophic dynamics, in the sense that higher-order terms are secular and grow exponentially at a rate determined by the instabilities of the associated quasi-geostrophic flow. While this may or may not be surprising, it does raise questions about the validity of quasi-geostrophic theory on moderate to long timescales, and about how quasi-geostrophy might be extended to higher order in a systematic fashion. We note with regards to the first of these that the non-uniformity may merely be a reflection of the lack of predictability of complex systems and the related difficulty of constructing approximations to specific trajectories for extended periods of time. The optimist might argue that, despite the non-uniformity, quasi-geostrophy still gives a meaningful leading-order long-term description in some statistical or climatological sense. On the other hand, the pessimist might worry that quasi-geostrophy might fail even in a statistical sense owing to the cumulative effect of small terms (much in the same way as the addition of a small amount of friction alters the long-term behaviour of a pendulum). This problem of long-term accuracy is an extremely difficult one which we do not attempt to deal with here.

The quasi-geostrophic equations are an example of a balance model. By a balance model we mean a reduced set of evolution equations that filters out fast oscillations and describes the slow-time evolution of a system. The concept of fast and slow motion, of course, only has a meaning in the presence of a timescale separation. Although our objective in this paper lies in small Rossby number balance, the considerations apply to any balance model that can be understood in these terms. Balance models involve two components:

(i) A balance relation, which filters the fast oscillations and thereby reduces the dimensionality of the system.

(ii) Balance, reduced or slow dynamics, which describes the slow-time evolution.

It may be that balance relations remain accurate while balance dynamics does not; in which case it is of interest to modify the conventional expansion in a manner that separates these two aspects. This turns out to be possible with the aid of a variant of the bounded derivative method (Kreiss 1979, 1980, 1985), the method of elimination of fast variables (Van Kampen 1985), and the normal-mode initialization procedures developed by Baer (1977), Machenhauer (1977), Leith (1980), Lorenz (1980) and others. The bounded derivative method has the advantage that rigorous results are available for some systems.

The balance relation for the shallow-water equations in the small Rossby number limit is given by the geostrophic relation $f\mathbf{k} \times \mathbf{v} = -g\nabla\eta$ to leading order (see section 4 for further definitions), while the balance dynamics is described by the usual quasi-geostrophic potential-vorticity equation. The geostrophic relation is often said to be diagnostic since it represents an instantaneous relation between fields. Following Van Kampen (1985) and others, we shall employ the term 'slave relation' to emphasize that the related variables are in fact independent from the standpoint of the full equations*. While various versions of higher-order balance systems have been proposed (e.g. Charney 1955; Phillips 1960; McWilliams and Gent 1980; Lynch 1989; Allen *et al.* 1990) which are formally and numerically more accurate than quasi-geostrophy, most are based on truncations of the equations of motion and therefore cannot be extended beyond first

* Hoskins *et al.* (1985) use the term 'invertibility principle' in this context.

order in the Rossby number†. Iterative procedures have been proposed by Allen (1993) and by M. E. McIntyre and W. A. Norton (personal communication) which, as in our approach, do not suffer from this limitation.

There are ambiguities inherent in the notion of balance which are, in a certain sense, present even at the leading order. Thus the geostrophic relation can be viewed as implying either that the velocity is determined by the height or that the height is determined by the velocity. For the shallow-water system one can also regard the free-surface height, η, and the divergent part of the velocity as slaved to the rotational part of the velocity, i.e.

$$\eta = U_\eta(\psi) \qquad \chi = U_\chi(\psi)$$

where ψ and χ are the stream function and velocity potential, respectively, and where

$$U_\eta = \frac{f\psi}{g} \qquad U_\chi = 0.$$

Alternatively one can use η as the slaving variable and write

$$\psi = V_\psi(\eta) \qquad \chi = V_\chi(\eta).$$

Other possibilities also exist. It turns out that it is best to view balance geometrically as a surface or manifold in an appropriately defined phase space (Leith 1980; Lorenz 1980). This is most easily illustrated in the case of doubly periodic flows when the Fourier coefficients of the dependent variables are used as coordinates of the phase space. The geostrophic relation defines a hyperplane or 'geostrophic manifold' in the (infinite dimensional) phase space. The notion of an 'approximately invariant manifold' arises for flows that are initially close to this geostrophic manifold and which remain in its proximity at subsequent times. The motion will then also be slow in the sense that the acceleration terms in the momentum equations remain small.

The assumption that there exists an infinite-order generalization of geostrophy is equivalent to assuming the existence of either a linear or nonlinear slaving condition

$$F(x, y; \varepsilon) = 0,$$

where the slaving and slaved variables x and y together represent a complete set of dependent variables. Approximate balance relations come in a variety of forms since the accuracy of an approximation does not determine the form uniquely. As an example we note that it is often argued that while conventional scaling leads, to order ε, to the Charney balance equation

$$f\nabla^2\psi + 2J(\psi_x, \psi_y) = g\nabla^2\eta \tag{2}$$

(J stands for Jacobian), the expressions

$$f\nabla^2\psi + 2\frac{g^2}{f^2}J(\eta_x, \eta_y) = g\nabla^2\eta \tag{3}$$

and

$$f\nabla^2\psi + \frac{g}{f}\{J(\psi_x, \eta_y) + J(\eta_x, \psi_y)\} = g\nabla^2\eta \tag{4}$$

† It seems worthwhile to point out that the 3-D geostrophic momentum approximation (Hoskins 1975) is formally not more accurate than quasi-geostrophy under conventional scaling (McWilliams and Gent 1980) but appears to have the merit of remaining accurate in the vicinity of weakly curved fronts where the conventional scaling breaks down (Craig 1993).

obtained by using geostrophy to replace all or some of the terms ψ in the (small) nonlinear term by η, have the same formal accuracy. Since there are, in fact, an infinity of balance relations of a given (formal) accuracy, the choice can be made on the basis of some other criteria (e.g. the variables observed, numerical efficiency, preservation of conservation laws and/or symmetries when coupled with dynamics, etc.).

Note that if η is given, (3) is linear in ψ, whereas the Charney balance equation is nonlinear and requires the satisfaction of a side condition to ensure invertibility, namely that the usual ellipticity condition $\nabla^2\psi \geq -f/2$ must be satisfied. However this is a somewhat artificial distinction, because inversion is always possible provided the scaling is respected, since the ratio of relative to absolute vorticity is then of order ε. A violation of the ellipticity condition therefore merely serves as a warning that the balance equations, and probably the rationale for balance itself, fail. It is emphasized that these conclusions depend on the particular scaling chosen. In the case of the shallow-water equations, Charney balance is formally accurate to leading order when the Rossby number is of order unity and the Froude number is small; this may represent a suitable scaling in the tropics, away from convective complexes (Charney 1963; Browning *et al.* 1980).

