

Local Mass Conservation and Velocity Splitting in PV-Based Balanced Models. Part II: Numerical Results

ALI R. MOHEBALHOJEH

*School of Mathematics and Statistics, University of St Andrews, St Andrews, United Kingdom, and Institute of Geophysics,
University of Tehran, Tehran, Iran*

MICHAEL E. MCINTYRE

Centre for Atmospheric Science, Department of Applied Mathematics and Theoretical Physics, Cambridge University, Cambridge,
United Kingdom*

(Manuscript received 11 January 2006, in final form 27 July 2006)

ABSTRACT

The effects of enforcing local mass conservation on the accuracy of non-Hamiltonian potential-vorticity-based balanced models (PBMs) are examined numerically for a set of chaotic shallow-water f -plane vortical flows in a doubly periodic square domain. The flows are spawned by an unstable jet and all have domain-maximum Froude and Rossby numbers $Fr \sim 0.5$ and $Ro \sim 1$, far from the usual asymptotic limits $Ro \rightarrow 0$, $Fr \rightarrow 0$, with Fr defined in the standard way as flow speed over gravity wave speed. The PBMs considered are the plain and hyperbalance PBMs defined in Part I. More precisely, they are the plain- $\delta\delta$, plain- $\gamma\gamma$, and plain- $\delta\gamma$ PBMs and the corresponding hyperbalance PBMs, of various orders, where “order” is related to the number of time derivatives of the divergence equation used in defining balance and potential-vorticity inversion. For brevity the corresponding hyperbalance PBMs are called the hyper- $\delta\delta$, hyper- $\gamma\gamma$, and hyper- $\delta\gamma$ PBMs, respectively. As proved in Part I, except for the leading-order plain- $\gamma\gamma$ each plain PBM violates local mass conservation. Each hyperbalance PBM results from enforcing local mass conservation on the corresponding plain PBM. The process of thus deriving a hyperbalance PBM from a plain PBM is referred to for brevity as plain-to-hyper conversion. The question is whether such conversion degrades the accuracy, as conjectured by McIntyre and Norton.

Cumulative accuracy is tested by running each PBM alongside a suitably initialized primitive equation (PE) model for up to 30 days, corresponding to many vortex rotations. The accuracy is sensitively measured by the smallness of the ratio $\epsilon = \|Q_{\text{PBM}} - Q_{\text{PE}}\|_2 / \|Q_{\text{PE}}\|_2$, where Q_{PBM} and Q_{PE} denote the potential vorticity fields of the PBM and the PEs, respectively, and $\|\cdot\|_2$ is the L_2 norm. At 30 days the most accurate PBMs have $\epsilon \approx 10^{-2}$ with PV fields hardly distinguishable visually from those of the PEs, even down to tiny details. Most accurate is defined by minimizing ϵ over all orders and truncation types $\delta\delta$, $\gamma\gamma$, and $\delta\gamma$. Contrary to McIntyre and Norton’s conjecture, the minimal ϵ values did not differ systematically or significantly between plain and hyperbalance PBMs. The smallness of ϵ suggests that the slow manifolds defined by the balance relations of the most accurate PBMs, both plain and hyperbalance, are astonishingly close to being invariant manifolds of the PEs, at least throughout those parts of phase space for which $Ro \leq 1$ and $Fr \leq 0.5$.

As another way of quantifying the departures from such invariance, that is, of quantifying the fuzziness of the PEs’ slow quasimanifold, initialization experiments starting at days 1, 2, . . . 10 were carried out in which attention was focused on the amplitudes of inertia-gravity waves representing the imbalance arising in 1-day PE runs. With balance defined by the most accurate PBMs, and imbalance by departures therefrom, the results of the initialization experiments suggest a negative correlation between early imbalance and late cumulative error ϵ . In such near-optimal conditions the imbalance seems to be acting like weak background noise producing an effect analogous to so-called stochastic resonance, in that a slight increase in noise level brings PE behavior closer to the balanced behavior defined by the most accurate PBMs when measured cumulatively over 30 days.

* The Centre for Atmospheric Science is a joint initiative of the Department of Chemistry and the Department of Applied Mathematics and Theoretical Physics (more information available online at www.atm.damtp.cam.ac.uk/).

Corresponding author address: A. R. Mohebalhojeh, School of Mathematics and Statistics, University of St Andrews, North Haugh, St Andrews KY16 9SS, United Kingdom.
E-mail: arm@mcs.st-and.ac.uk

1. Introduction

The problem of local mass conservation in non-Hamiltonian potential vorticity (PV)-based balanced models, called PBMs for brevity, was addressed in Mohebalhojeh and McIntyre (2007, hereafter Part I). It was proved that the highly accurate “plain PBMs” studied previously by McIntyre and Norton (2000, hereafter MN00), and by Mohebalhojeh and Dritschel (2001, hereafter MD01), all violate local mass conservation and therefore exhibit velocity splitting. It was then shown how to construct a new class of PBMs, called hyperbalance equations or hyperbalance PBMs, that exactly satisfy local mass conservation and are therefore free of velocity splitting, and free of an internal inconsistency related to velocity splitting (Mohebalhojeh 2002). Hyperbalance equations are natural generalizations of the Bolin–Charney balance equations to an arbitrary formal order of accuracy. They are constructed by taking a plain PBM of given type and order, then incorporating it into a generalized Bolin–Charney procedure in a certain way. This process, which we call plain-to-hyper conversion, preserves the formal order of accuracy.

The formal order is not to be confused with asymptotic order in a small-parameter limit (e.g., Warn et al. 1995), nor with actual numerical accuracy. Here we consider numerical accuracies for chaotic vortical flows such as that in Fig. 1 whose Froude and Rossby numbers, numerically around 0.5 and 1, respectively, cannot be considered asymptotically small. The numerical accuracies nevertheless exhibit what might be called “quasi-asymptotic behavior.” This means that the error ϵ goes through a minimum as the formal order is increased, where ϵ measures the difference between PBM and PE (primitive equation) behavior. For a given flow over a given time interval there is therefore a “most accurate” PBM of given type.

Questions about the smallest attainable ϵ values are of interest, not only for improving our understanding of the fluid dynamics of atmospheres and oceans, but also for understanding the best design of numerical algorithms to cope simultaneously with balance and imbalance in model atmospheres and oceans (e.g., Mohebalhojeh and Dritschel 2000, 2004). Regardless of numerical errors, however, the ϵ values attainable are limited by the imbalance due to the spontaneous-adjustment emission of inertia–gravity waves by unsteady vortical flows, a physically real and ubiquitous phenomenon. It was this consideration that led MN00 to conjecture that, in view of the local mass adjustments involved in spontaneous-adjustment emission, the most accurate balanced models should be expected to violate local

mass conservation as well as energy and momentum conservation. In other words, MN00 argued that there should be a trade-off between accuracy and local mass conservation. The numerical experiments reported here were designed, therefore, to test MN00’s conjecture as part of a first attempt to examine the numerical accuracies of hyperbalance PBMs, in flow regimes like that of Fig. 1. This in turn required extreme care with issues of numerical error, especially truncation error, so as to deliver numerical accuracies able to detect slight departures from local mass conservation.

The paper is organized as follows. Section 2 summarizes the equations governing plain and hyperbalance PBMs for three different truncations of arbitrary order. Section 3 describes the numerical methods. Section 4 demonstrates that those methods are delicate enough to discriminate between local mass conservation and its violation. Section 5 specifies the initialization of the PEs used as the standard of accuracy, and makes a preliminary assessment of the accuracy of the plain- $\delta\delta$ and hyper- $\delta\delta$ PBMs defined in Part I and section 2, suggesting that plain-to-hyper conversion can sometimes degrade accuracy but sometimes improve it. Section 6 confirms that suggestion with a detailed intercomparison across all the plain and hyperbalance PBMs of the first to fifth orders. Section 7 presents a complementary assessment of accuracy using initialization experiments that estimate levels of imbalance. Section 8 presents some brief concluding remarks.

2. Plain and hyperbalance PBMs: Equations

The PEs to be used as the standard of accuracy are the f -plane shallow-water equations, (2.1a)–(2.1c) of Part I, in a doubly periodic domain with no bottom topography. Apart from implicit numerical diffusion all the models share the same PV conservation equation

$$\frac{\partial Q}{\partial t} = -\mathbf{v} \cdot \nabla Q, \quad (2.1)$$

where \mathbf{v} is the velocity and where Q is the exact PV given by the formula of Rossby (1936),

$$Q = \frac{f + \zeta}{H + g^{-1}\Phi}. \quad (2.2)$$

Here f is the Coriolis parameter, ζ the relative vorticity, H the area-mean layer depth, g the acceleration due to gravity, and Φ the geopotential anomaly.

