On the accuracy of the semi-geostrophic approximation

By M. J. P. Cullen*

European Centre for Medium-Range Weather Forecasts, UK

(Received 28 June 1999; revised 23 October 1999)

SUMMARY

The semi-geostrophic model has been widely used to understand atmospheric flows such as fronts and developing cyclones. However, there have been a number of demonstrations of its lack of accuracy. This paper presents theory and computations to demonstrate that the semi-geostrophic model is an accurate approximation to the primitive equations either on horizontal scales larger than the Rossby radius of deformation or when the ratio of horizontal to vertical scales is greater than $f/N$.

KEYWORDS: Accuracy Balance Semi-geostrophic

1. INTRODUCTION

The semi-geostrophic approximation to the primitive equations was introduced by Eliassen (1948) and developed and popularized by Hoskins (1975). The semi-geostrophic equations have a very simple structure which allows solution by analytic techniques using the geostrophic coordinate transformation, and by geometrical techniques. Cullen et al. (1987) showed that the geometrical solution technique allows the equations to give understanding of a wide range of mesoscale phenomena, such as the limit on inland penetration of sea-breezes, barrier jets upstream of orography, and the control of convective mass transport by the large-scale circulation.

Unfortunately, the semi-geostrophic model has been, to some extent, discredited by demonstrations of its lack of accuracy in solving problems where the primitive equations have smooth solutions and accurate comparison is possible. Examples include the comparison of solutions given by a set of different balanced models by Allen et al. (1990), and comparisons of baroclinic wave simulations by Snyder et al. (1991). The primitive equations in three dimensions have much more complex solutions than the shallow-water equations, including a wider variety of inertio-gravity waves and turbulence. The purpose of a simpler model is to restrict the solutions to a subset of real flows in order to aid understanding. We therefore do not want a highly accurate representation of the primitive-equation solutions (or indeed of the shallow-water equation solutions) by a simple model in all circumstances. However, we need estimates of accuracy in order to assess the usefulness of the predictions made by the simple models, in particular to assess in which asymptotic regimes the results are likely to be physically relevant.

The original assessment of accuracy by Hoskins (1975) was for axisymmetric vortices, the worst case for the semi-geostrophic approximation; the error was 10% for cyclonic vortices with Rossby number 0.55 and anticyclonic vortices with Rossby number 0.2. These estimates are consistent with the more recent assessments of accuracy referred to above. In particular, the data used by Allen et al. (1990) contain strong anticyclonic vortices which are shown to be the main source of error in the semi-geostrophic approximation.

In this paper we estimate the error by comparing the evolution of the potential vorticity with that of the equivalent potential vorticity derived from the primitive equations. Clearly a semi-geostrophic model cannot predict those parts of the primitive-equation solution which are independent of the potential vorticity. To give a complete picture, we also compare the complete solution of the primitive equations with that

* Corresponding address: European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, UK.
derived by diagnosing the potential vorticity from the complete solution and inverting to obtain the dynamical fields. This difference is a measure of the motions which are independent of the potential vorticity. For consistency, we carry out this procedure using the semi-geostrophic inversion procedure. The results show semi-geostrophic theory should be most accurate for length-scales $L$ greater than the Rossby radius of deformation $L_R$, or for aspect ratios less than $f/N$. In this regime the potential vorticity is dominated by variations in static stability. We then demonstrate by computations with a shallow-water model that the theoretical prediction that the error will be $O(ReBu^2)$ for $Bu = L_R/L \ll 1$ is observed in practice, where $Ro$ is the Rossby number and $Bu$ is the Burger number. A quasi-geostrophic model has $O(Re)$ errors for $Bu = O(1)$ and does not gain extra accuracy for small $Bu$ because of the use of a reference value of the static stability. We also compare the results with those given by a shallow-water version of the nonlinear balance equation. The semi-geostrophic model has larger errors for $Bu = O(1)$, in agreement with the studies referred to above, but the errors become comparable for $Bu \lesssim 0.2$. We also show that the rate of transfer of energy to \textquote{unbalanced} motions in the primitive equations is $O(ReBu^2)$, given initial data in exact balance and using the semi-geostrophic definition of \textquote{balance}.

The balance equations as used, for instance, by Holm (1996) and McWilliams et al. (1998) are accurate for all Burger numbers. For ease of reference, we analyse the balance equations using the same methods as applied above to semi-geostrophic theory. The split between balanced and unbalanced parts of the solution to the primitive equations is now different, because the inversion procedure is different. The rate of generation of unbalanced energy from balanced initial data is also different. The evolution error of the potential vorticity is shown to be $O(Re^3)$ for $Bu \ll 1$ or $O(Fr^3/Bu)$ for $Bu \gg 1$, where $Fr$ is the Froude number. However, the balance equations do not have such a simple solution structure, and spontaneous violations of the solvability conditions are possible. It remains to be demonstrated that they can give equivalent insights to those obtained by the semi-geostrophic model.

The main result is that semi-geostrophic theory is accurate (i.e. second-order or better) if one horizontal length-scale is larger than the Rossby radius of deformation, while it is asymptotically valid (i.e. first-order accurate) for all small Rossby number (and in fact small Froude number) cases. The other horizontal length-scale can be much smaller, according to the original frontal-scale analysis due to Hoskins and Bretherton (1972). Thus the mesoscale predictions discussed above should be useful provided they have one ratio of horizontal scale to vertical scale greater than $N/f$, e.g. mountain barrier length compared with barrier height, or the length of a straight segment of coastline compared with the depth of the sea-breeze circulation.