The non-uniformity of the conventional expansion and the proposed modification are most easily illustrated via a purely formal discussion which is presented in sections 2 and 3. The formal discussion emphasizes the fact that the methodology applies to any system involving fast and slow motion. The method is then applied, in section 4, to the rapidly rotating shallow-water equations, and the results are summarized in section 5.

2. Regular perturbations of complex systems

Consider first an initial-value problem consisting of a system of coupled ordinary or partial differential equations with a single timescale of the form

$$\frac{\partial \mathbf{u}}{\partial t} = \mathbf{N}(\mathbf{u}; \varepsilon), \tag{5}$$

in the limit of small ε. Here the dependent variable $\mathbf{u}$ is a vector function possibly depending on spatial variables and time t, while $\mathbf{N}(\mathbf{u}; \varepsilon)$ is a nonlinear operator or function of $\mathbf{u}$. To fix the scaling it is assumed that $\mathbf{N}(\mathbf{u}; 0) \neq 0$.

Substituting

$$\mathbf{u} = \mathbf{u}^{(0)} + \varepsilon\mathbf{u}^{(1)} + \ldots$$

into (5) and equating powers of ε gives the following sequence:

$$\frac{[\partial \mathbf{u}^{(0)}]}{\partial t} = \mathbf{N}(\mathbf{u}^{(0)}; 0) \tag{6}$$

$$\frac{[\partial \mathbf{u}^{(1)}]}{\partial t} = \mathscr{L}(\mathbf{u}^{(0)})\mathbf{u}^{(1)} + \mathscr{F}(\mathbf{u}^{(0)}) \tag{7}$$

etc., expanding on $\mathbf{N}(\mathbf{u}; \varepsilon)$ about $\mathbf{u} = \mathbf{u}^{(0)}$ and $\varepsilon = 0$. Here $\mathscr{L}$ is a linear operator which depends on $\mathbf{u}^{(0)}$. We note that it is not necessary, neither is it expected, that the series converges for $\varepsilon \neq 0$. All that is required is that it be asymptotic, i.e.

$$\frac{\|\mathbf{u} - \mathbf{u}^{(0)} - \ldots - \varepsilon^m \mathbf{u}^{(m)}\|}{\varepsilon^m} \to 0$$

as $\varepsilon \to 0$ for each m, where $\|\mathbf{u}\|$ denotes an appropriate norm. In other words, the error in taking a given number of terms can be made as small as desired by taking ε sufficiently small. For a fixed but small ε it often happens in practice that the error decreases with m up to some order and thereafter increases.

The leading-order dynamics is therefore described by (6), and it might be expected that a more accurate description could be obtained by proceeding to higher order. However, there is a real possibility that $\mathbf{u}^{(1)}$ becomes unbounded in time since (7) is linear in $\mathbf{u}^{(1)}$, and $\mathscr{L}$ depends on time and perhaps space through $\mathbf{u}^{(0)}$. This is in fact the case if the zeroth-order solution is chaotic and exhibits sensitivity to the initial conditions, since the homogeneous version,

$$\frac{\partial \phi}{\partial t} = \mathscr{L}(\mathbf{u}^{(0)}(t))\phi,$$

represents the linearization of the zeroth-order system about a solution. The quantities ϕ and hence $\mathbf{u}^{(1)}$ then grow exponentially on average at a rate determined by the largest Liapunov exponent, λ, which is a measure of the strength of the strongest instability* (e.g. Guckenheimer and Holmes 1983). In contrast to the usual initial growth rates of classical linear instability theory, the Liapunov exponents measure the instability of the final or long-term solution. The quantity $\mathbf{u}^{(1)}$ consequently grows without bound, becoming of the same order as $\mathbf{u}^{(0)}$ in times, T, of the order of $-(\log \varepsilon)/\lambda$, which, for moderate values of ε, is a few doubling times of the strongest instability. The direct expansion is therefore not uniformly valid in the sense that it is not asymptotic for times of order T or larger. The quantity $\mathbf{u}^{(2)}$ also evolves according to an equation of the form of (7) with a forcing depending linearly on $\mathbf{u}^{(1)}$, and so might amplify even more quickly than $\mathbf{u}^{(1)}$. (Secular growth is also possible when $\lambda = 0$, though it will then be linear rather than exponential in time.)

While the non-uniform validity is perhaps not surprising, given the supposed instability of the zeroth-order flow, it renders the direct expansion useless on long timescales. Secular behaviour also often signals a breakdown of the zeroth-order dynamics on long timescales. In some problems the time interval can be extended by appropriately modifying the expansion (using the method of multiple scales for example) by including feedbacks to the leading-order dynamics. Such methods hold little promise here owing to the fact that the secular growth is much stronger than the algebraic or weak exponential growth usually treated (e.g. in weakly nonlinear problems), and that the methods require a reasonably detailed knowledge of the solutions, which is not available in general.

The only approximation procedure that we are aware of which would avoid secular growth involves truncating the Taylor-series expansion of $\mathbf{N}$ in ε to obtain the sequence

$$\frac{\partial \mathbf{u}}{\partial t} = \mathbf{N}_0(\mathbf{u}) + \varepsilon \mathbf{N}_1(\mathbf{u}) + \ldots + \varepsilon^m \mathbf{N}_m(\mathbf{u}).$$

Here $0 \leqslant m \leqslant n$, where n is the number of models in the hierarchy; n may be finite or infinite depending on the ε-dependence of $\mathbf{N}$. Other systems can be obtained by transforming to a new variable, viz.

$$\tilde{\mathbf{u}} = \mathrm{G}(\mathbf{u}; \varepsilon),$$

where $\mathbf{G}$ is an order unity one-to-one transformation. This permits a certain degree of control over the hierarchy through the choice of $\mathbf{G}$.

* The solutions do not have exponential time dependence since $\mathscr{L}$ depends on t.