For the plain PBMs, the PV inversion operators are defined by Eqs. (4.1) of Part I, which we rewrite as:

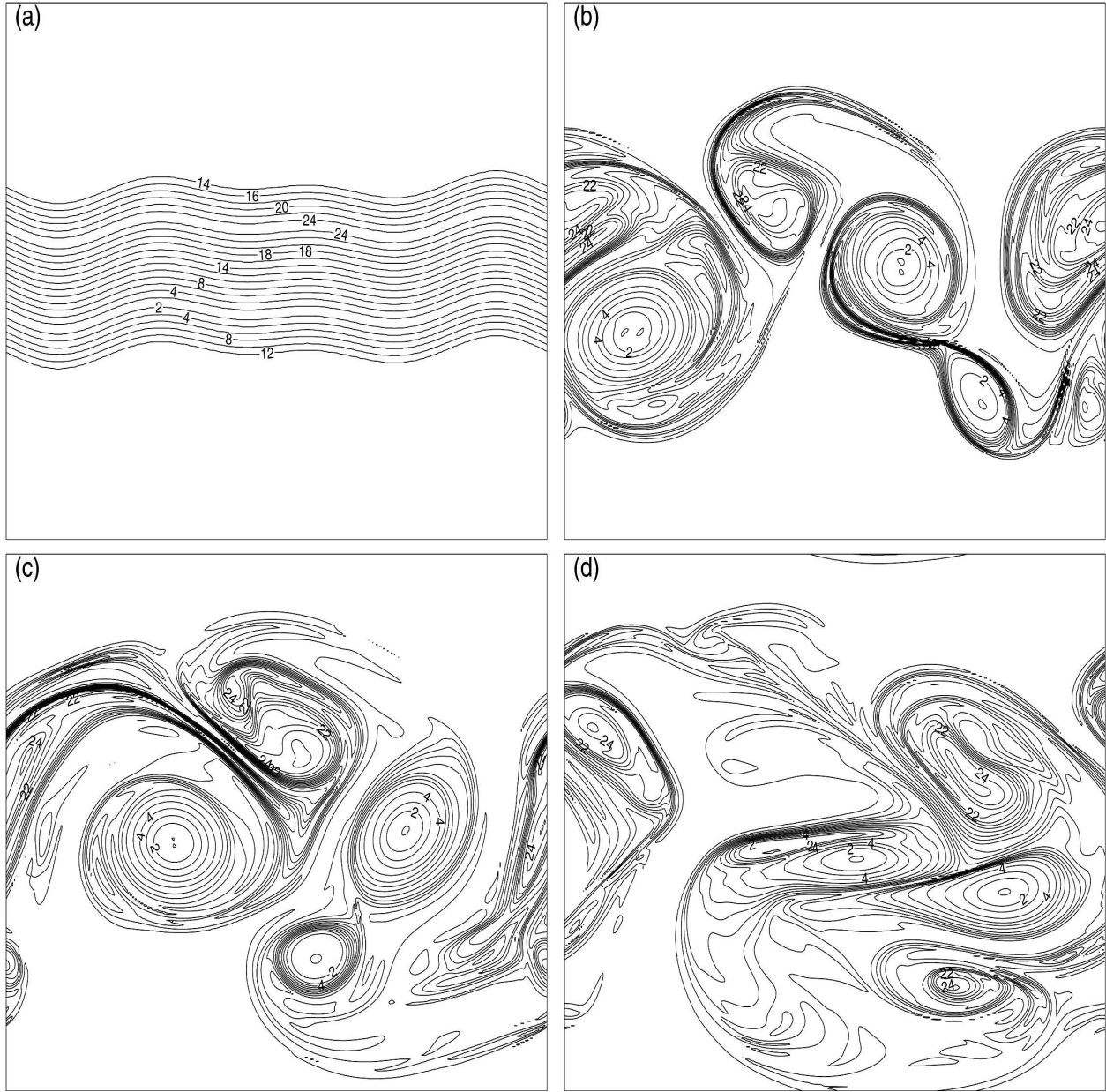


FIG. 1. Maps of PV at (a) $t = 0$, (b) $t = 5$, (c) $t = 10$, and (d) $t = 15$ days from one of the most accurate PBMs, the third-order plain- $\delta\delta$ PBM with semi-Lagrangian time stepping on a 256×256 grid. These maps are practically indistinguishable from the corresponding maps from the PEs. The Coriolis parameter $f = 4\pi \text{ day}^{-1}$. A single $2\pi \times 2\pi$ cell of the doubly periodic domain is shown; the Rossby length or radius $L_D = cf = 0.50$ in the same length units, so that the domain-crossing time is 1 day at the gravity wave speed $c = (gH)^{1/2}$ based on the area-mean layer depth H . In (a) the PV values run from $(f/4\pi H)$ times 1.11 to 25.04, the contour interval being 2 in (a), and 1 in (b), (c), and (d); selected contours are labeled.

$$\delta_{AI}^{(n)} = f\zeta_{AI}^{(n-1)} - \nabla^2\Phi_{AI}^{(n-1)} - \nabla \cdot (\mathbf{v} \cdot \nabla \mathbf{v})_{AI}^{(n-1)} \quad (n = 1, \dots, M + 1) \quad (2.3a)$$

$$\Phi_{AI}^{(n)} = -gH\delta_{AI}^{(n-1)} - \nabla \cdot (\mathbf{v}\Phi)_{AI}^{(n-1)} \quad (n = 1, \dots, M) \quad (2.3c)$$

$$\zeta_{AI}^{(n)} = -f\delta_{AI}^{(n-1)} - \nabla \cdot (\mathbf{v}\zeta)_{AI}^{(n-1)} \quad (n = 1, \dots, M) \quad (2.3b)$$

$$\mathbf{v}_{AI}^{(n)} = \mathbf{v}_{AI\psi}^{(n-1)} + \mathbf{v}_{AIx}^{(n)} = \mathbf{z} \times \nabla \nabla^{-2}\zeta_{AI}^{(n)} + \nabla \nabla^{-2}\delta_{AI}^{(n)} \quad (n = 0, \dots, M), \quad (2.3d)$$

where \mathbf{z} is the unit vertical vector and M a nonnegative integer, together with

$$f + \zeta_{AI} = (Q + q_{AI})(H + g^{-1}\Phi_{AI}) \quad (2.4)$$

and either the $\delta\delta$ truncation of order $M + 1$ with $M \geq 0$,

$$\begin{cases} \delta_{AI}^{(M)} = 0 \\ \delta_{AI}^{(M+1)} = 0, \end{cases} \quad (2.5a) \quad (2.5b)$$

or the $\gamma\gamma$ truncation of order $M - 1$ with $M \geq 1$,

$$\begin{cases} \delta_{AI}^{(M)} = -\nabla \cdot (\mathbf{v} \cdot \nabla \mathbf{v})_{AI}^{(M-1)} \\ \delta_{AI}^{(M+1)} = -\nabla \cdot (\mathbf{v} \cdot \nabla \mathbf{v})_{AI}^{(M)}, \end{cases} \quad (2.6a) \quad (2.6b)$$

or the $\delta\gamma$ truncation of order M with $M \geq 0$,

$$\begin{cases} \delta_{AI}^{(M)} = 0 \\ \delta_{AI}^{(M+1)} = -\nabla \cdot (\mathbf{v} \cdot \nabla \mathbf{v})_{AI}^{(M)}. \end{cases} \quad (2.7a) \quad (2.7b)$$

Superscripts (n) denote diagnostic estimates of n th time derivatives; for further explanation see (4.3), (5.1), (5.2), and (5.3)ff. Part I. The nominal orders $M + 1$, $M - 1$, and M have no absolute significance and are defined purely for consistency with the conventions established in MN00 and MD01. When $M = 0$, Eqs. (2.3b) and (2.3c) are empty sets of equations and are therefore absent from (2.3). The suffix AI has the same meaning as in Part I, namely that the variables are regarded as auxiliary variables within the PV inversion process. The velocity field \mathbf{v} resulting from the plain PV inversion, that is, from solving the foregoing equations when the PV field $Q(\mathbf{x})$ is given, is defined by

$$\mathbf{v} = \mathbf{v}_{AI}\{\mathbf{x}; Q(\cdot)\}, \quad (2.8)$$

where $\mathbf{v}_{AI} = \mathbf{v}_{AI}^{(0)} = \mathbf{v}_{AI\psi}^{(0)} + \mathbf{v}_{AI\chi}^{(0)}$. As explained below Eq. (5.6) of Part I, the small quantity q_{AI} in (2.4), to be called the ‘‘PV offset,’’ is independent of the spatial coordinate \mathbf{x} and is determined as part of the plain inversion calculation when the total mass and the far-field or boundary Kelvin circulation are specified. The corresponding plain PBM consists of the PV inversion operator just defined together with the prognostic equation (2.1). For the most accurate plain PBMs, q_{AI} is often small enough to be lost in the numerical noise. It is, however, conceptually important because it measures the internal inconsistency pointed out in Mohebalhojeh (2002). When total mass and far-field or boundary Kelvin circulation are held constant, plain PBM evolution makes q_{AI} vary in time even though constant in space. Because we insist that $DQ/Dt = 0$ when the velocity field \mathbf{v} is given by (2.8), we generally have $D(Q + q_{AI})/Dt \neq 0$. In a plain PBM, there is no single velocity field that both advects and evaluates the PV exactly.