2. APPROXIMATION OF BOUSSINESQ INCOMPRESSIBLE FLOW BY BALANCED SYSTEMS

We use as \textquote{exact} equations the non-hydrostatic, Boussinesq equations in Cartesian geometry, with constant Coriolis parameter. These are sufficiently general to support the analyses carried out in this paper.

\[
\frac{\partial \mathbf{v}}{\partial t} + f k \times \mathbf{v} + \nabla \phi = \frac{\partial}{\partial t} k \\
\frac{\partial \theta}{\partial t} = 0 \\
\nabla \cdot \mathbf{v} = 0.
\]
The boundary conditions are that there is no normal flow at the rigid upper and lower boundaries $\delta \Gamma$ of a region $\Gamma$ in $(x, y, z)$, and there are periodic boundary conditions in $x$ and $y$. Here $z$ is a pseudo-height coordinate, as in Hoskins (1975), $v = (u, v, w)$ is the wind velocity, $\phi$ is the geopotential, and $\theta$ is the potential temperature, with basic state value $\theta_0$. $f$ and $g$ are the Coriolis parameter and acceleration due to gravity, and $k$ is a unit vector in the vertical direction. Energy conservation is expressed as

$$E = \int_\Gamma \left\{ \frac{1}{2}(u^2 + v^2 + w^2) - g \frac{\theta}{\theta_0} z \right\} \, d\tau = \text{a constant},$$

(2)

where $d\tau$ represents a volume element. These equations also conserve the Ertel potential vorticity

$$q = (\nabla \times v + f k) \cdot \nabla \theta.$$

(3)

We will also illustrate the theory by using the shallow-water form of the equations, using $h$ as the depth of the fluid:

$$\frac{Dv}{Dt} + f k \times v + g \nabla h = 0$$

$$\frac{\partial h}{\partial t} + \nabla \cdot hv = 0.$$  

(4)

The region of integration is now a periodic box $\Gamma$ in $(x, y)$. Energy conservation is expressed as

$$E = \int_\tau \left\{ \frac{1}{2}h(u^2 + v^2) + \frac{1}{2}gh^2 \right\} \, d\tau = \text{a constant},$$

(5)

where $d\tau$ is an area element. The potential vorticity is

$$q = \frac{(k \cdot \nabla \times v + f)}{h}.$$  

(6)

In either case, we can write the evolution equation for the potential vorticity as

$$\frac{\partial q}{\partial t} + v \cdot \nabla q = 0.$$  

(7)

In order to assess how well the solutions of these equations can be approximated by those of a simpler model, we must define specific asymptotic regimes characterized by a small parameter. Assume horizontal and vertical length-scales $L$ and $H$, and velocity scale $U$. The Brunt–Väisälä frequency for Eqs. (1) is $N^2 = (g/\theta_0) \partial \theta/\partial z$. Then the Rossby number is defined by $Ro = U/(f L)$, the Froude number for Eqs. (1) by $Fr = U/(NH)$, and for Eqs. (4) by $Fr = U/\sqrt{gh}$. The Burger number is defined as $Bu = Ro/Fr$. Note that only $f$, $g$, and $h_0$, the mean value of $h$, are invariant in time. The other scaling parameters may change in magnitude during the flow evolution. In particular, $Fr$ is quite robust for the shallow-water equations if the initial perturbations to $h$ are small compared with the mean value, since it then only varies with $U$. However, it is much less robust for the three-dimensional equations where $N$ and $H$ can vary enormously in time, due to wave breaking for instance.

We now assume that (1) or (4) is approximated by a balanced model which conserves potential vorticity. $q$ is taken as the 'slow' variable, as by Vallis (1996). In the balanced model, the velocity, written as $u_p$, can be derived from $q$ by an invertibility relation
(Hoskins et al. 1985). After initializing both models with the same \( q \), all the other fields required to solve the balanced model can be derived from \( q \). The initial data for the exact equations is assumed to satisfy the scaling assumed for the asymptotic regime chosen, and these scale assumptions are assumed to remain valid in time. Then, if the asymptotic regime is characterized by a small parameter \( \epsilon \), we can estimate the error in the evolution of the potential vorticity by seeking an estimate

\[
\mathbf{u} = \mathbf{u}_b[1 + O(\epsilon^n)].
\]

(8)

If this is satisfied, the evolution error in the potential vorticity will also be \( O(\epsilon^n) \). Since the potential vorticity is assumed to be an unapproximated slow variable, the error in its evolution is the important error to estimate in evaluating the accuracy of the balanced approximation. Another, separate, error measure is the accuracy to which a balanced state approximates the real state. One suitable measure of this is the difference between the total energy \( E \) of the real state and the energy \( e \) of a balanced state having the same potential vorticity.

This type of analysis, though useful in understanding the behaviour of the atmosphere in different regimes, falls short of a mathematically rigorous analysis, such as those carried out by Babin et al. (1996) and Embid and Majda (1996). A rigorous approach would be to initialize both sets of equations with the same values of all variables derived by the invertibility principle from a given \( q \). It would then be necessary to show that the solution of the exact equations diverged at a rate \( O(\epsilon^n) \) from that of the balanced equations, with given values of \( f \) and \( g \), but no constraints on horizontal or vertical scale other than those satisfied by the initial data. Such an analysis has to deal with the robustness of the asymptotic regimes. The ability of the exact equations to generate small scales very fast makes rigorous estimates difficult.