3. Singular perturbations of complex systems

Next consider singular systems of ordinary or partial differential equations for which the order in time changes when the small parameter vanishes, i.e. systems involving a combination of fast and slow modes of the form

$$\frac{\partial \mathbf{s}}{\partial t} = \mathbf{S}(\mathbf{s}, \mathbf{f}; \varepsilon) \tag{8}$$

and

$$\frac{\partial \mathbf{f}}{\partial t} + \frac{\Gamma \mathbf{f}}{\varepsilon} = \mathbf{F}(\mathbf{s}, \mathbf{f}; \varepsilon) \tag{9}$$

where $\mathbf{S}$ and $\mathbf{F}$ are nonlinear vector operators or functions that are non-zero when $\varepsilon, \mathbf{f} = 0$. Here $\mathbf{f}$ and $\mathbf{s}$ denote fast and slow modes, respectively, and Γ is an invertible linear operator (independent of ε, time and space) such that the equation

$$\frac{\partial \mathbf{f}}{\partial t} + \frac{\Gamma \mathbf{f}}{\varepsilon} = 0$$

has normal-mode solutions of the form

$$\phi_\nu(\mathbf{x}) \exp(\mathrm{i}\omega_\nu t \varepsilon^{-1}),$$

where ω_ν is real and

$$|\omega_\nu| \geqslant 1 \quad \text{for all } \nu$$

Kreiss 1985). This ensures that the partition into fast and slow modes is unambiguous and that their interaction will be weak. As will be discussed in the next section, the planar rotating shallow-water equations under the usual quasi-geostrophic scaling fall into this class, although the spherical system does not. We note that when ε is not small or when the constraint on the frequency is not satisfied then the distinction between the fast and slow modes becomes blurred. Spatially infinite or semi-infinite systems with a non-dispersive 'fast' branch (e.g. sound waves or Kelvin waves) with $\omega_k = kc$, where k is the modulus of the spatial wave number, do not satisfy the frequency constraint, since long waves with $k = \mathrm{O}(\varepsilon)$ are not fast and can, in principle, interact strongly with the slow modes and participate fully in the evolution. In this case the notions of balance and slaving are more difficult to support, as is the relevance of asymptotic arguments to the initialization problem of numerical weather prediction.

We note that there is nothing to be gained in displaying the slow dispersion relation explicitly, since (8) is fully nonlinear and linear dispersion relations are relevant only in the small-amplitude limit. Similarly Γ, which plays a central role in the reduction, is well defined and significant only when ε is small, so the notion of slow motion and balance can be defined unambiguously only in the presence of a small parameter.

(*a*) *Direct expansion*

If slow solutions to (8) and (9) are sought by substituting the expansions

$$\mathbf{s} = \mathbf{s}^{(0)} + \varepsilon \mathbf{s}^{(1)} + \ldots \qquad \text{and} \quad \mathbf{f} = \mathbf{f}^{(0)} + \varepsilon \mathbf{f}^{(1)} + \ldots$$

into (8) and (9), we obtain

$$\frac{\partial \mathbf{s}^{(0)}}{\partial t} = \mathbf{S}_0(\mathbf{s}^{(0)}) \tag{10}$$

$$\mathbf{f}^{(0)} = 0 \tag{11}$$

at leading order; while at the next order we have

$$\frac{\partial \mathbf{s}^{(1)}}{\partial t} = \mathscr{L}(\mathbf{s}^{(0)})\mathbf{s}^{(1)} + \mathscr{N}(\mathbf{s}^{(0)})\mathbf{f}^{(1)} + \mathscr{G}(\mathbf{s}^{(0)}) \tag{12}$$

$$\mathbf{f}^{(1)} = \Gamma^{-1}\mathbf{F}_0(\mathbf{s}^{(0)}). \tag{13}$$

Here $\mathbf{S}_0(\mathbf{s}) = \mathbf{S}(\mathbf{s}, 0; 0)$, $\mathbf{F}_0(\mathbf{s}) = \mathbf{F}(\mathbf{s}, 0; 0)$, and $\mathscr{L}$, $\mathscr{N}$ and $\mathscr{G}$ follow directly on expanding $\mathbf{S}$ about $\mathbf{s} = \mathbf{s}^{(0)}$, $\mathbf{f} = 0$, and $\varepsilon = 0$, respectively. In particular, $\mathscr{L}$ is the linearization of $\mathbf{S}(\mathbf{s}, 0; 0)$ about $\mathbf{s}^{(0)}$. Eliminating $\mathbf{f}^{(1)}$ then gives

$$\frac{\partial \mathbf{s}^{(1)}}{\partial t} = \mathscr{L}(\mathbf{s}^{(0)})\mathbf{s}^{(1)} + \mathscr{F}(\mathbf{s}^{(0)}) \tag{14}$$

which has the same form as (7). Since the homogeneous version of this equation describes the divergence of the trajectories of (10), $\mathbf{s}^{(1)}$ amplifies exponentially on average when the leading-order system is chaotic, implying that the expansion becomes disordered (i.e. $\mathbf{s}^{(1)} = O(\varepsilon^{-1})$, etc.) on timescales determined by the largest Liapunov exponent.

(*b*) *Slaving*

The assumption that an arbitrary flow with slow forcing and small Rossby number can be described asymptotically by an infinite-order variant of quasi-geostrophic theory (after an initial transient phase), would appear to be equivalent to assuming the existence of a 'slow manifold' $\mathscr{M}$, which can be represented by a graph

$$\mathbf{f} = \mathbf{U}(\mathbf{s}; \varepsilon) \tag{15}$$

that is approximately invariant in the sense that trajectories are attracted to, and subsequently remain within, thin neighbourhoods of $\mathscr{M}$. The manifold is 'slow' in the sense that nearby trajectories are balanced and satisfy the bounded derivative condition, i.e. the first n time derivatives of the dependent variables, with respect to the slow time, relax to order unity (Kreiss 1985). While this definition is somewhat weaker than earlier versions (Leith 1980; Lorenz 1980), which required that the manifold also be invariant*, the existence of $\mathscr{M}$ remains in doubt; although a weaker result due to Kreiss (1979, 1980, 1985) indicates that the bounded derivative condition defines a sequence of manifolds $\mathscr{M}_n$ that are approximately invariant for some finite (possibly small) slow time. If the $\mathscr{M}_n$ could also be shown to be attracting, then the time interval could be extended to infinity and they would represent a sequence of approximately invariant manifolds. If no such sequence with these properties exists then the relevance of quasi-geostrophy, the bounded derivative method, and related initialization schemes, to the long-term dynamics needs to be re-examined.

It is perhaps worth emphasizing that Kreiss's results do not make explicit use of the dissipative structure of the equations, which is probably required to control the fast modes. Also the slow manifold as defined here is distinct from inertial or approximately inertial manifolds (Debussche and Temam 1991a, b), although the formal constructions have similarities. Roughly speaking, an inertial manifold is a finite-dimensional invariant manifold that attracts orbits exponentially so that there is a decomposition of the form $\mathbf{u} = \mathbf{u}_n + \mathbf{u}'$ where $\mathbf{u}_n$ is finite dimensional (e.g. it has a finite spectral representation) and

* i.e. that trajectories initially in $\mathscr{M}$ remain in $\mathscr{M}$, or equivalently that there is an exact balance relation such that initially balanced flows remain balanced.

the residual $\mathbf{u}'$ is slaved after a transient period, i.e. $\mathbf{u}' = \mathbf{f}(\mathbf{u}_n)$. Inertial manifolds involve slaving of strongly dissipated modes and, in contrast to the slow manifold as defined in meteorology and oceanography, make no reference to a frequency separation.