As shown in section 6 of Part I, each plain PBM may be converted to a corresponding hyperbalance PBM as follows. We replace (2.8) by

$$\mathbf{v} = \mathbf{v}\{\mathbf{x}; Q(\cdot)\}, \quad (2.9)$$

where $\mathbf{v} = \mathbf{v}_{AI\psi}^{(0)} + \nabla\nabla^{-2}\delta$ and δ is defined by the implicit elliptic equation

$$\mathcal{L}\delta = f\nabla \cdot (\mathbf{v}\zeta) - \nabla^2\nabla \cdot (\mathbf{v}\Phi) + \frac{\mathcal{D}\mathcal{R}}{\mathcal{D}Q} \odot (\mathbf{v} \cdot \nabla Q), \quad (2.10)$$

with $\zeta = \zeta_{AI} = \zeta_{AI}^{(0)}$ and $\Phi = \Phi_{AI} = \Phi_{AI}^{(0)}$ in (2.2) and in (2.10) and with

$$\begin{aligned} \mathcal{R} &= \mathcal{R}\{\mathbf{x}; Q(\cdot)\} \\ &= -\delta_{AI}^{(1)} - \nabla \cdot [(\mathbf{v}_{AI\psi} + \mathbf{v}_{AI\chi}) \cdot \nabla(\mathbf{v}_{AI\psi} + \mathbf{v}_{AI\chi})]. \end{aligned} \quad (2.11)$$

Here, as in Part I, \mathcal{L} is the modified Helmholtz operator $gH\nabla^2 - f^2$ and \mathcal{D} signifies variational differentiation, with \odot the corresponding inner product involving integration over the physical domain; see Eq. (3.8) of Part I. Thus the hyperbalance PBM is defined by (2.9)–(2.11) together with (2.1)–(2.3) and a truncation, such as (2.5), (2.6), or (2.7), for some nonnegative integer M . As verified in Part I the velocity field \mathbf{v} defined by (2.9) now conserves mass locally, advects Q via (2.1), and evaluates Q via (2.2).

For computational purposes it will be found convenient to specify the total mass and far-field or boundary Kelvin circulation, ab initio, to match those of the PE run used as the standard of accuracy. We must then replace (2.2) by (2.4), remembering that $\zeta = \zeta_{AI}$ and $\Phi = \Phi_{AI}$. This no longer represents any internal inconsistency because, provided that the advecting velocity is still given by (2.9), we now have a q_{AI} that is exactly constant in time as well as space. Thus the replacement of (2.2) by (2.4) does no more than adjust Q by an additive constant so as to make it consistent with the specified mass and circulation. Within the hyperbalance equations this is no more than a change of notation, merely replacing the symbol Q by the symbol $(Q + q_{AI})$. We have $DQ/Dt = 0$ and $D(Q + q_{AI})/Dt = 0$ simultaneously, and there is no internal inconsistency. So any failure of q_{AI} to be constant in a numerical run will be entirely due to rounding or truncation error.

3. Numerical setup and methodology

The geometry of the experiments is a $[-\pi, \pi] \times [-\pi, \pi]$ doubly periodic box, allowing pseudospectral computation via fast Fourier transforms (FFTs) both

for the PBMs and for the PEs used as the standard of accuracy; details for the PEs may be found in Mohebalhojeh and Dritschel (2000). The Rossby length or radius $L_D = cf$ is specified in the same length units, where the Coriolis parameter f is set to 4π and the gravity wave speed $c = (gH)^{1/2}$ is based on the area-mean layer depth H , which will be held constant to keep total mass constant. The Kelvin circulation around the boundary will likewise be held constant at the value $8\pi f$, corresponding to zero relative circulation or zero domain area-average relative vorticity. For most of the experiments $L_D = 0.50$, though in order to expose certain sensitivities we also use $L_D = 0.51, 0.52, \dots, 0.59, 0.60$. In all cases the domain maximum Froude and Rossby numbers are somewhere near $Fr = 1/2$ and $Ro \sim 1$; for example, when $L_D = 0.50$ we have $Fr \approx 0.52$ and $Ro \approx 1.0$. Because the angular velocity of the background rotation is 2π , the unit of time is 1 day, which is also the domain-crossing time for a gravity wave when $L_D = 0.50$.

The test flows (e.g., Fig. 1) are similar to those extensively reported in Dritschel et al. (1999), Mohebalhojeh and Dritschel (2000), and Dritschel and Mohebalhojeh (2000). For the initial conditions the PV is first prescribed in a contour representation, then converted to a grid representation that is used to initiate all PBM runs. In this sense all the PBM runs have identical initial conditions. In its contour representation the initial PV field has a zigzag profile in the y direction, with piecewise constant gradients, undulated by a small-amplitude disturbance. The disturbance uses a superposition of Fourier harmonics in the x direction, with wavenumbers 2 and 3. The maximum and minimum PV values in the zigzag profile, $Q = 25.04$ and 1.11 , are on contours (not shown in Fig. 1a) that undulate around $y = \pm 0.5$, respectively.¹ The L_2 norm of Q at initial time $t = 0$ is $\|Q\|_2 = 13.7$ and stays nearly constant during all of the most accurate integrations, typically to within a few percent. The jet undergoes a rapid rollup evolution by which at $t = 5$ well-defined vortices are formed (Fig. 1b). The subsequent evolution is dominated by vortex interactions.

The plain inversions for the PBMs, including the third-order plain- $\delta\delta$ PBM that produced Fig. 1, are all performed pseudospectrally using FFTs to move between regular grids in spectral space and physical space as in MD01. Linear operations are done on the spectral

grid, for instance taking derivatives, inverting Laplacians ∇^2 as in (2.3d), and inverting the modified Helmholtz operators \mathcal{L} that arise from substitution of (2.3b) and (2.3c) into (2.3a). The multiplications to evaluate nonlinear terms are done on the physical grid without dealiasing. The linear and nonlinear operations are performed alternately in an iterative loop, taking advantage of the speed of the FFT and the smoothness of the inverse operators ∇^{-2} and \mathcal{L}^{-1} . Iterations are terminated when the i th and $(i + 1)$ th iterates of both Φ_{AI} and δ_{AI} are close enough, more precisely when $\|\Phi_{AI}^{[i]} - \Phi_{AI}^{[i+1]}\|_\infty < \kappa_{\Phi 1}$ and $\|\delta_{AI}^{[i]} - \delta_{AI}^{[i+1]}\|_\infty < \kappa_{\delta 1}$ where $\|\cdot\|_\infty$ denotes the maximum or L_∞ norm and where $\kappa_{\Phi 1}$ and $\kappa_{\delta 1}$ are preassigned small numbers. We take $\kappa_{\Phi 1}$ in the range $[10^{-12}, 10^{-10}] gH$ and $\kappa_{\delta 1}$ in the range $[10^{-10}, 10^{-8}]$. All are exceedingly small fractions of typical values. For instance, in the example of Fig. 1 we have $(gH)^{-1}\|\Phi_{AI}\|_\infty = 0.38, 0.43, 0.46, 0.53$ at $t = 0, 5, 10, 15$, respectively, and $\|\delta_{AI}\|_\infty = 0.22, 0.29, 0.19, 0.39$. Stringent convergence criteria are, however, essential if our accuracy tests are to be meaningful.

To advect Q across the physical grid we have used two different time stepping algorithms. These are the contour-advective semi-Lagrangian (CASL) algorithm of Dritschel and Ambaum (1997) and a standard semi-Lagrangian (SL) algorithm. The SL algorithm used here involves a backward trajectory computation using the midpoint method (Temperton and Staniforth 1987) for all the grid points, and a bicubic Lagrange interpolation to update Q values on the physical grid.

To integrate the hyperbalance equations we must deal with $\mathcal{R}\{\mathbf{x}; Q(\cdot)\}$, the nonlocal function appearing in (2.10) and (2.11). Except in the lowest-order cases there is no known analytic, closed form expression for \mathcal{R} representing the nonlocal dependence of \mathcal{R} upon the Q field. Still less is there any such expression for the variational derivative $\mathcal{D}\mathcal{R}/\mathcal{D}Q$ appearing in the diagnostic equation (2.10). However, exploiting the fact that the time evolution defined by the hyperbalance equations entails that the actual rate of change of \mathcal{R} at fixed \mathbf{x} , under that evolution, is equal to $-(\mathcal{D}\mathcal{R}/\mathcal{D}Q) \odot (\mathbf{v} \cdot \nabla Q)$, we have developed an iterative time stepping technique to avoid having to deal directly with the variational derivative. A brief account of the technique and the solution procedure for a hyperbalance PBM is given next. Since on the numerical level that we are forced to deal simultaneously with PV inversion and time evolution, we need to make time dependences explicit and will therefore use notations such as $Q = Q(\mathbf{x}, t)$ from here onward.

The procedure involves an outer loop and two inner loops. Given a Q field, inner loop (1) solves the plain

¹ Figure 1a actually represents the gridded PV. The corresponding contour representation can be found at top left of Fig. 2a of Dritschel et al. (1999). The zigzag profile of unperturbed PV at $t = 0$ can be found in Fig. 1a of Dritschel et al. (1999) and in Fig. 1a of Mohebalhojeh and Dritschel (2000).

PV inversion equations, that is, (2.3)–(2.4) together with (2.5) or another truncation, using the iterative pseudospectral method already described. To keep the boundary circulation at its specified value $8\pi f$ we need to allow $q_{AI} \neq 0$, in case Q is imperfectly specified and in any case to allow the computational system to tolerate rounding and truncation errors, wherever they may occur. Inner loop (2) similarly solves (2.10) for δ when all other fields and the last term are given. As with inner loop (1), the process has to be iterative because the unknowns appear implicitly. In the case of (2.10), δ appears explicitly on the left but also implicitly in the first two terms on the right, through (2.9).