In order to understand how well the potential-vorticity evolution is approximated, it is helpful to relate it to the other variables. This can be done using classical geostrophic-advection theory which constructs a geostrophically balanced state with the same potential vorticity as the given data. If \( \text{Ro} \ll \text{Fr} \) and \( Bu \ll 1 \), the underlying balanced state has approximately the same pressure and potential-temperature field as the general state. We can also express this condition as \( L \gg L_R \), where \( L_R \) is the Rossby radius of deformation. If \( L \ll L_R \), so \( \text{Fr} \ll \text{Ro}, \text{Bu} \gg 1 \), the underlying balanced state has approximately the same vertical component of vorticity as the general state but a different pressure and potential-temperature field. This can be quantified using the theory of geostrophic adjustment for the linearized shallow-water equations set out by Haltiner and Williams (1980). Consider the case where the pressure is a function of \( x \) only. Let \( v \) and \( h \) be complex wave amplitudes. The balanced state is given by

\[
\begin{align*}
\nu_b &= \frac{1}{1 + (L_R/L)^2} \{ (L_R/L)^2 v + i(fL)^{-1} h \} \\
\eta_b &= \frac{1}{1 + (L_R/L)^2} \{ -i(L_R/L)^2 fLv + h \}.
\end{align*}
\]

(9)

Equations (9) show that, if \( Bu \ll 1 \),

\[
\eta_b = h[1 - O((L_R/L)^2)] + O((L_R/L)^2)h_v,
\]

(10)

where \( h_v \) is the pressure field deduced by the inverse geostrophic relation from the wind field. In the large Burger number case

\[
\nu_b = v[1 - O((L/L_R)^2)] + O((L/L_R)^2)v_g,
\]

(11)

where \( v_g \) is the geostrophic wind.

We will generalize these estimates in the next sections.
3. ACCURACY OF THE SEMI-GEOSTROPHIC MODEL

(a) Formulation and method of analysis

The semi-geostrophic approximation to (1) takes the form

\[
\begin{align*}
\frac{Dv_g}{Dt} + f k \times v + \nabla \phi &= 0 \\
\frac{D\theta}{Dt} &= 0 \\
\left( f v_g, -f u_g, g \frac{\theta}{\theta_0} \right) &= \nabla \phi \\
\nabla \cdot v &= 0.
\end{align*}
\] (12)

The third equation is a statement of the geostrophic and hydrostatic relations. These equations are valid for variable Coriolis parameter, but for the purposes of this paper we exploit the ability to analyse the equations in a simple way for the special case \( f \) constant. In that case we introduce the geostrophic and isentropic coordinates

\[
X \equiv (X, Y, Z) = \left( x + f^{-1} v_g, y - f^{-1} u_g, g f^{-2} \frac{\theta}{\theta_0} \right).
\] (13)

We can then rewrite (12) as

\[
\begin{align*}
\frac{DX}{Dt} &= (u_g, v_g, 0) = -f^{-1} k \times \nabla \phi \\
X &= \nabla \left( f^{-2} \phi + \frac{1}{2} (x^2 + y^2) \right) \\
\nabla \cdot v &= 0.
\end{align*}
\] (14)

We can interpret this equation as describing motion of particles in \((X, Y, Z)\) space with a velocity \((u_g, v_g, 0)\). This can be shown to be non-divergent as a function of \((X, Y, Z)\), so that the Jacobian

\[
q = \frac{\partial(X, Y, Z)}{\partial(x, y, z)}
\] (15)

is conserved following the motion. This is the potential-vorticity form of the equations introduced by Hoskins. Details of the derivation are given by Cullen and Purser (1989).

We now seek to estimate the accuracy of semi-geostrophic theory by comparing the evolution of \(q\) with that of an equivalent quantity derived from the primitive equations. The simplest approach would be to compare \(q\) as defined by (15) with the Ertel potential vorticity (3) calculated from a solution of the primitive equations. However, this introduces an unnecessary error, because these two quantities are different. Instead, we define

\[
X^* \equiv (X^*, Y^*, Z^*) = \left( x + f^{-1} v, y - f^{-1} u, g f^{-2} \frac{\theta}{\theta_0} \right).
\] (16)

Then we can rewrite (1) as

\[
\begin{align*}
\frac{DX^*}{Dt} + f^{-1} k \times \nabla \phi &= 0 \\
\frac{Dw}{Dt} + \frac{\partial \phi}{\partial z} &= g \frac{\theta}{\theta_0} \\
\nabla \cdot v &= 0.
\end{align*}
\] (17)
Equations (17) can be written as a conservation law for a Jacobian of a similar form to (15)

\[ q^* = \frac{\partial (X^*, Y^*, Z^*)}{\partial (x, y, z)}. \]  

(18)

Equation (18) is a Monge–Ampere equation which could be solved directly for the rest of the fields given suitable boundary conditions. However, Cullen and Purser (1984) showed that, when fronts form, they act as an intrusion of high potential vorticity into the interior of the fluid, and prevent conservation of the potential vorticity integrated over the physical domain. Conservation on each fluid particle remains. It is therefore better to reconstruct the rest of the fields by first inverting (18) to give

\[ \rho^* = \frac{\partial (x, y, z)}{\partial (X^*, Y^*, Z^*)}. \]  

(19)

This defines the volume of fluid which has specific values of \( X^* \). Then construct a state of the fluid satisfying

\[ (X^*, Y^*, Z^*) = \nabla P \equiv \nabla \left\{ f^{-2} \phi^* + \frac{1}{2} (x^2 + y^2) \right\}, \]  