Substituting (15) into (9) and using (8) yields the 'superbalance equation' (Lorenz 1980)

$$\mathcal{R}(\mathbf{U}) \equiv \varepsilon \mathcal{T}\mathbf{S}(\mathbf{s}, \mathbf{U}; \varepsilon) + \Gamma\mathbf{U} - \varepsilon\mathbf{F}(\mathbf{s}, \mathbf{U}; \varepsilon) = 0, \tag{16}$$

where $\mathcal{T}$ denotes the linearization of $\mathbf{U}$ about $\mathbf{s}$. Similarly substituting (15) into (8) gives

$$\frac{\partial \mathbf{s}}{\partial t} = \mathbf{S}(\mathbf{s}, \mathbf{U}(\mathbf{s}; \varepsilon); \varepsilon) \tag{17}$$

for the evolution of the slow variables alone. If solutions to (16) could be found that were analytic in ε, then an invariant slow manifold would exist; (15) would represent an exact interaction law between the fast and slow modes, while (17) would give an exact description of the slow dynamics. Exact slaving seems unlikely, however (Lorenz 1986; Vautard and Legras 1986; Warn and Menard 1986), although the expansion

$$\mathbf{U}\Big|_{\mathrm{O}(\varepsilon^n)} = \mathbf{U}^{(0)}(\mathbf{s}) + \varepsilon\mathbf{U}^{(1)}(\mathbf{s}) + \ldots + \varepsilon^n\mathbf{U}^{(n)}(\mathbf{s}) \tag{18}$$

generated from (16) leads to the bounded derivative manifolds $\mathcal{M}_n$ which provide a sequence of approximate slave relations. Here the notation $(\cdot)|_{\mathrm{O}(\varepsilon^n)}$ denotes the $\mathrm{O}(\varepsilon^n)$ approximation. Substituting (18) into (8), expanding and truncating then gives a hierarchy of reduced or slow systems.

(*c*) *Modified expansion*

The expansion above can be seen to be equivalent to constructing slave relations with the aid of the expansions

$$\mathbf{s} = \mathbf{s} \qquad \mathbf{f} = \mathbf{f}^{(0)} + \varepsilon\mathbf{f}^{(1)} + \ldots \tag{19}$$

i.e. by expanding only the fast modes in both the fast and slow equations and then using the equations

$$\frac{\partial \mathbf{f}}{\partial t} = (\mathcal{T}^{(0)} + \varepsilon\mathcal{T}^{(1)} + \ldots)\frac{\partial \mathbf{s}}{\partial t} = (\mathcal{T}^{(0)} + \varepsilon\mathcal{T}^{(1)} + \ldots)\mathbf{S}(\mathbf{s}, \mathbf{f}; \varepsilon), \tag{20}$$

where $\mathcal{T}^{(n)}$ denotes the linearization of $\mathbf{f}^{(n)} = \mathbf{f}^{(n)}(\mathbf{s})$ with respect to $\mathbf{s}$, before equating powers of ε. The full slow equation is then used to generate reduced dynamics (cf. Van Kampen 1985).

The first few models in the hierarchy are:

$$\mathrm{O}(1)\text{: } \mathbf{f} = 0, \qquad \frac{\partial \mathbf{s}}{\partial t} = \mathbf{S}(\mathbf{s}, 0; 0) \tag{21}$$

$$\mathrm{O}(\varepsilon)\text{: } \mathbf{f} = \varepsilon\Gamma^{-1}\mathbf{F}(\mathbf{s}, 0; 0), \qquad \frac{\partial \mathbf{s}}{\partial t} = \mathbf{S}(\mathbf{s}, \mathbf{f}; \varepsilon)\Big|_{\mathrm{O}(\varepsilon)} \tag{22}$$

$$\vdots$$

$$\mathrm{O}(\varepsilon^n)\text{: } \mathbf{f} = \varepsilon\Gamma^{-1}\{\mathbf{F}(\mathbf{s}, \mathbf{f}; \varepsilon) - \mathcal{T}\mathbf{S}(\mathbf{s}, \mathbf{f}; \varepsilon)\}\Big|_{\mathrm{O}(\varepsilon^{n-1})}, \qquad \frac{\partial \mathbf{s}}{\partial t} = \mathbf{S}(\mathbf{s}, \mathbf{f}; \varepsilon)\Big|_{\mathrm{O}(\varepsilon^n)}. \tag{23}$$

Although (23) appears to be implicit in $\mathbf{f}$, it is not so because each term in the expansion

(19) of $\mathbf{f}$ can be solved for in a sequential manner. The method is demonstrated in section 4 by an application to the rapidly rotating shallow-water equations.

It should be noted that the expansion method depends only on the choice of the slow variable. The evolution equations do not have to be in the normal form (8)–(9), nor is it necessary to restrict considerations to the variables $\mathbf{f}$ and $\mathbf{s}$. Any set of variables $(\mathbf{x}, \mathbf{y})$ can be used provided that they are independent to leading order and that the slaving variable $\mathbf{x}$ projects onto the slow modes. It turns out that the final form of the resulting balance approximations and reduced dynamics generally depends on this choice.

(*d*) *Iterative approximation*

An alternative approach to approximating the superbalance equation (16) is an iteration procedure, defined by the equation

$$\mathbf{U}_{n+1} = \varepsilon\Gamma^{-1}\{\mathbf{F}(\mathbf{s}, \mathbf{U}_n; \varepsilon) - \mathcal{T}_n\mathbf{S}(\mathbf{s}, \mathbf{U}_n; \varepsilon)\}, \tag{24}$$

with the starting point $\mathbf{U}_0 = 0$, where $\mathcal{T}_n$ denotes the linearization of $\mathbf{U}_n$ with respect to $\mathbf{s}$. (We denote iterates by subscripts, and terms in the asymptotic expansion by superscripts.) This can be seen to be equivalent to iterating the fast variables but not the slow variables. The nth-iterate of the system (8)–(9) is evidently

$$\frac{\partial \mathbf{s}}{\partial t} = \mathbf{S}(\mathbf{s}, \mathbf{f}_n; \varepsilon) \tag{25}$$

$$\frac{\partial \mathbf{f}_{n-1}}{\partial t} + \frac{\Gamma \mathbf{f}_n}{\varepsilon} = \mathbf{F}(\mathbf{s}, \mathbf{f}_{n-1}; \varepsilon) \tag{26}$$

with the starting point $\mathbf{f}_0 = 0$. Equation (26) may be written in the form

$$\mathbf{f}_n = \varepsilon\Gamma^{-1}\{\mathbf{F}(\mathbf{s}, \mathbf{f}_{n-1}; \varepsilon) - \mathcal{T}_{n-1}\mathbf{S}(\mathbf{s}, \mathbf{f}_{n-1}; \varepsilon)\} \tag{27}$$

which is equivalent to (24) with the identification $f_n = U_n$. The iterative approximation has recently been proposed by Allen (1993) in the context of small Rossby number balance.