Before starting the outer loop at the current time t , with given $Q(\mathbf{x}, t)$, a preliminary call to inner loop (1) is made. It produces the fields

$$\zeta_{AI}(\mathbf{x}, t), \quad \mathbf{v}_{AI\psi}(\mathbf{x}, t), \quad \delta_{AI}(\mathbf{x}, t), \quad \mathbf{v}_{AI\chi}(\mathbf{x}, t), \quad \delta_{AI}^{(1)}(\mathbf{x}, t),$$

and $\Phi_{AI}(\mathbf{x}, t)$ together with a value for q_{AI} . These are all held fixed throughout the subsequent running of the outer loop, whose sole purpose is to evaluate the $(\mathcal{D}\mathcal{R}/\mathcal{D}Q) \odot (\mathbf{v} \cdot \nabla Q)$ and δ fields. We denote successive guesses for δ by $\delta^{[k]}$ and the corresponding contributions to (2.9) by $\mathbf{v}_\chi^{[k]}$, going into the k th cycle of the outer loop. To start the loop we take as first guess $\delta^{[1]} = \delta_{AI}(\mathbf{x}, t)$, already available from the preliminary call to inner loop (1). The k th cycle of the outer loop consists of the following stages:

- (i) Advect Q forward in time using the current guess for the velocity field $\mathbf{v} = \mathbf{v}_{AI\psi} + \mathbf{v}_\chi^{[k]}$. A short time step is used, a fraction α of the time step Δt used for time stepping the PBM, the value of α being chosen from numerical experimentation on the convergence of the outer loop. Denote the result by $Q^{[k]}(\mathbf{x}, t + \alpha\Delta t)$.
- (ii) Call inner loop (1) with $Q = Q^{[k]}(\mathbf{x}, t + \alpha\Delta t)$. Denote the resulting fields by $\zeta_{AI}^{[k]}(\mathbf{x}, t + \alpha\Delta t)$, $\mathbf{v}_{AI\psi}^{[k]}(\mathbf{x}, t + \alpha\Delta t)$, $\delta_{AI}^{[k]}(\mathbf{x}, t + \alpha\Delta t)$, $\mathbf{v}_{AI\chi}^{[k]}(\mathbf{x}, t + \alpha\Delta t)$, $\delta_{AI}^{(1)[k]}(\mathbf{x}, t + \alpha\Delta t)$, and $\Phi_{AI}^{[k]}(\mathbf{x}, t + \alpha\Delta t)$. The new value of q_{AI} is discarded.
- (iii) Use the foregoing to evaluate $\{(\mathcal{D}\mathcal{R}/\mathcal{D}Q) \odot (\mathbf{v} \cdot \nabla Q)\}^{[k]}$ as $[-\mathcal{R}\{\mathbf{x}; Q^{[k]}(\cdot, t + \alpha\Delta t)\} + \mathcal{R}\{\mathbf{x}; Q(\cdot, t)\}]/(\alpha\Delta t)$ from (2.11).
- (iv) Call inner loop (2) to obtain $\delta^{[k+1]}$ and $\mathbf{v}_\chi^{[k+1]}$ by solving (2.10). The termination criterion at the j th iteration is $\|\delta^{[k+1,j]} - \delta^{[k+1,j+1]}\|_\infty < \kappa_{\delta 2}$ for some preassigned small number $\kappa_{\delta 2}$.
- (v) Check for numerical convergence of the outer loop. If its termination criterion is not satisfied then go back to stage (i) above. The termination criterion is taken to be $\|\delta^{[k]} - \delta^{[k+1]}\|_\infty < \kappa_{\delta 3}$ for some preassigned small number $\kappa_{\delta 3}$.

Values of $\kappa_{\delta 2}$ and $\kappa_{\delta 3}$ will both be taken in the range $[10^{-8}, 10^{-7}]$. To keep computational costs within bounds, values toward the higher ends of the ranges assigned to $\kappa_{\Phi 1}$, $\kappa_{\delta 1}$, $\kappa_{\delta 2}$, and $\kappa_{\delta 3}$ have been used for the fourth- and fifth-order PBMs, and toward the lower ends for the first-, second-, and third-order PBMs. The main time step $\Delta t = 0.002$ in all cases, and $\alpha = 0.25$ except in some of the fourth- and fifth-order cases where $\alpha = 0.125$.

4. Numerics and local mass conservation

It was proved in Part I, section 4, that for $M \geq 2$ every plain PBM violates local mass conservation. However, it demands extreme care to see this violation numerically. In this section we establish that with the foregoing values of $\kappa_{\Phi 1}$, $\kappa_{\delta 1}$, $\kappa_{\delta 2}$, and $\kappa_{\delta 3}$ and with sufficiently fine grids our numerical procedures are, indeed, accurate enough for the purpose. (Readers willing to take that assertion on faith can skip this section.)

During the time integration of each PBM, departures from local mass conservation are measured by computing a mass residual:

$$m_{\text{res}}^n = \frac{h^{n+1} - h^{n-1}}{2\Delta t} + H\delta^n + \nabla \cdot (\mathbf{v}h)^n, \quad (4.1)$$

where h stands for $g^{-1}\Phi$ and the unbracketed superscript n denotes a value at the n th time step. As with all the other nonlinear terms in our computations, the term $\nabla \cdot (\mathbf{v}h)$ in (4.1) is evaluated pseudospectrally without dealiasing. Thus when $m_{\text{res}} = 0$ we have a discrete analog of the local mass-conservation equation, (2.1c) of Part I. The time step $\Delta t = 0.002$ is dictated by accuracy considerations in the PE time integrations, which use semi-implicit time stepping but in which we want to represent inertia–gravity waves accurately. Such Δt values are significantly smaller than the values demanded by PV advection in the PBMs. We use the same small value $\Delta t = 0.002$ in the PBMs 1) because the mass residual is sensitive to Δt , and 2) because we want comparisons between PBMs and the parent PEs to be done in as consistent a way as possible.

We have examined the behavior of m_{res} under both CASL and SL advection. It turns out that the CASL algorithm is unsuitable for assessing the accuracy of local mass conservation. This is because its ability to cope with sharp-edged PV distributions has a price in terms of noisiness in the way it handles mass rearrangement. The noise has properties that are far from obvious; a brief discussion is given in the appendix.

In the main part of the paper, therefore, we use only

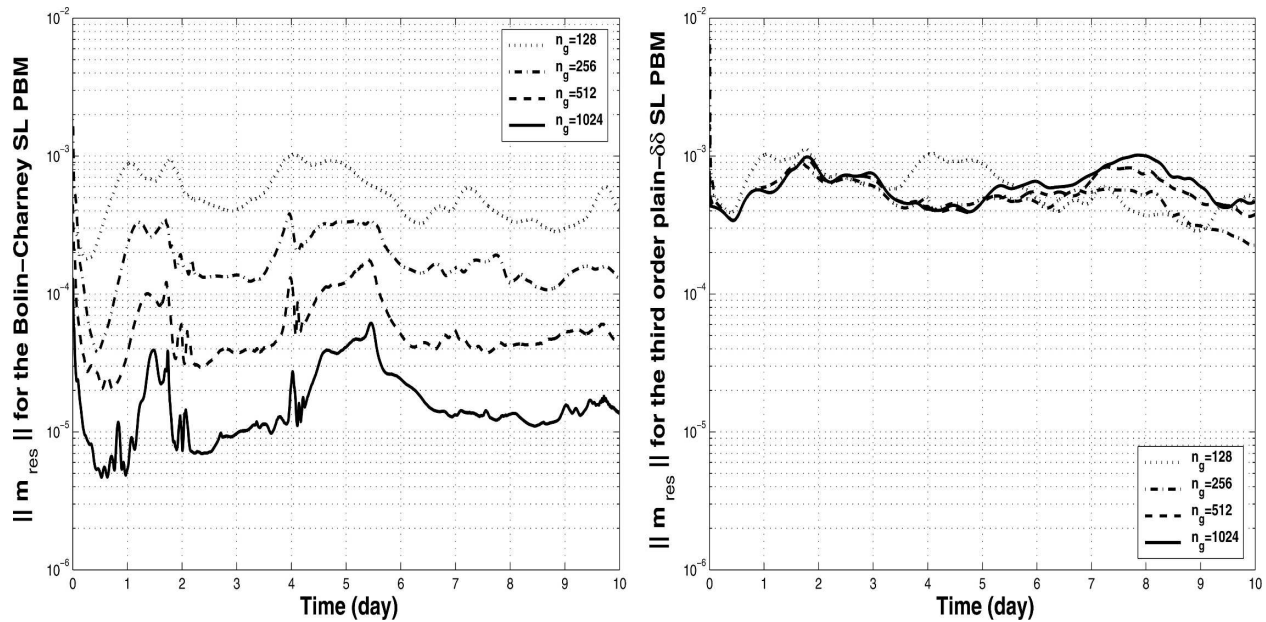


FIG. 2. The L_2 norm of mass residual vs time for resolutions $n_g = 128, 256, 512,$ and 1024 . The results are for (left) the Bolin–Charney SL PBM, equivalently the first-order hyper- $\delta\delta$ PBM, and (right) the third-order plain- $\delta\delta$ SL PBM.

the SL algorithm. Being slightly diffusive, it keeps the PV distributions slightly smoother and gives the numerics a better chance to converge toward the continuum limit. In the hyperbalance equations, the diffusion now implicit in the prognostic equation slightly changes the term $(\mathcal{D}\mathcal{R}/\mathcal{D}Q) \odot (\mathbf{v} \cdot \nabla Q)$. But that change is automatically taken into account at stage (iii) of the numerical procedure since the SL algorithm is, of course, used at stage (i) as well as in the main time stepping.