(20)

with \( P \) convex, by carrying out a projection which moves fluid particles from positions \( x \) to new positions \( x^* \) in \( \Gamma \) while preserving their values of \( X^* \). Write this projection as

\[
\Pi (u, v, w, \theta) = (u_b, v_b, 0, \theta_b)
\]

\[ x^* = x + (\xi, \eta, \zeta) \]

\[ u_b (x^*) = u (x) + f \eta \]

\[ v_b (x^*) = v (x) - f \xi \]

\[ \theta_b (x^*) = \theta (x), \]  

(21)

where the projection is generated by a displacement \( \Xi = (\xi, \eta, \zeta) \) of the fluid which satisfies \( J (x + \Xi, x) = 1 \). The only boundary condition that can be specified is the statement

\[ (x^*, y^*, z^*) \in \Gamma, \]  

(22)

which is ensured by setting \( \Xi \cdot n = 0 \) on \( \delta \Gamma \). This is significantly different from the boundary conditions required in other forms of potential-vorticity inversion. Shutts and Cullen (1987) showed that this can be interpreted as minimizing the energy integral

\[ E = \int_{\tau} \left\{ \frac{1}{2} (u^2 + v^2) - g \frac{\theta}{\theta_0} z \right\} \, d\tau. \]  

(23)

with respect to particle displacements that conserve \( X^* \).

For small departures from balance, the displacement can by estimated by a linearization. Set

\[ \nabla^2 p = \nabla \cdot \left( f v, -f u, g \frac{\theta}{\theta_0} \right) \]

\[ \frac{\partial p}{\partial n} = \left( f v, -f u, g \frac{\theta}{\theta_0} \right) \cdot n \quad \text{on} \ \delta \Gamma \]  

(24)

\[ \Xi = (\xi, \eta, \zeta) = -f^{-2} \left\{ \nabla p - \left( f v, -f u, g \frac{\theta}{\theta_0} \right) \right\}. \]
Then (21) shows that, if the displacement is small compared with the scale on which \( \mathbf{u} \) varies and does not affect the pressure field, then

\[
(\mathbf{u}_b(\mathbf{x}^*), \mathbf{v}_b(\mathbf{x}^*)) = f^{-1} \left( -\frac{\partial p}{\partial y}, \frac{\partial p}{\partial x} \right).
\]  
(25)

In the general case, the solution will have to be obtained by iteration. Because there will usually be changes to the pressure, the displacements in the iteration will have to be scaled by a factor of order \( \omega^{-2} \), where \( \omega \) is the fastest inertio-gravity wave frequency resolved, rather than by \( f^{-2} \) as in (24).

Writing the original, conserved, energy of the primitive-equation solution as \( E \), and the energy after minimization as \( e^* \), then \( (E - e^*) \) is a measure of the unbalanced energy, where balance is defined by semi-geostrophic theory. The change to the energy in each iteration is

\[
\delta E = \alpha \int_\Gamma \left\{ -f^2 u^2 - f^2 v^2 - \left( g \frac{\theta}{\theta_0} \right)^2 \right\} + \left( f u, -f u, \frac{\theta}{\theta_0} \right) \cdot \nabla p \, d\tau.
\]  
(26)

Since, for small displacements, \( \nabla \cdot \mathbf{\Xi} = 0 \), we have

\[
\int_\Gamma \left( \mathbf{\Xi} \cdot \nabla p \right) \, d\tau = 0.
\]  
(27)

We can then add \( f^2 \) times \( \int_\Gamma (\mathbf{\Xi} \cdot \nabla p) \, d\tau \) to (26) and substitute from (24) to obtain

\[
\delta E = \int_\Gamma - \left\{ \left( f u, -f u, \frac{\theta}{\theta_0} \right) - \nabla p \right\}^2 \, d\tau.
\]  
(28)

This is negative definite, and vanishes when the balanced state is reached. The relative changes to the wind and mass fields can be estimated by considering the effect on the two sides of the thermal wind equation, \( f(\partial v/\partial z) \) and \( (g/\theta_0)(\partial \theta/\partial x) \). The local change in \( v \) is approximately \( (\partial v/\partial x) + f \xi \), and in \( \theta \) is approximately \( \zeta (\partial \theta/\partial z) \). Assuming \( \xi/\zeta \) has magnitude \( L/H \), the relative magnitude of the changes to the thermal wind balance is \( 1/Bu^2 \). Thus the change to the pressure field in the low Burger number regime is \( O(Bu^2) \) as given by the linear theory.

The magnitude of the displacement \( \mathbf{\Xi} \) is determined by the amount by which the data fails to satisfy the geostrophic and hydrostatic relations themselves, which is \( O(Ro) \). If \( L \gg L_R \), the projection primarily alters the velocity field. The correction required will be of order \( RoU \), where \( U \) is the velocity-scale, and the horizontal displacement required will thus be \( \xi = O(RoU/f) \). Thus \( \xi/L = O(Ro)^2 \).

Note that, following Shutts and Cullen (1987), (26) can be used to deduce

\[
\delta^2 E = \alpha \int_\Gamma \mathbf{\Xi} \mathbf{\Lambda} \mathbf{\Xi}^T \, d\tau
\]  
(29)

at the minimization point \( E = e^* \), where \( \det \mathbf{\Lambda} = q^* \) and \( \mathbf{\Lambda} \) is the matrix with elements \( \partial X_i/\partial x_j \). Thus we can estimate

\[
E - e^* \leq \alpha \int_\Gamma \lambda |\mathbf{\Xi}|^2 \, d\tau,
\]  
(30)

where \( \lambda \) is the largest eigenvalue of \( \mathbf{\Lambda} \).
(b) Accuracy of the evolution of potential vorticity