There are obvious similarities between the two approaches, as can be seen by comparing (23) with (25) and (27). The modified expansion contains only terms up to $O(\varepsilon^n)$, while the nth-order iterative approximation will generally contain higher-order terms as well. In the special case where $\mathbf{F}$ and $\mathbf{S}$ are both linear in $\mathbf{f}$ and contain no explicit dependence on ε, the two approaches are equivalent.

Both approaches have the merit of avoiding the problem of secular growth of the slow variable. But it must be emphasized that the problem of long-term accuracy has not been solved. The accuracy of a given model refers to the accuracy of the balance relation—not of the balance dynamics.

4. The shallow-water system

(*a*) *Preliminaries*

In a frame rotating with constant angular velocity, the equations of motion can be written in the dimensionless form

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\frac{1}{\varepsilon}(\mathbf{k} \times \mathbf{v} + \nabla\eta) \tag{28}$$

$$Fr\frac{\mathrm{d}\eta}{\mathrm{d}t} + Fr\,\eta\,\nabla\cdot\mathbf{v} = -\frac{1}{\varepsilon}\nabla\cdot\mathbf{v} \tag{29}$$

where $\mathbf{v}$ is the horizontal velocity of the fluid, η is the free-surface displacement, $\mathrm{d}/\mathrm{d}t$ is the material derivative, and $\mathbf{k}$ is the unit vector in the vertical. Here ε is the Rossby number defined by (1), and Fr, defined by the expression

$$Fr = \frac{f^2L^2}{gH}$$

is the (square) rotational Froude number; f, H, and g are the Coriolis parameter, the mean fluid depth and the acceleration of gravity, respectively, and V and L are typical velocity and length scales. It is assumed that the Rossby number is small and that the rotational Froude number is unity. Other choices are also possible (e.g. Browning *et al.* 1980).

If we introduce the potential vorticity, q, defined by the equations

$$1 + \varepsilon q = \frac{1 + \varepsilon\nabla^2\psi}{1 + \varepsilon\eta} \iff q = \frac{\nabla^2\psi - \eta}{1 + \varepsilon\eta} \tag{30}$$

the divergence, D, given by

$$D = \nabla\cdot\mathbf{v} = \nabla^2\chi \tag{31}$$

and the geostrophic departure, Ω (Browning and Kreiss 1987; Lynch 1989), given by

$$\Omega = \nabla^2\psi - \nabla^2\eta \tag{32}$$

where the Helmholtz decomposition, $\mathbf{v} = \mathbf{k}\times\nabla\psi + \nabla\chi$, has been used and ψ and χ are the stream function and velocity potential, then the equations of motion can be written in the form

$$\frac{\partial q}{\partial t} = -J(\psi, q) - \nabla\chi\cdot\nabla q, \tag{33}$$

$$\frac{\partial D}{\partial t} - \frac{\Omega}{\varepsilon} = -J(\chi, \nabla^2\psi) - \frac{1}{2}\nabla^2|\nabla\chi|^2 + \nabla^2 J(\chi, \psi) + 2J(\psi_x, \psi_y) \tag{34}$$

$$\frac{\partial\Omega}{\partial t} - \frac{\mathcal{H}D}{\varepsilon} = -J(\psi, \nabla^2\psi) - \nabla\cdot(\nabla^2\psi\,\nabla\chi) + \nabla^2\{J(\psi, \eta) + \nabla\cdot(\eta\,\nabla\chi)\}. \tag{35}$$

Here $J(\alpha, \beta) \equiv \alpha_x\beta_y - \alpha_y\beta_x$, subscripts denoting partial derivatives, and $\mathcal{H} \equiv \nabla^2 - 1$. Suppose also that the geometry and boundary conditions ensure that ∇^2 and $\mathcal{H}$ are invertible (as is the case for doubly periodic flows and localized flows that vanish sufficiently rapidly at infinity). If q and (D, Ω) are identified as slow and fast variables, respectively, then (33) to (35) are seen to be of the form (8)–(9).

(*b*) *Direct expansion*

If all quantities are expanded as power series in the Rossby number then (30) to (35) give

$$\chi^{(0)} = 0 \qquad \Omega^{(0)} = 0 \qquad q^{(0)} = \mathcal{H}\psi^{(0)}$$

and

$$\frac{[\partial q^{(0)}]}{\partial t} = -J(\psi^{(0)}, q^{(0)})$$

at the leading order, which is the usual shallow-water quasi-geostrophic system. At the next order we obtain

$$\Omega^{(1)} = -2J(\psi_x^{(0)}, \psi_y^{(0)}) \tag{36}$$

and

$$\Lambda\chi^{(1)} = J(\psi^{(0)}, \nabla^2\psi^{(0)}), \tag{37}$$

where $\Lambda \equiv \mathcal{H}\nabla^2$, which are versions of the usual balance and omega equations, and

$$\frac{[\partial q^{(1)}]}{\partial t} = -J(\psi^{(0)}, q^{(1)}) - J(\psi^{(1)}, q^{(0)}) - \nabla\chi^{(1)}\cdot\nabla q^{(0)}, \tag{38}$$

with

$$q^{(1)} = \nabla^2\psi^{(1)} - \eta^{(1)} - \psi^{(0)}q^{(0)} \tag{39}$$

from the definition of q. Since (39) and (36) imply that

$$\mathcal{H}\psi^{(1)} = q^{(1)} + \psi^{(0)}q^{(0)} + 2\nabla^{-2}J(\psi_x^{(0)}, \psi_y^{(0)}),$$

(38) becomes

$$\frac{[\partial q^{(1)}]}{\partial t} = \mathcal{L}(q^{(0)})q^{(1)} + \mathcal{F}(q^{(0)}),$$

where

$$\mathcal{L}(q^{(0)}) = -J(\psi^{(0)}, (\cdot)) + J(q^{(0)}, \mathcal{H}^{-1}(\cdot))$$

and

$$\mathcal{F}(q^{(0)}) = -\nabla q^{(0)}\cdot\nabla\{\Lambda^{-1}J(\psi^{(0)}, \nabla^2\psi^{(0)})\} + \\ + J(q^{(0)}, 2\Lambda^{-1}J(\psi_x^{(0)}, \psi_y^{(0)}) + \mathcal{H}^{-1}(\psi^{(0)}q^{(0)})).$$

Thus, as noted in the last section, the first-order correction evolves according to a linear forced problem with the associated homogenous part given by the linearization of the zeroth-order problem. The quantity $q^{(1)}$ therefore grows exponentially, on average, when the underlying quasi-geostrophic system is chaotic. The direct expansion is therefore secular.