The left-hand panel of Fig. 2 presents the L_2 norm $\|m_{\text{res}}\|_2$ of the mass residual for the Bolin–Charney SL PBM, equivalently the first-order hyper- $\delta\delta$ PBM (Part I, footnote 4), at resolutions corresponding to $n_g = 128, 256, 512, 1024$ grid points across the domain. The first 10 days of integration are shown and the initial condition is the same as in Fig. 1. By comparison with the results for the corresponding CASL PBM (appendix, Fig. A1), the mass residual is markedly smoother and smaller, and it decreases as resolution n_g increases. A systematic decrease of $\|m_{\text{res}}\|_2$ by a factor of 2 to 3 at each doubling of resolution is observed. The results for the L_2 norm of mass tendency (not shown) exhibit convergence in the sense of negligible sensitivity to resolution. We approximate the mass tendency $\partial h/\partial t$ as the first term on the rhs of (4.1), written $\Delta h/\Delta t$ for brevity, throughout. The smallness of $\|m_{\text{res}}\|_2$ can be judged by comparing it with $\|\Delta h/\Delta t\|_2$. Over the 10-day runs the ratio $\|m_{\text{res}}\|_2/\|\Delta h/\Delta t\|_2$ has typical values of the order of 0.01, 0.003, 0.001, 0.0004, respectively, for $n_g = 128, 256, 512, 1024$.

Also shown in Fig. 2, on the right, are the corresponding results for the third-order plain- $\delta\delta$ SL PBM. This PBM has special interest since it will prove to be one of the most accurate for the test flows. The plain PBM exhibits an irregular behavior with no systematic decrease. This is just what we expect from what was proved in section 4 of Part I for high-order plain PBMs, for which $M \geq 2$. That is, in the continuum limit, the Bolin–Charney model conserves mass locally but high-order plain PBMs do not; that is, such plain PBMs have nonvanishing mass residuals.

In Fig. 3, $\|m_{\text{res}}\|_2$ is shown against time for seven PBMs all at resolution $n_g = 256$. The PBMs are the Bolin–Charney model (heavy solid curves duplicating the dash-dot curve on the left of Fig. 2—note the different scales at left and right in Fig. 3) together with the first- to fifth-order plain- $\delta\delta$ PBMs and the third-order hyper- $\delta\delta$ PBM (thin solid curve on the left in Fig. 3). Since the Bolin–Charney model and the third-order hyper- $\delta\delta$ PBM both conserve mass locally, that is, have vanishing mass residual in the continuum limit, the heavy solid and thin solid curves measure the effects of rounding and truncation errors, here dominated by truncation error. The violation of local mass conservation by the first-, second-, and third-order plain- $\delta\delta$ PBMs shows up clearly, being most conspicuous in the first-order case, the dashed curve high on the left. Except for a brief period between days 4 and 6, the fourth- and fifth-order plain- $\delta\delta$ $\|m_{\text{res}}\|_2$ values shown on the right are close to the Bolin–Charney values. At this

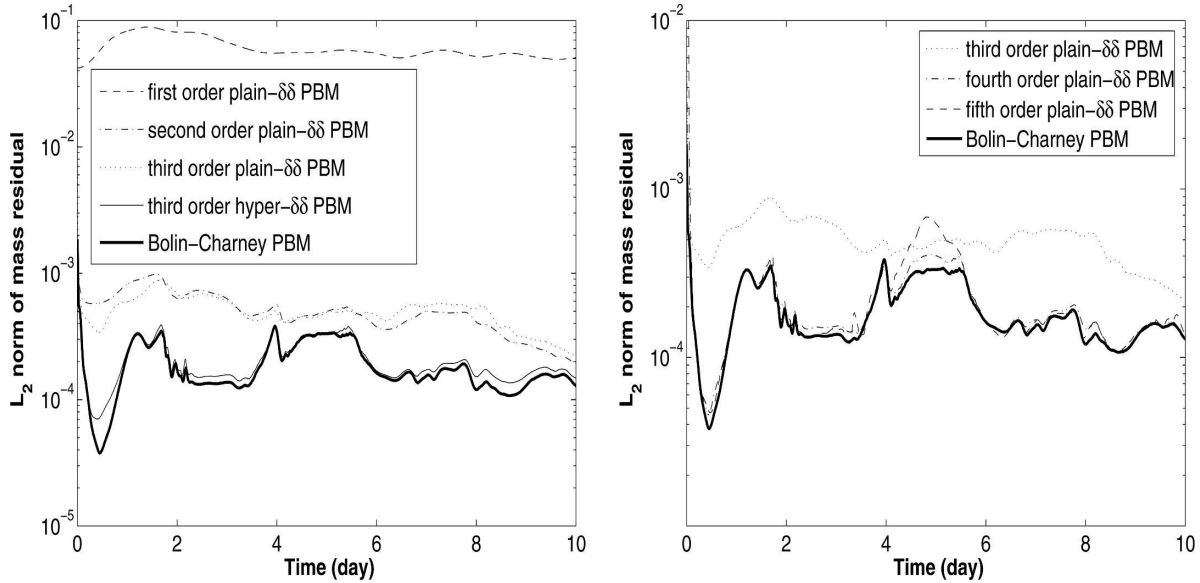


FIG. 3. (left) The L_2 norms of mass residual vs time for Bolin–Charney (thick solid), the first-order plain- $\delta\delta$ (dashed), the second-order plain- $\delta\delta$ (dash-dotted), the third-order plain- $\delta\delta$ (dotted), and the third-order hyper- $\delta\delta$ (thin solid) PBMs. (right) The fourth-order plain- $\delta\delta$ (dash-dotted) and the fifth-order plain- $\delta\delta$ (dashed) PBMs. For ease of comparison, in the right panel the results for the third-order plain- $\delta\delta$ and Bolin–Charney PBMs have been replotted. Recall that the Bolin–Charney PBM is the same as the first-order hyper- $\delta\delta$ PBM (Part I, footnote 4).

resolution, for the fourth and fifth orders the continuum-limit values of $\|m_{res}\|_2$ are evidently being masked by numerical truncation error. By contrast, the dotted and thin-solid curves on the left clearly show that the resolution $n_g = 256$ is good enough to see the effect of plain-to-hyper conversion for the third-order $\delta\delta$ PBM truncation, as well as the more dramatic effect at first order (for which the conversion produces the Bolin–Charney model, that is, takes the high dashed into the low heavy curve). As expected—and this is another powerful check on the correctness of the numerical codes—the conversion succeeds in making $\|m_{res}\|_2$ decrease to values close to that for the Bolin–Charney model, that is, to values governed solely by rounding and truncation error.

5. PE initialization and preliminary comparisons

Before embarking on a systematic assessment of PBM accuracy we need to consider how to initialize the PE runs that serve as the standard of accuracy. We have experimented extensively with initialization methods. Details are omitted for brevity. Not surprisingly, the experiments confirmed that there is great sensitivity to initialization. The main result was, however, that no systematic bias or disadvantage arose from using consistent initialization in the sense that the PBM whose accuracy is being judged is also used to initialize the

PEs. Some further questions regarding initialization and inertia–gravity wave activity in the PEs will be investigated in section 7. In addition to using consistent initialization throughout, we now extend the time span of each run to 30 days and reduce the sensitivity to initialization by starting from the same initial jet but with an undulation 10 times smaller than that in Fig. 1. Resolution is $n_g = 256$ throughout.

We begin by asking whether plain-to-hyper conversion has any obvious systematic effect, such as the degradation of accuracy expected from MN00’s argument. Of course it might also be argued that better conservation should improve accuracy. Since it was known from the start that the third-order plain- $\delta\delta$ PBM often performs exceedingly well in parameter regimes like that of Fig. 1, with order unity Froude and Rossby numbers, we begin by looking at the cumulative effect on accuracy, over 30 days, of plain-to-hyper conversion in the third-order $\delta\delta$ case. At the same time we check for sensitivity to parameter conditions.

We measure accuracy using the dimensionless quantity

$$\epsilon = \epsilon(t) = \frac{\|Q_{PBM} - Q_{PE}\|_2}{\|Q_{PE}\|_2}, \quad (5.1)$$

where Q_{PBM} and Q_{PE} denote the PV fields of the PBM and the PEs, respectively, and $\|\cdot\|_2$ is the L_2 norm.