To estimate the error in the evolution implied by semi-geostrophic theory, assume that we have computed a solution of the primitive equations (17), with particle positions \( x(t) \), and particle values \( X^*(t) \). At each time \( t \), we project to the balanced state using (21). This involves displacing the particles to new positions \( x^*(t) \) while preserving their values of \( X^*, Y^* \) and \( Z^* \). Writing \( D^*/Dt \) to express a derivative following the ‘minimum energy’ particle positions \( x^* \), and writing the ‘velocity’ that achieves this as \( \mathbf{V} = (U, V, W) \), the condition \( J(x + \varepsilon, x) = 1 \) then implies \( \nabla \cdot \mathbf{V} = 0 \). The boundary conditions (22) imply that \( \mathbf{V} \cdot \mathbf{n} = 0 \) on \( \delta \Gamma \). Then the equations expressing the evolution of this balanced state are

\[
\frac{D^* X^*}{Dt} + (f^{-1} \mathbf{k} \times \nabla \phi)^* = 0
\]

\[
\frac{D^* \alpha}{Dt} = 0
\]

\[
(X^*, Y^*, Z^*) = \nabla P.
\]

The * after the term in \( \nabla \phi \) indicates that this term is calculated at \( x(t) \) from the solution of (17), and added to the solutions of (31) at positions \( x^* \). This means that (31) cannot be solved without prior knowledge of the solution of (17). However, (31) can be solved uniquely once \( (\mathbf{k} \times \nabla \phi)^* \) is specified. The displacement means that \( \nabla \phi \) is in general no longer the gradient of a scalar and, thus, (31) does not conserve energy. Thus, the balanced energy \( e^* \) can change in time. Note that this formulation is very close to the generalized Lagrangian mean form of equations introduced by Andrews and McIntyre (1978) and Buhler and McIntyre (1998) in the context of understanding wave/mean-flow interaction.

Equations (31) represent the exact evolution of \( q^* \). The semi-geostrophic evolution of a state with potential vorticity \( q^* \) can be obtained according to (12) by replacing the term \( (\mathbf{k} \times \nabla \phi)^* \) in (31) with \( \mathbf{k} \times \nabla \phi_b \), where \( \phi_b = f^2 \{ P - \frac{1}{2} (x^2 + y^2) \} \) and \( P \) is the scalar potential appearing in (31).

Write

\[
(\nabla \phi)^* = \nabla \phi - \Xi \cdot \nabla (\nabla \phi) + O(|\Xi|^2).
\]

Using \( \nabla \phi = O(f U) + O(Ro) \), and the previous estimate \( \xi = O(Ro^2 L) \), the second term can be estimated as \( Ro^2 L \times f U/L \). We can then say that

\[
(\nabla \phi)^* - \nabla \phi_b = \nabla \phi[1 + O(Ro^2)] - \nabla \phi_b.
\]

As shown in the previous subsection, the difference \( \nabla (\phi - \phi_b) \) can be estimated as \( RoBu^2 \nabla \phi_b \). Thus the overall error in the prediction of the balanced motion is \( O(Ro^2) \), provided that \( Bu^2 \leq Ro \). For a Rossby number of 0.1, this requires a Froude number of greater than 0.3. If this corresponds to a wind of 10 m s\(^{-1}\) and a length-scale of 1000 km, this implies a vertical scale less than 3.3 km for \( N^2 = 10^{-2} \text{s}^{-2} \).

In our computations we will take \( Ro \) fixed at 0.04, and vary \( Bu \) from 1 to 0.2, so that we cover the range \( Bu = O(1) \) to \( O(\sqrt{Ro}) \). In this range the prediction is that the error will be \( O(Ro Bu^2) \), and we will demonstrate the second order dependence on \( Bu \) for fixed \( Ro \).

(c) Rate of energy transfer between balanced and unbalanced flow

We first estimate the relative contributions to \( E \) of the kinetic- and potential-energy terms in (23). The available potential-energy terms can be estimated from the
difference between $\int g \theta z / \theta_0 \, d\tau$ and the same quantity evaluated from a minimum energy configuration with $\theta$ a function of $z$ only. This can be shown to give $(NH)^2 \phi_b$, where $\phi_b$ measures the horizontally varying part of the pressure field. The kinetic energy density can be estimated using the geostrophic relation as $(fL)^{-2} \phi_b$. The kinetic energy is thus $O(Bu^2)E$.

The evolution of the balanced energy $e^*$ can be calculated from (31). Replace $X^*$ by the original variables, multiply the first equation by $(u_b, v_b)$ and the second by $z$ and add. Using $W = D^* z / D\tau$, and the fact that $\nabla \cdot \nabla s = -f U v_b + f V u_b - g W / \theta_0$ integrates to zero, we find that

$$\frac{de^*}{dt} = \int (u_b, v_b, 0) \cdot (\nabla \phi)^* \, d\tau.$$

This is zero if $(\nabla \phi)^*$ can be written as $\nabla \pi$ for some scalar $\pi$, because $\nabla \cdot (u_b, v_b, 0) = 0$. In general, (33) gives $(\nabla \phi)^* = \nabla \phi [1 + O(Ro^2)]$. Thus we have

$$\frac{de^*}{dt} = O(f Ro^2) \int (u_b^2 + v_b^2) \, d\tau$$

$$= O(f (Ro Bu)^2) E. \quad (35)$$

This suggests that energy will only transfer between the balanced and unbalanced flow on a time-scale $f^{-1} (Ro Bu)^{-2}$. Thus in well initialized computations, the rate of generation of unbalanced energy should be $O(Ro Bu^2)$. We test this in our computations also.