(c) *Potential-vorticity slaving*

The secular behaviour noted above can be avoided using the modified expansion discussed in section 3, i.e. by expanding all quantities except the slaving variable in a series in the Rossby number. The choice (q, Ω, D) with q as the slaving variable is perhaps the simplest, since the equations are then in the normal form (8)–(9). Moreover, it leads to reduced dynamics that maintains the material conservation of q.

Expanding all quantities except q in (30)–(35) gives $\Omega^{(0)} = 0$ and $\chi^{(0)} = 0$, as before, and

$$\nabla^2\psi^{(0)} - \psi^{(0)} = q,$$

which provide the leading-order slave relations for the velocity and height in terms of q.

At the next order (36) and (37) are recovered, while (39) is replaced by the equation

$$\nabla^2\psi^{(1)} - \eta^{(1)} = \eta^{(0)}q.$$

The quantities $\psi^{(1)}$ and $\chi^{(1)}$ are therefore determined, given q. If we take

$$\psi = \psi^{(0)} + \varepsilon\psi^{(1)} \qquad \chi = \chi^{(0)} + \varepsilon\chi^{(1)}$$

then the reduced system takes the concise form

$$\frac{\partial q}{\partial t} + \mathbf{v}\cdot\nabla q = 0, \tag{40}$$

where $\mathbf{v}$ is determined from the relations

$$\Lambda\chi = \varepsilon J(\gamma, \nabla^2\gamma) \tag{41}$$

and

$$\Lambda\psi = \nabla^2 q + \varepsilon\{\nabla^2(\gamma q) + 2J(\gamma_x, \gamma_y)\} \tag{42}$$

with $\gamma \equiv \psi^{(0)} = \mathcal{H}^{-1}q$. We note that potential vorticity is exactly conserved, and that mass is conserved to $O(\varepsilon)$. A commendable feature of this system is that (41) and (42) are linear in χ and ψ.

The Charney balance system (Allen *et al.* 1990) has the form

$$\frac{\partial\nabla^2\psi}{\partial t} + J(\psi, \nabla^2\psi) + \nabla\cdot(\nabla^2\psi\,\nabla\chi) + \frac{D}{\varepsilon} = 0 \tag{43}$$

$$\frac{\partial\eta}{\partial t} + J(\psi, \eta) + \nabla\cdot(\eta\,\nabla\chi) + \frac{D}{\varepsilon} = 0 \tag{44}$$

$$\nabla^2\psi - \nabla^2\eta = -2\varepsilon J(\psi_x, \psi_y). \tag{45}$$

Although the system seems to be of second order in time, it is not so, owing to the fact that the balance constraint involves only ψ and η and can be used to eliminate one of the time derivatives. However, the system is of second order in time if the planetary vorticity gradient β is included, since the balance constraint then also involves the divergent part of the velocity. The system (43)–(45) can be expressed in terms of potential-vorticity advection, (40), together with two diagnostic relations which determine χ and ψ. These diagnostic relations are identical to (41)–(42) to the first order. Thus the Charney balance system, although formally of the same accuracy as (40)–(42), differs by the inclusion of higher-order terms which guarantee mass conservation in addition to material conservation of potential vorticity.

Lynch (1989) also derived a first-order balance model by dropping the time tendencies in (34)–(35) while keeping the potential-vorticity evolution equation. Formally, it has the same accuracy as (40)–(42). Unlike the Charney balance system, Lynch's method—like our own—does not introduce spurious modes when the β-effect is included.

One advantage of the present approach is that we may proceed systematically to higher order. At second order, for example, we obtain (40) with

$$\Lambda\chi = \varepsilon G(\gamma) + \varepsilon^2 C(\gamma) \tag{46}$$

and

$$\Lambda\psi = \nabla^2 q + \varepsilon R(\gamma) + \varepsilon^2 P(\gamma) \tag{47}$$

where $G(\gamma)$, $C(\gamma)$, $R(\gamma)$ and $P(\gamma)$ are defined by the equations

$$
\begin{aligned}
G(\gamma) &= J(\gamma, \nabla^2\gamma) \\
R(\gamma) &= \nabla^2(\gamma q) + 2J(\gamma_x, \gamma_y) \\
C(\gamma) &= 2J(\partial_x\{\mathcal{H}^{-1}J(\gamma, q)\}, \gamma_y) + 2J(\gamma_x, \partial_y\{\mathcal{H}^{-1}J(\gamma, q)\}) + J(\Lambda^{-1}R(\gamma), \nabla^2\gamma) + \\
&\quad + J(\gamma, \nabla^2\Lambda^{-1}R(\gamma)) + \nabla\cdot\{\nabla^2\gamma\nabla(\Lambda^{-1}G(\gamma))\} - \\
&\quad - \nabla^2[J(\gamma, \mathcal{H}^{-1}R(\gamma) - \gamma q) + J(\Lambda^{-1}R(\gamma), \gamma) + \nabla\cdot\{\gamma\nabla(\Lambda^{-1}G(\gamma))\}] \\
P(\gamma) &= \nabla^2\{(\mathcal{H}^{-1}R(\gamma) - \gamma q)q\} + \mathcal{H}^{-1}J(\mathcal{H}^{-1}J(\gamma, q), \nabla^2\gamma) + \\
&\quad + \mathcal{H}^{-1}J(\gamma, \nabla^2\mathcal{H}^{-1}J(\gamma, q)) - J(\Lambda^{-1}G(\gamma), \nabla^2\gamma) + \nabla^2J(\Lambda^{-1}G(\gamma), \gamma) + \\
&\quad + 2J(\partial_x(\Lambda^{-1}R(\gamma)), \gamma_y) + 2J(\gamma_x, \partial_y(\Lambda^{-1}R(\gamma))),
\end{aligned}
$$

where ∂_x and ∂_y denote partial derivatives with respect to x and y, respectively. Here we have used the relation $\gamma_t = -\mathcal{H}^{-1}J(\gamma, q)$ to eliminate the time derivatives of $\Omega^{(1)}$ and $D^{(1)}$, consistent with the accuracy of the model.

(*d*) *Height slaving*

As mentioned earlier, any variable that projects onto the slow modes at leading order can, in principle, be used as a slaving variable. From this perspective, there is nothing particularly exceptional about the potential vorticity. For example, if one chooses height as the slaving variable, then a mass-conserving slow dynamics will result if a slave relation is constructed in the manner suggested above and substituted into (29).