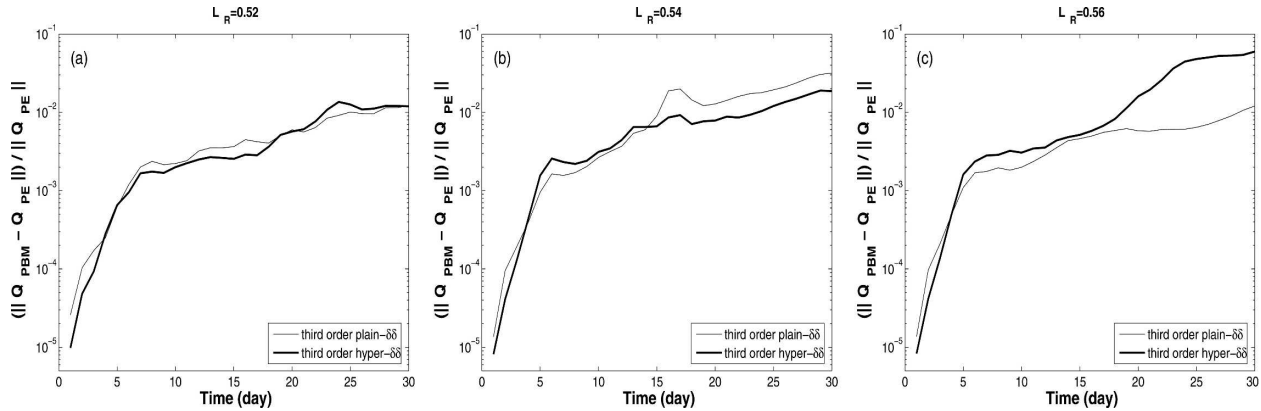


FIG. 4. The relative errors $\epsilon(t)$ in PV for the third-order plain- $\delta\delta$ (thin solid) and hyper- $\delta\delta$ (thick solid) PBMs. Consistent initialization and resolution $n_g = 256$ are used throughout. The results are for (a) $L_D = 0.52$, (b) $L_D = 0.54$, and (c) $L_D = 0.56$. Corresponding results for $L_D = 0.55$ and 0.60 can be seen as pairs of dashed curves at top left and right, respectively, of Fig. 6 below.

This is an exceedingly sensitive measure of accuracy. It will be found that the best-performing PBMs yield values $\epsilon \sim 10^{-2}$ after 30 days. Such small values are possible only if the two PV fields match each other very closely, down to tiny details.

Figure 4 shows $\epsilon(t)$ for three pairs of runs out of a sequence in which the Rossby length L_D is varied slightly, taking the eleven values $L_D = 0.50, 0.51, 0.52, \dots, 0.60$. Although the Froude and Rossby numbers change somewhat across this range of L_D values, they are anything but small and it is reasonable to regard all of these runs as being in much the same parameter regime. Perhaps surprisingly, we see that plain-to-hyper conversion does not always degrade accuracy. Nor does it always improve it.

The case $L_D = 0.54$ (middle panel of Fig. 4) shows a definite improvement, when judged on cumulative accuracy, contrary to what might be expected from MN00's argument. Other cases, such as $L_D = 0.52$ (left panel), also $L_D = 0.55, 0.59$ (not shown in Fig. 4), indicate no clear advantage for either PBM, though for $L_D = 0.55$ (dashed curves at top left in Fig. 6 below) the hyperbalance PBM has the advantage at later times. Yet others, such as $L_D = 0.56$ (right panel), also $L_D = 0.50, 0.51, 0.53, 0.57, 0.58, 0.60$, show to varying degrees an advantage of the plain over the hyperbalance PBM. For instance in the case $L_D = 0.60$ (dashed curves at top right in Fig. 6 below) the plain PBM could be said to win, though only by a fine margin, for nearly the whole time span. In the case $L_D = 0.50$ (solid curves at top left in Fig. 5 below) the plain wins by a substantial margin. Especially when looking at the right panel of Fig. 4 one might be tempted to say that, on average, the plain PBM has a slight advantage over the hyperbalance PBM. But no such conclusion is supported when the

other orders and truncations are looked at systematically, as is done next.

6. A systematic intercomparison of PBMs

Figure 5 presents the $\epsilon(t)$ behavior for a complete set of 30 runs each over 30 days, at $L_D = 0.50$, covering all three truncations from first order to fifth order. All are at resolution $n_g = 256$. We continue to focus mainly on cumulative errors over 30 days, that is, on the best-performing PBMs as measured by $\epsilon(30)$. The quasi-asymptotic behavior is evident, with third- or fourth-order performing best. Very similar quasi-asymptotic behavior will be shown by the different measures of accuracy considered in section 7. We conclude that there is no point in going to higher order than fifth.

Among the cases represented in Fig. 5 we see that there are just two best performers, one plain and the other hyperbalance. They are the third-order plain- $\delta\delta$ PBM (thin solid curve at top left) and the fourth-order hyper- $\gamma\gamma$ PBM (heavy dashed curve at middle right), for both of which $\epsilon(30)$ is very close to 10^{-2} . It should be remembered, of course, that the nominal orders are defined purely for consistency with earlier published work and have no absolute significance. All that matters is the best performance, as defined by minimizing ϵ over all available PBMs.

If it were computationally feasible to consider fractional orders and a much larger set of different truncations, then one would probably find a minimum value of $\epsilon(30)$ that is somewhat smaller than 10^{-2} . However, the results of Fig. 5, along with similar results for other cases, are enough to suggest that in the present parameter regime such an $\epsilon(30)$ value would not be drastically smaller.

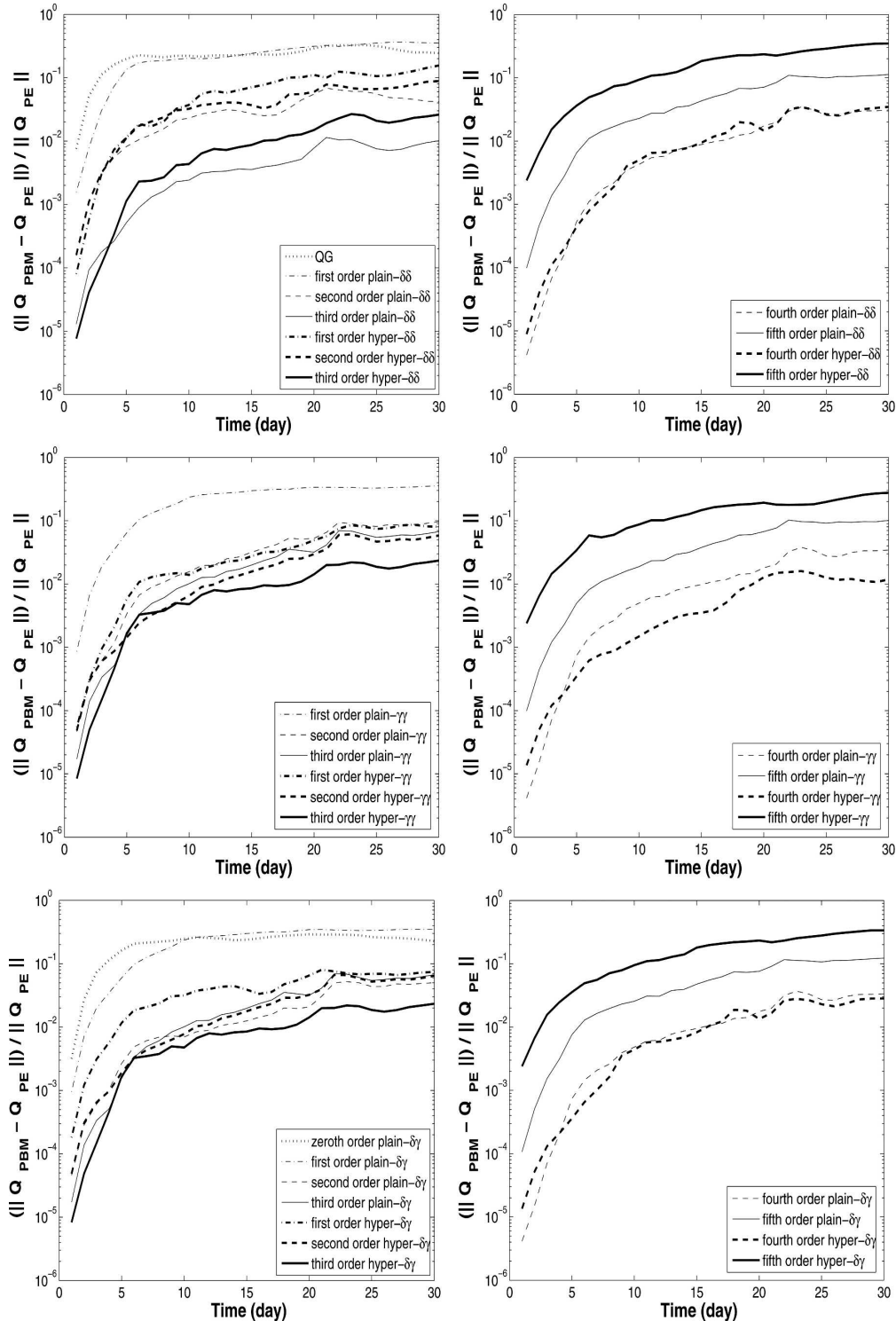


FIG. 5. The thin curves show $\epsilon(t)$ for the plain PBMs of (left) first to third order, and (right) fourth and fifth order. The thick curves are for the corresponding hyperbalance PBMs: the (top) $\delta\delta$, (middle) $\gamma\gamma$, and (bottom) $\delta\gamma$ truncations. Insets give formal orders of accuracy as defined in the text. $L_D = 0.50$ throughout. Also shown are the results for the QG model (dotted curve at upper left) and the zeroth-order plain- $\delta\gamma$ PBM (dotted curve at bottom left). At large time, in this case, the two best performances are those of the third-order plain- $\delta\delta$ PBM (thin solid curve at upper left) and the fourth-order hyper- $\gamma\gamma$ PBM (thick dashed curve at middle right), with the plain- $\delta\delta$ best by a very small margin. As before, the resolution is $n_g = 256$.