Calculating the equilibrium value of $E - e^*$ that is consistent with the scaling, since $v - v_b = O(Ro) v$, the kinetic energy difference is $O(Ro) v^2 = O(Ro Bu^2) E$. The potential-energy difference can be shown to be of the same order. Thus, $E - e^* = O(Ro Bu^2) E$. Combining this with Eq. (35) shows that $E - e^*$ varies on the 'slow' time-scale $(Ro f)^{-1}$.

Using (32), we can also seek a sharper estimate by calculating the condition that $(\nabla \phi)^* = \nabla \pi$ to leading order in $\mathcal{E}$. This gives

$$\frac{\partial \xi}{\partial y} \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial \eta}{\partial y} \frac{\partial^2 \phi}{\partial x \partial y} - \frac{\partial \xi}{\partial x} \frac{\partial^2 \phi}{\partial y \partial x} - \frac{\partial \eta}{\partial x} \frac{\partial^2 \phi}{\partial y^2} = 0. \quad (36)$$

This is satisfied for flows independent of one horizontal coordinate. This fits with the original semi-geostrophic scaling introduced by Hoskins and Bretherton (1972), which permits $L$ to be arbitrarily small in one direction as long as it stays large in the other. The energy estimate shows that a genuinely two-dimensional flow will conserve its balanced energy, and thus the partition of energy between balanced and unbalanced motion is fixed for all time.

It will be of interest to seek other cases where (36) is either zero, or smaller than would be expected from general estimates. In such cases $de^*/dt$ will be small and, if $E - e^*$ is initially small, it will only grow slowly so that the flow stays closer to balance than would normally be expected.

The accuracy estimates above agree with the standard asymptotic error estimate, showing that the latter is consistent. This is because the terms neglected in semi-geostrophic theory are $Du_{ag} / D\tau$. Since $u_{ag} = O(Ro) u_g$ and $D / D\tau = O(Ro) f$, the neglected terms are $O(Ro)^2$ smaller than the largest retained terms.
4. Accuracy of the balance equations

(a) Formulation and method of analysis

We summarize the accuracy estimates for the balance equations using a similar framework to that set out above for semi-geostrophic theory. This helps to relate the behaviour of the two models. Vallis (1996) also presents analyses of various forms of the balance equation based on potential vorticity.

Write the decomposition of the velocity field into rotational and divergent parts as
\[
\mathbf{v} = (u_r, v_r, 0) + (u_d, v_d, w)
\]
\[
\nabla_h \cdot (u_r, v_r, 0) = 0
\]
\[
\nabla_h \times (u_d, v_d, 0) = 0.
\]

Since \( \nabla \cdot \mathbf{v}_d = 0 \), \( w \) is the only independent variable in \( \mathbf{v}_d \), with \( u_d \) and \( v_d \) determined from it. Following Holm (1996), we write the Hamiltonian form of the balance equations as
\[
\frac{D}{Dt} (u_r, v_r) + (u_r \nabla_h u_d + v_r \nabla_h v_d) + (-f v, f u) + \nabla_h \phi = 0
\]
\[
\frac{D \theta}{Dt} = 0
\]
\[
\frac{\partial \phi}{\partial z} - g \frac{\theta}{\theta_0} + v_r \cdot \frac{\partial v_d}{\partial z} = 0.
\]

These equations conserve an energy integral (with only the rotational wind appearing in the kinetic energy), and the potential vorticity
\[
q_{BE} = (\nabla \times \mathbf{v}_r + f \mathbf{k}) \cdot \nabla \theta
\]
in a material sense. \( \mathbf{v}_d \) is determined implicitly. If \( q_{BE} \) is given, \( \mathbf{v}_r \) and \( \theta \) can be calculated by using the diagnostic relation obtained from taking the horizontal divergence of (38). \( \mathbf{v}_d \) can be determined implicitly from the time derivative of this relation. Write these calculations in the form
\[
(\mathbf{v}_r, \mathbf{v}_d, \theta) = B(q_{BE}).
\]

Then a natural projection from the solutions of the primitive equations (1) to those of (38) is to set
\[
\Pi_{BE}(\mathbf{v}, \theta) = B(q).
\]

(b) Accuracy of the evolution of potential vorticity

If Eqs. (1) and (38) are started with the same potential vorticity, the error in evolution will simply be the error in the advecting velocity, since both models conserve potential vorticity. Holm (1996) (Eq. (5.10)) showed that taking the horizontal divergence and vertical derivative of the momentum equation in (38) gives an omega equation of the form
\[
-S(z) \nabla^2 w - f^2 \frac{\partial^2 w}{\partial z^2} + \text{LOT} = 0,
\]
where \( S(z) \) is of order \( N^2 \), and LOT denotes terms of lower order in derivatives of \( w \). The largest terms in LOT will typically be \( \nabla^2 (v_r \cdot \nabla \theta) \) and \( f \partial / \partial z (\nabla v_r \cdot \nabla v_r) \). Assuming
that the magnitude of these terms can be related by the thermal wind equation, they are comparable and of order \( fU^2/(HL^2) \). The coefficients of \( w \) in (38) have magnitudes \( N^2/L^2 \) and \( f^2/H^2 \), with ratio \( Ro^2: Fr^2 \). Thus for \( Bu \ll 1 \), we have \( w \approx Ro(UH/L) \), and for \( Bu \gg 1 \), \( w \approx (Fr/Bu)(UH/L) \).