Allen (1993) used height as the slaving variable and implicitly used an iterative approximation of the superbalance equation (see section 3(d)) by iterating all variables except height. A direct comparison of our expansion procedure with Allen's iterative procedure reveals the similarity of the two methods in this case.

We start by rewriting the shallow-water equations (28) and (29) in the form (with $Fr = 1$)

$$\frac{1}{\varepsilon}(\mathbf{v} - \mathbf{k}\times\nabla\eta) = \mathbf{k}\times\left(\frac{\partial\mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right) \tag{48}$$

$$\frac{\partial\eta}{\partial t} + \nabla\cdot(\eta\mathbf{v}) + \frac{D}{\varepsilon} = 0. \tag{49}$$

This is not in the standard form of (8) and (9). Yet we may still regard (49) as the equation for the slaving variable η, and expand all variables except η. At the leading order we obtain geostrophic balance, i.e.

$$\mathbf{v}^{(0)} = \mathbf{k}\times\nabla\eta \qquad \nabla^2\psi^{(0)} = \nabla^2\eta \qquad D^{(0)} = 0. \tag{50}$$

At O(1) we obtain

$$\mathbf{v}^{(1)} = \mathbf{k}\times\left\{\frac{\partial\mathbf{v}^{(0)}}{\partial t} + (\mathbf{v}^{(0)}\cdot\nabla)\mathbf{v}^{(0)}\right\}, \tag{51}$$

with

$$\frac{\partial\eta}{\partial t} + \nabla\cdot(\eta\mathbf{v}^{(0)}) + D^{(1)} = 0. \tag{52}$$

The term $D^{(1)}$ may be eliminated from (52) by using (51) and (50), thus yielding the quasi-geostrophic system

$$(\nabla^2 - 1)\frac{\partial \eta}{\partial t} + J(\eta, \nabla^2 \eta - \eta) = 0 \tag{53}$$

as the leading-order model in the hierarchy. The slaving relations for the velocity, to $O(\varepsilon)$, are

$$D^{(1)} = -\nabla^2 \frac{\partial \eta}{\partial t} - J(\eta, \nabla^2 \eta) \tag{54}$$

$$\nabla^2 \psi^{(1)} = -2J(\eta_x, \eta_y) \tag{55}$$

where $\partial \eta / \partial t$ is evaluated from (53). The balance model to $O(\varepsilon)$ is obtained from the condition

$$\mathbf{v}^{(2)} = \mathbf{k} \times \left\{ \frac{\partial \mathbf{v}^{(1)}}{\partial t} + (\mathbf{v}^{(1)} \cdot \nabla)\mathbf{v}^{(0)} + (\mathbf{v}^{(0)} \cdot \nabla)\mathbf{v}^{(1)} \right\} \tag{56}$$

together with (49), taking $\mathbf{v} = \mathbf{v}^{(0)} + \varepsilon \mathbf{v}^{(1)}$, $D = \varepsilon D^{(1)} + \varepsilon^2 D^{(2)}$. Here $\mathbf{v}^{(0)}$, $\mathbf{v}^{(1)}$ and $D^{(1)}$ are known in terms of η from (50), (54) and (55), while $D^{(2)}$ may be obtained from (56). Since the height equation includes the term D/ε, the $O(\varepsilon)$-system requires the determination of $D^{(2)}$, for which one must evaluate the time derivative of $\nabla^2 \psi^{(1)}$. This complexity arises because height, unlike potential vorticity, is not unequivocally slow; in the linear equations height can undergo fast oscillations whereas potential vorticity cannot.

We now consider the iterative procedure. At leading order this yields geostrophic balance, i.e.

$$\mathbf{v}_0 = \mathbf{k} \times \nabla \eta \qquad \nabla^2 \psi_0 = \nabla^2 \eta \qquad D_0 = 0. \tag{57}$$

At O(1) we obtain

$$\mathbf{v}_1 = \mathbf{k} \times \nabla \eta + \varepsilon \mathbf{k} \times \left\{ \frac{\partial \mathbf{v}_0}{\partial t} + (\mathbf{v}_0 \cdot \nabla)\mathbf{v}_0 \right\} \tag{58}$$

with

$$\frac{\partial \eta}{\partial t} + \nabla \cdot (\eta \mathbf{v}_0) + \frac{D_1}{\varepsilon} = 0. \tag{59}$$

Elimination of D_1 in (59) via (58) gives the quasi-geostrophic equations (53) once again. The slaving relations, to $O(\varepsilon)$, are

$$D_1 = -\varepsilon \nabla^2 \frac{\partial \eta}{\partial t} - \varepsilon J(\eta, \nabla^2 \eta) \tag{60}$$

$$\nabla^2 \psi_1 = \nabla^2 \eta - 2\varepsilon J(\eta_x, \eta_y) \tag{61}$$

where $\partial \eta / \partial t$ in (60) is determined from (53). The $O(\varepsilon)$-model is obtained from the condition

$$\mathbf{v}_2 = \mathbf{k} \times \nabla \eta + \varepsilon \mathbf{k} \times \left\{ \frac{\partial \mathbf{v}_1}{\partial t} + (\mathbf{v}_1 \cdot \nabla)\mathbf{v}_1 \right\} \tag{62}$$

with

$$\frac{\partial \eta}{\partial t} + \nabla \cdot (\eta \mathbf{v}_1) + \frac{D_2}{\varepsilon} = 0. \tag{63}$$

The balance model (63) is identical to Allen's (1993) model (B.10).

The only difference between the expansion and iteration procedures at this order is that the condition (62), from which D_2 is determined, contains the cross-term $(\mathbf{v}^{(1)}\cdot\nabla)\mathbf{v}^{(1)}$ while the condition (56) does not. This leads to an extra (higher-order) term $\varepsilon^2\nabla\cdot(\mathbf{v}^{(1)}\nabla^2\psi^{(1)})$ in the iterative balance model (63). We conclude that the two procedures are largely similar, and have the same formal accuracy.

5. Discussion and conclusion

In this paper we have considered approximations to the shallow-water equations in the limit of small Rossby number. The conventional expansion in powers of the Rossby number has been shown to lead to a problem of secularities: if the lowest-order evolution is chaotic then variables which project onto the slow modes at leading order grow exponentially in time at first order, while the remaining variables grow at the next order. The expansion therefore ceases to be uniformly valid in time. This point has also been made by Allen (1993) when considering approximations to the dynamics of a rotating vortex; in that case the secular growth is only linear in time. Numerical solutions by Allen and Newberger (1993) of an unstable baroclinic jet also indicate a pronounced growth of inaccuracy for the second-order quasi-geostrophic model that is consistent with our findings.