A reviewer has asked how the balanced models based on geostrophic balance perform. For this we have presented in Fig. 5 the $\epsilon(t)$ behavior for the standard quasigeostrophic (QG) model² (dotted curve at upper left) and the zeroth-order plain- $\delta\gamma$ PBM (dotted line at bottom left), which is the same as the balanced model introduced by Bleck (1974). It is clear that these models fall well short of the accuracy achieved by the best-performing PBMs.

Figure 6 shows two other cases, $L_D = 0.55$ on the left and $L_D = 0.60$ on the right. At large times, the best performance for $L_D = 0.55$ is that of the fourth-order plain- $\delta\delta$ PBM (thin solid curve at upper left) and for $L_D = 0.60$ the fourth-order hyper- $\gamma\gamma$ PBM (thick solid curve at middle right). Notice that these two PBMs differ from the two best performers in Fig. 5, though once again we have minimum $\epsilon(30)$ values close to 10^{-2} . It seems clear by now that, in order to get substantially smaller $\epsilon(30)$ values, one would have to go to parameter regimes with substantially smaller Froude or Rossby number.

Another feature of interest both in Fig. 5 and in Fig. 6 is that judgments made from $\epsilon(30)$ seem to be robust over a large part of the time span, in many cases the entire time span following the first 5 days or so when the jet instability grows and saturates, though in the $L_D = 0.55$ case it is only the latter part following the first 15 days (top left in Fig. 6).

In summary, three main points have emerged. First, plain-to-hyper conversion can either improve or degrade the performance of a particular PBM, depending on the flow considered. There are some systematic effects when $\epsilon(30)$ is above its minimum but, more importantly, no systematic effects near the minimum. Second, no particular PBM is the overall winner in all cases. The PBM or PBMs that minimize $\epsilon(30)$ are not the same PBMs in all cases, even when we confine attention to similar flows in similar parameter regimes. Third, near minimum $\epsilon(30)$ there is no systematic preference for plain PBMs over hyperbalance PBMs, as far as accuracy is concerned. So accuracy is not a countervailing consideration against the theoretical advantages of the hyperbalance equations.

7. Initialization experiments

Hitherto we have paid little attention to the earlier stages, the first few days when the jet instability grows and saturates and goes over into the chaotic vortex re-

gime. The transition is of course conspicuous in all the $\epsilon(t)$ plots, as the break in error growth rates shortly after day 5. Also noticeable, in a majority of the best-performing cases, is a crossover suggesting a negative correlation between early and late error. Quite often, the PBMs with the smallest late ϵ values have early ϵ values slightly larger than those of their nearest competitors. Examples can be seen in Figs. 5 and 6 at middle right and in Fig. 5 at top left, though Fig. 6 at top left is exceptional up to day 4.

Let us look more closely, then, at what is going on in the early stages near the beginning of the chaotic vortex regime. Some of the early error, though of course not all of it, must be due to slight imbalance in the PEs. To estimate such imbalance we have carried out what might be called initialization or slow-quasimanifold-fuzziness experiments (MD01). By comparison with the 30-day experiments, these experiments provide a more or less instantaneous view of the dynamics.

More precisely, the experiments are done using short, here 1-day, SL time integrations as follows. For a given PV field at some initial time $t = t_i$, usually taken as 5 days, a PV inversion is done using the inversion operator that defines one of the PBMs, plain or hyperbalance. The resulting fields are used to initialize the PEs, which are then integrated forward to time $t = t_i + 1$, which we find to be sufficient for the generation of imbalance as well as sufficient to allow inertia-gravity waves to propagate well away from their sources. Recall that the domain-crossing time for a gravity wave is of the order of 1 day. Resolution $n_g = 256$ is used throughout. During the time integration, the PEs' state vector $\mathbf{X} = \{u, v, (g/H)^{1/2}h\}$ is stored at equal time intervals, which we take as 0.1 day. At the end of each such time interval the PV field given by the PE integration is inverted, using the same PBM as before, to give another state vector $\mathbf{X}_{\text{bal}} = \{u_{\text{bal}}, v_{\text{bal}}, (g/H)^{1/2}h_{\text{bal}}\}$. This represents a state more or less close to balance depending on the choice of PBM. The difference $\mathbf{X}_{\text{imb}} = \mathbf{X} - \mathbf{X}_{\text{bal}}$ gives us, correspondingly, a more or less accurate estimate of the imbalance. To measure its magnitude we use the same linearized energy norm as in MD01, the square of that norm being averaged over the time interval $(t_i, t_i + 1)$,

$$\|\mathbf{X}_{\text{imb}}\|_2 = \left[\frac{1}{2} H \int_{t_i}^{t_i+1} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left(u_{\text{imb}}^2 + v_{\text{imb}}^2 + \frac{g}{H} h_{\text{imb}}^2 \right) dx dy dt \right]^{1/2}. \quad (7.1)$$

Note that this is a slight redefinition of the symbol $\|\cdot\|_2$, which in the preceding sections involved only spatial averaging and only a single scalar field.

² The QG balanced model inverts and advects the linearized PV, $Q_\epsilon = \zeta - fh/H$ using geostrophic balance.

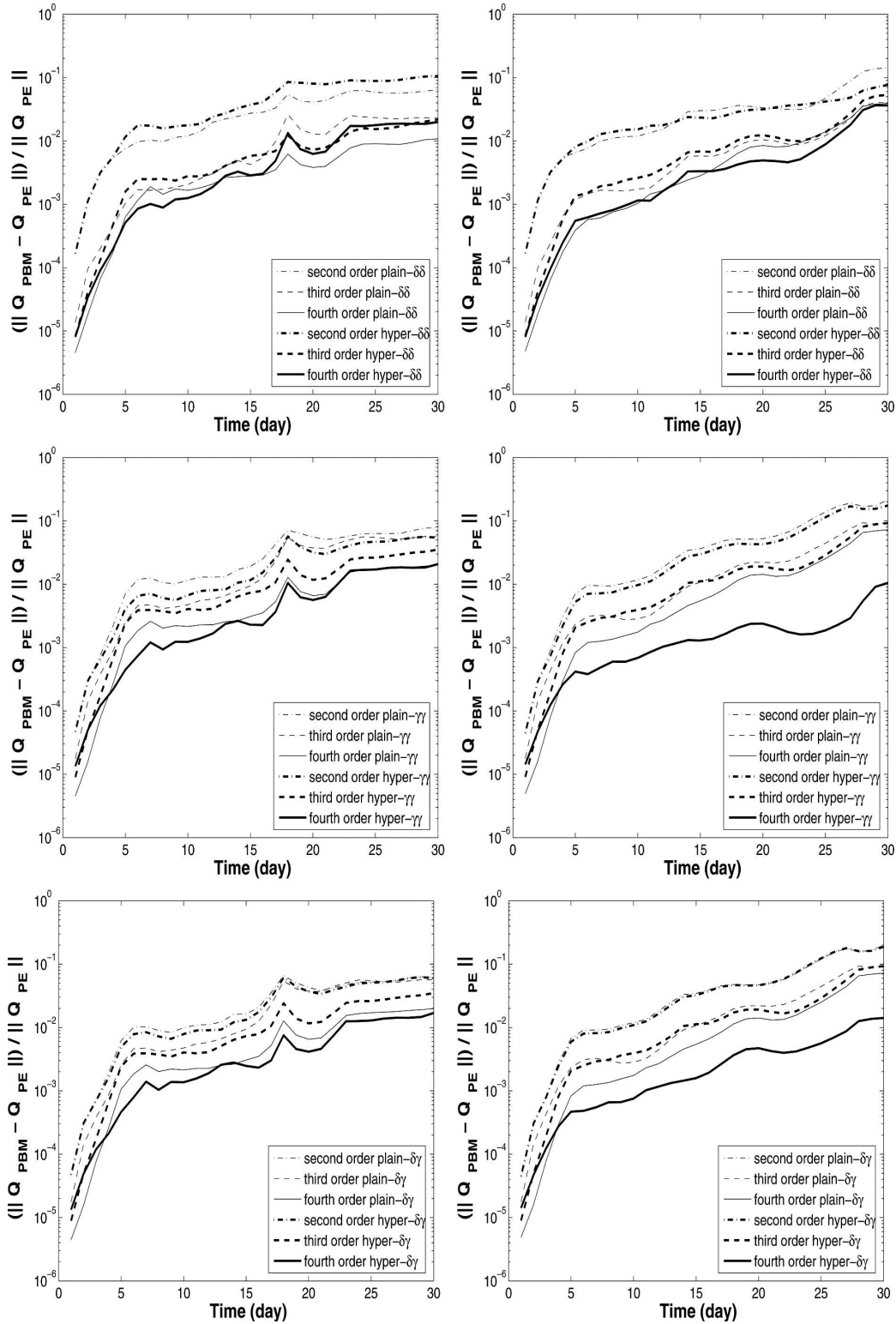


FIG. 6. The thin and thick curves show $\epsilon(t)$ for plain and hyperbalance PBMs of second to fourth order in (left) the case $L_D = 0.55$, and (right) $L_D = 0.60$. The first- and fifth-order PBMs (not shown) all perform less well, just as was found in the $L_D = 0.50$ case, Fig. 5. At large times, the best performance for $L_D = 0.55$ is that of the fourth-order plain- $\delta\delta$ PBM (thin solid curve at upper left), and for $L_D = 0.60$ the fourth-order hyper- $\gamma\gamma$ PBM (thick solid curve at middle right).