If we derive an equation of the form (42) directly from (1), we obtain

\[
\frac{\partial^2 w}{\partial t^2} - \frac{\partial^2 w}{\partial z^2} - S(z) \nabla^2 w - f^2 \frac{\partial^2 w}{\partial z^2} + \text{LOT} = 0. \tag{43}
\]

If the solution of (42) is substituted directly into (43), and we assume that the time variations in LOT are on an advective timescale \( U/L \), then the error in \( w \) resulting from using (42) is \( O(Fr^2) \) if \( Bu \gg 1 \) and \( O(Ro^2) \) if \( Bu \ll 1 \). Thus the error in the evolution of \( q \) will be, respectively, \( O(Fr^3/Bu) \) and \( O(Ro^3) \). For \( Bu = 1 \) the error is \( O(Ro^3) = O(Fr^3) \).

In the special case of the shallow-water equations, the advective term \( v \cdot \nabla h \) is \( O(Ro) \) times the expected value, because geostrophic advection of the height field is zero. The solution for \( Bu \ll 1 \) is then close to a steady state, because the height is the slow variable. Thus the neglect of the time derivative in (43) will have less effect and the solution will be \( O(Ro^2) \) more accurate than expected. This argument does not apply to three-dimensional low Burger number flows that are baroclinic, as the advection of potential temperature by the geostrophic wind will not vanish.

In general, (43) has fast-wave solutions which will grow as a result of the time dependence of LOT, even if absent from the initial data. The asymptotic analysis of Ford et al. (2000) shows that their growth is at \( O(Fr^2) \), but their interaction back on the potential vorticity, which only depends on the time-averaged effect of the waves, is at \( O(Fr^4) \). Applying the same argument for \( Bu \ll 1 \) shows that the rate of growth of inertio-gravity wave energy will be \( O(Ro^2) \). Embid and Majda (1996) and Babin et al. (1996, 1997) carried out rigorous analyses of this effect using time averaging over the fast waves.

5. Computational tests

(a) Numerical models used

The experiments are designed to test the predictions that the errors in semi-geostrophic theory scale as \( O(Bu^2) \) for fixed \( Ro \), and the rate of generation of unbalanced energy from initialized data is also \( O(Bu^2) \). We use a set of spherical shallow-water models. The semi-geostrophic model is that described and used by Mawson (1996). It uses semi-Lagrangian advection of the primitive variables \((u_g, v_g, h)\), and an implicit method of calculating \((u, v)\) to ensure that the geostrophic relation is satisfied at each new time level. The variables \((u, v, h)\) are held on a C-grid, and \((u_g, v_g)\) are held on a D-grid. The implicit equations are solved by a multigrid method. The data are initialized by first choosing analytic height and geostrophic wind fields, and then carrying out the discrete initialization procedure set out by Mawson (1996). The initial values of \( u \) and \( v \) are set by making an initial time step, and calculating the \( u \) and \( v \) needed to preserve geostrophic balance.

The primitive-equation model is a shallow-water version of the implicit version of the UK Meteorological Office model (Cullen et al. 1997). It uses a C-grid, semi-Lagrangian advection, and a multigrid solution of the implicit equations. The nonlinear balance equations are solved by adapting the primitive-equation code. Equation (42) in shallow-water form becomes

\[
g h_0 \nabla^2 \nabla \cdot v - f^2 \nabla \cdot v + \text{LOT} = 0. \tag{44}
\]
This is used to calculate the divergence at each time step, and then the condition \( v_{d}^{t+\delta t} = v_{d}^{t} \) is enforced by correcting \( v \), and \( h \) iteratively, while conserving the potential vorticity. This can be achieved by setting

\[
\begin{align*}
g h_0 \nabla^2 D - f^2 D &= \nabla \cdot \left( v_{d}^{(n)\epsilon+\delta t} - v_{d}^{(n)} \right)/\delta t^2 \\
\nabla \cdot (U, V) &= D \\
h^{(n+1)} &= h^{(n)} - h_0 D \delta t \\
u^{(n+1)} &= u^{(n)} + f V \delta t \\
v^{(n+1)} &= v^{(n)} - f U \delta t
\end{align*}
\]

\( D \) is a correction with the same dimensions as a divergence.

The initial values are set by carrying out a short initial time step and performing the necessary iterative solution.

\( (b) \) Experiment design

The initial data are chosen to give a wave-number 2 pattern in each hemisphere, with no height perturbation close to the equator (Fig. 1). This is legal initial data for the semi-geostrophic model, which has to have inertially stable data. The data are initialized for the semi-geostrophic model, and then passed to the primitive-equation model. Both models are run for 2 days. The results from the primitive-equation model are then initialized using the semi-geostrophic initialization procedure, and compared with the semi-geostrophic results. This gives the evolution error of the semi-geostrophic model. The unbalanced energy in the primitive equation model is estimated from the difference made by initializing the day-2 results. The procedure is shown diagrammatically in Fig. 2. The evolution error is \( B - V \), and the unbalanced part of the primitive-equation solutions is \( P - V \).

The same design is used to measure the error of the nonlinear balanced model. The initialized data used in the first experiment is re-initialized using the nonlinear balance model. The primitive and nonlinear balanced models are run for 2 days and
Figure 2. Experimental set-up: A—analytic initial data, I—data initialized for balanced model, B—forecast using balanced model, P—forecast using primitive-equation model, V—initialized end state from primitive-equation model.

<table>
<thead>
<tr>
<th>$\phi_0$ (m)</th>
<th>$Re = U/fL$</th>
<th>$Fr = U/\sqrt{(g\phi_0)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5760</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>2880</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>1440</td>
<td>0.04</td>
<td>0.08</td>
</tr>
<tr>
<td>720</td>
<td>0.04</td>
<td>0.11</td>
</tr>
<tr>
<td>360</td>
<td>0.04</td>
<td>0.16</td>
</tr>
<tr>
<td>180</td>
<td>0.04</td>
<td>0.22</td>
</tr>
</tbody>
</table>

the results compared, as above. Since the two initialization procedures are different, the two primitive-equation runs are also slightly different.