Secularities are avoided if, instead of expanding all variables in an asymptotic series, only the fast variables are expanded. This is equivalent to introducing a slaving condition and then solving the resulting superbalance equation using a perturbation expansion. (Thus the accuracy of a given model refers to the accuracy of the balance relation, not of the balance dynamics: the problem of long-term accuracy of approximate models has not been solved.) Allen's (1993) 'iterated geostrophic intermediate models' represent another way of approximating the superbalance equation, with no iteration of the slow variables. The resulting models closely resemble ours. In the shallow-water equations, potential vorticity is a natural slow variable, whereas the divergence and geostrophic imbalance are natural fast variables. Like the bounded derivative method the expansion can be carried out to any order, in stark contrast to traditional 'truncations', such as the Charney balance equations, which are difficult to extend beyond first order. To highlight this feature, an explicit system has been presented which is accurate to second order in ε, namely to two orders beyond quasi-geostrophy.

The notion of a preferred field to which other variables are slaved is, of course, not new. Apart from Allen (1993), it is explicit, for instance, in the work of Hoskins *et al.* (1985) and M. E. McIntyre and W. A. Norton (personal communication). The 'hypo-geostrophic' model of McWilliams and Gent (1980) is also an example of this type of expansion; they did not expand the height variable, although they gave no formal justification for their procedure. The advantage of the present approach lies in the general framework that is presented wherein the expansion may be carried to an arbitrarily high order. Although any variable that has a slow component at leading order can be a candidate slaving variable, the expansion is likely to be simplest when the slow variable is orthogonal to the fast variables in the linear system; i.e. when $\mathbf{s} = 0$ for the fast oscillations. For the shallow-water equations, this suggests the choice of potential vorticity, or possibly of something that differs from the potential vorticity beyond zeroth order—for example the quasi-geostrophic potential vorticity. The interest in potential vorticity stems from its advective character, which is entirely consistent with the present perspective since advection is unambiguously 'slow' for small Rossby number.

The modified expansion procedure that we have presented is general, and is currently being applied to the stratified primitive equations. Under conventional Rossby number scaling, quasi-geostrophy will always emerge at lowest order, independently of the choice of the slaving variable, whereas the higher-order systems depend on this choice. This lack of uniqueness is highlighted by the fact that most of the conservation properties of the original system appear to be lost and only those directly associated with the slaving variable (e.g. potential vorticity or height) can easily be retained. This is clearly an unsatisfactory aspect of the present theory. It may be that the freedom in the choice of the slow variable can be used to enforce additional conservation properties. This is a possibility that remains to be explored.

Acknowledgements

The authors are grateful for the stimulating atmosphere of the 1993 Woods Hole GFD Summer School, where this work was developed. Helpful comments were received from John Allen and Roger Samelson. TW and TGS acknowledge support from NSERC and from AES. GKV acknowledges support from NSF (ATM 9317485) and ONR. OB acknowledges receipt of a Simcoe Special Fellowship.

References

Allen, J. S. 1993 Iterated geostrophic intermediate models. *J. Phys. Oceanogr.*, **23**, 2447–2461

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

Allen, J. S., Barth, J. and Newberger, P. 1990 On intermediate models for barotropic continental shelf and slope flow fields. Part I: formulation and comparison of exact solutions. *J. Phys. Oceanogr.*, **20**, 1017–1042

Baer, F. 1977 Adjustments of initial conditions required to suppress gravity oscillations in nonlinear flows. *Beitr. Phys. Atmos.*, **50**, 350–366

Browning, G. and Kreiss, H.-O. 1987 Reduced systems for the shallow water equations. *J. Atmos. Sci.*, **44**, 2813–2822

Browning, G., Kasahara, A. and Kreiss, H.-O. 1980 Initialization of the primitive equations by the bounded derivative method. *J. Atmos. Sci.*, **37**, 1424–1436

Charney, J. G. 1955 The use of the primitive equations of motion in numerical prediction. *Tellus*, **7**, 22–26

1963 A note on large scale motion in the tropics. *J. Atmos. Sci.*, **20**, 607–609

Craig, G. C. 1993 A scaling for the three-dimensional semi-geostrophic approximation. *J. Atmos. Sci.*, **50**, 3350–3355

Debussche, A. and Temam, R. 1991a Inertial manifolds and slow manifolds. *Appl. Math. Lett.*, **4**, 73–76

1991b Inertial manifolds and slow manifolds in meteorology. *Differential and Integral Equations*, **4**, 897–931

Guckenheimer, J. and Holmes, P. J. 1983 *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Springer-Verlag

Hoskins, B. J. 1975 The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmos. Sci.*, **32**, 233–242

Hoskins, B. J., McIntyre, M. E. and Robertson, A. 1985 On the use and significance of isentropic potential vorticity maps. *Q. J. R. Meteorol. Soc.*, **111**, 887–946

Kampen Van, N. G. 1985 Elimination of fast variables. *Phys. Rep.*, **124**, 69–160

Kreiss, H.-O. 1979 Problems with different time scales for ordinary differential equations. *SIAM J. Numer. Anal.*, **16**, 980–998

1980 Problems with different time scales for partial differential equations. *Comm. Appl. Math.*, **33**, 399–439

1985 Problems with different timescales. In *Multiple timescales*. Eds. J. Brackbill and B. Cohen. Academic

Leith, C. E.	1980	Nonlinear normal mode initialization and quasi-geostrophic theory. *J. Atmos. Sci.*, **37,** 399–439
Lorenz, E. N.	1980	Attractor sets and quasi-geostrophic equilibrium. *J. Atmos. Sci.*, **37,** 1685–1699
	1986	On the existence of the slow manifold. *J. Atmos. Sci.*, **43,** 1547–1557
Lynch, P.	1989	The slow equations. *Q. J. R. Meteorol. Soc.*, **115,** 201–219
Machenhauer, B.	1977	On the dynamics of gravity oscillations in a shallow water model with application to normal mode initialization. *Beitr. Phys. Atmos.*, **50,** 253–271
McWilliams, J. C. and Gent, P. R.	1980	Intermediate models of planetary circulations in the atmosphere and ocean. *J. Atmos. Sci.*, **37,** 1657–1678
Pedlosky, J.	1987	*Geophysical fluid dynamics*, 2nd ed. Springer-Verlag
Phillips, N. A.	1960	On the problem of initial data for the primitive equations. *Tellus*, **12,** 122–126
Vautard, R. and Legras, B.	1986	Invariant manifolds, quasigeostrophy, and initialization. *J. Atmos. Sci.*, **43,** 565–584
Warn, T. and Menard, R.	1986	Nonlinear balance and gravity-inertial wave saturation in a simple atmospheric model. *Tellus*, **38A,** 285–294