The base resolution for the experiments was a latitude–longitude grid with 96 points around latitude circles and 65 points between the poles. The results were also generated using a higher resolution of 192 × 129 points. Though the general behaviour as the parameters were varied was the same for both resolutions, in the semi-geostrophic tests there were significant differences in individual results. The complete semi-geostrophic experiment was thus also run at the higher resolution, and a further check carried out using a 288 × 193 grid for one set of parameters. The nonlinear balance model is more compatible with the primitive-equation model. Sample runs with the 192 × 129 grid showed that it was unnecessary to repeat the whole experiment at higher resolution.

The experiments were designed to test the effect of varying the Burger number for fixed Rossby number. Thus, the same perturbation height field was used for all runs, but the mean value was varied from 5760 m down to 182.5 m. The amplitude of the superposed wave was ±170 m, so that the lowest mean value used is just sufficient to avoid the height becoming zero. The horizontal velocity had a maximum value of about 10 m s$^{-1}$, the gravity wave speed associated with the mean height varied from 240 m s$^{-1}$ to 42 m s$^{-1}$. Table 1 lists the values used, together with typical Froude and Burger numbers.

(c) Results

The results for the height-evolution errors are shown in Fig. 3, with the wind errors in Fig. 4. The results for the measures of imbalance are shown in Figs. 5 and 6. The experiment was carried out as above with one exception. It proved impossible to initialize the day-2 results from the primitive-equation model at high resolution and large values of the mean height using the semi-geostrophic initialization scheme, which was based on a local iteration. This was because of the development of significant height
Figure 3. Root-mean-square height-evolution errors (m) after 48 hours plotted against gravity-wave speed (m s\(^{-1}\)). Stars indicate balance-equation results on a 192 \(\times\) 129 grid, the diamond a semi-geostrophic result on a 288 \(\times\) 193 grid.

Figure 4. Root-mean-square wind-evolution errors (m s\(^{-1}\)) after 48 hours plotted against gravity-wave speed (m s\(^{-1}\)). Notation as Fig. 3.
Figure 5. Root-mean-square height imbalances (m) after 48 hours plotted against gravity-wave speed (m s\(^{-1}\)). Notation as Fig. 3.

Figure 6. Root-mean-square wind imbalances (m s\(^{-1}\)) after 48 hours plotted against gravity-wave speed (m s\(^{-1}\)). Notation as Fig. 3.
perturbations at the equator and the availability of only a local iteration to remove negative potential vorticity, as described by Mawson (1996). A different, non-local, initialization algorithm would be required to resolve this problem. The two parts of the error are therefore plotted together in the relevant parts of Figs. 3 to 6. Results for individual cases run at higher resolution are also plotted to validate the results.

It is readily seen that the \(O(Bu^2)\) behaviour expected for the evolution error of the semi-geostrophic model is clearly demonstrated. The nonlinear balanced model also shows errors reducing with \(Bu\) for \(Bu \ll 1\). As discussed in section 4, this is because the solution is close to a steady state. The much lower errors for the balanced model for larger values of mean height are consistent with the results of Allen et al. (1990). The two models have comparable errors once the gravity-wave speed is of the order of 50 m s\(^{-1}\).

The measures of imbalance also show the expected behaviour. In the semi-geostrophic case the \(O(Bu^2)\) behaviour is clearly seen at high resolution, though the low-resolution results are less reliable in this measure. In the balance-equation results, the rate of generation of gravity-wave energy clearly reduces for small \(Bu\).

6. Conclusions

We have demonstrated the expected rate of convergence for small Burger number in the semi-geostrophic model, and shown that the errors are comparable to those of the nonlinear balance equation in the small Burger number regime. This follows from the use of the exact potential temperature and continuity equations. Quasi-geostrophic theory uses a reference-state static stability. Semi-geostrophic theory is thus an improvement on quasi-geostrophic theory in this regime. The approximations made to the vorticity equation in semi-geostrophic theory are shown to be relatively unimportant for small Burger number. Since, in addition, semi-geostrophic theory is also accurate in the ‘frontal’ case where one horizontal length-scale can be small, it improves on quasi-geostrophic theory in two important respects. These results should help to formalize the applicability of the wide range of analytic and geometric predictions that can be made using semi-geostrophic theory.

We note, in addition, that neither model shows greater accuracy in predicting the potential vorticity than in predicting the total evolution. The rate of growth of imbalance from balanced initial data, and the evolution errors in the potential-vorticity evolution are of comparable size for all cases tested.

Acknowledgements

The author wishes to thank staff at The Met. Office for making the shallow-water codes available and providing advice on their use. The theoretical part of the work was done during the Isaac Newton Institute programme on the ‘Mathematics of Atmosphere/Ocean Dynamics’ held from June to December 1996. The author also wishes to thank two anonymous referees for very constructive comments.

References


Eliassen, A. 1948 The quasi-static equations of motion. *Geofys. Publ.*, 17, no. 3


Haltiner, G. J. and Williams, R. T. 1980 *Numerical prediction and Dynamic Meteorology*, 2nd edition, John Wiley and Sons


McWilliams, J. C., Yavneh, I., Cullen, M. J. P. and Gent, P. R. 1998 The breakdown of large-scale flows in rotating, stratified fluids. *Phys. Fluids*, 10, 3178–3184