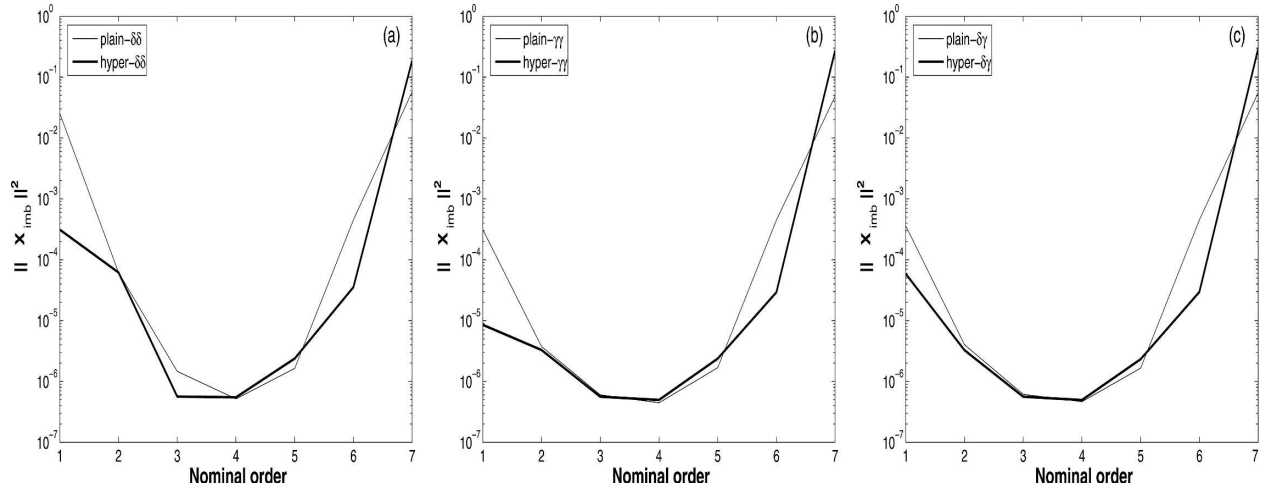


FIG. 7. The time-averaged squared L_2 norm of the imbalance for the $L_D = 0.50$ flow from $t = 5$ to $t = 6$, as estimated more or less accurately via the different PBMs (see text). The estimates via plain PBMs are shown by the thin solid curves, and those via hyperbalance PBMs by thick solid curves. The results are for the plain- and hyper- (a) $\delta\delta$, (b) $\gamma\gamma$, and (c) $\delta\gamma$ balances.

Figure 7 presents $\|\mathbf{X}_{\text{imb}}\|_2^2$ in this new sense for all the 42 PBMs of first to seventh order, for $L_D = 0.50$ and $t_i = 5$. In all cases the initial PV at $t = 5$ is taken from the SL integration of the third-order plain- $\delta\delta$ PBM. For orders higher than the seventh, none of the PV inversions converged. Not surprisingly, these estimates of imbalance reach their minima, broadly speaking, for those PBMs that have already been seen to perform best on cumulative accuracy. The minima can be regarded as our best estimates of the truly irreducible imbalance associated with the fuzziness of the PEs' slow quasimanifold.

But there are other aspects less obvious at first sight. Consider the third-order points plotted in the left panel of Fig. 7 and recall the crossover behavior seen in the top left panel of Fig. 5. The same PBM that conspicuously has the smallest late errors in this case, the third-order plain- $\delta\delta$ PBM, induces the greater early imbalance in the PEs. The corresponding hyperbalance PBM has, conversely, larger late errors and smaller early imbalance, in this particular case. There is a similar though less conspicuous relation between the middle panel of Fig. 7 and the middle right panel of Fig. 5 except that plain and hyper are interchanged.

The crossover might seem paradoxical at first sight. However, it can be made sense of if we suppose that, for the best performing PBMs, which induce low levels of imbalance in the PEs, the imbalance might be acting as background noise in the PE integrations—noise so weak that slightly increasing it brings the PE behavior closer to PBM behavior, in the manner of so-called stochastic resonance (e.g., Hasselmann 1976). This calls for further investigation.

The initialization experiments provide us with an ensemble of PV distributions, for each of which we have available the pairs of balanced states produced by plain and hyperbalance inversion. It is interesting to ask how close those pairs of states might be to each other, as measured by the norm (7.1). We have therefore computed the quantity $\|\mathbf{X}_{\text{bal(p)}} - \mathbf{X}_{\text{bal(h)}}\|_2$ where $\mathbf{X}_{\text{bal(p)}}$ and $\mathbf{X}_{\text{bal(h)}}$ denote the balanced states produced by plain and hyperbalance PV inversion, respectively, for a given order and truncation type. When values of $\|\mathbf{X}_{\text{bal(p)}} - \mathbf{X}_{\text{bal(h)}}\|_2$ are plotted in the manner of Fig. 7 (plots omitted for brevity), we again see the quasi-asymptotic behavior, with deep minima this time all at fourth order, in the $L_D = 0.50$ case, reaching exceedingly small numerical values $< 10^{-8}$.

These smallest values are numerically far smaller than typical values of the corresponding mean imbalance $\|\mathbf{X}_{\text{imb}}\|_2^2 = \frac{1}{2}(\|\mathbf{X}_{\text{imb(p)}}\|_2 + \|\mathbf{X}_{\text{imb(h)}}\|_2)$. For instance, for the fourth-order $\delta\gamma$ truncation we find a deep minimum in the ratio $\|\mathbf{X}_{\text{bal(p)}} - \mathbf{X}_{\text{bal(h)}}\|_2 / \|\mathbf{X}_{\text{imb}}\|_2^2$, with a value of 0.02. Away from the minimum, the values on either side reach ~ 0.2 at first order and ~ 0.3 at seventh order. We may say that, at optimal order, fourth in this case, the slow manifolds of the plain and hyperbalanced PBMs very nearly coincide. Similar results are found for the other truncation types.

The departure of PE from PBM behavior is initiated by the term $\mathbf{v}_{\text{imb}} \cdot \nabla Q$ in the PV equation, where \mathbf{v}_{imb} is the velocity field in \mathbf{X}_{imb} . Figure 8 shows a measure of this, the quantity $\|\mathbf{v}_{\text{imb}} \cdot \nabla Q / |\nabla Q|\|_2$. It shows a pattern similar to that in Fig. 7. Of course there are further terms contributing to the departure of PE from PBM behavior. Once the Q field of the PEs has departed

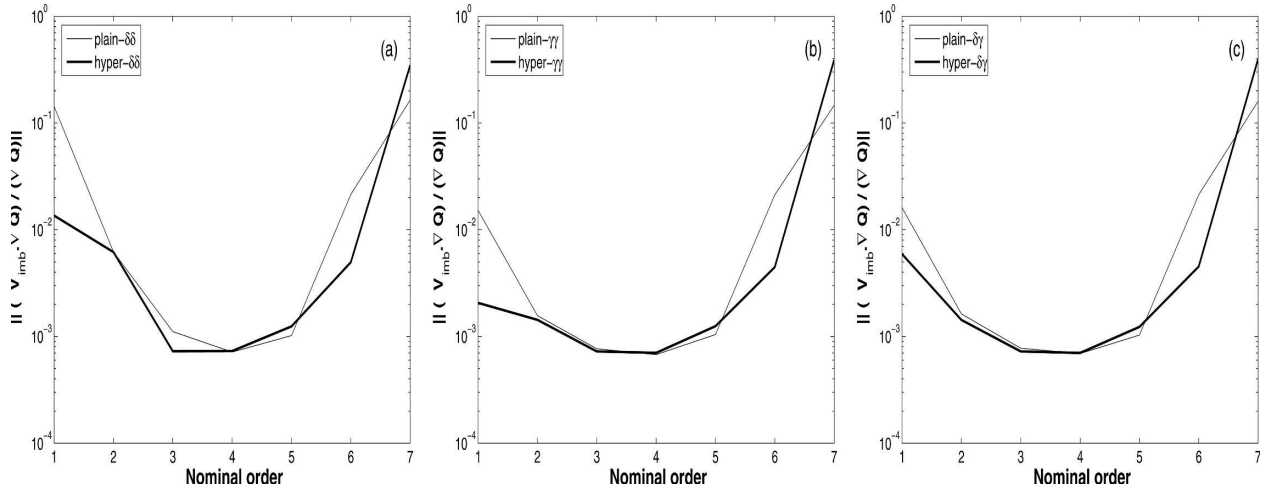


FIG. 8. The time-averaged L_2 norm of $(\mathbf{v}_{\text{imb}} \cdot \nabla Q)/|\nabla Q|$ for the plain balances (thin solid) and the corresponding hyperbalances (thick solid). The results are for the plain- and hyper- (a) $\delta\delta$, (b) $\gamma\gamma$, and (c) $\delta\gamma$ balances. The initial PV field is that obtained from the third-order plain- $\delta\delta$ SL PBM at $t = 5$.

from that of the PBM, by an amount Q' say, there will be an additional contribution $\mathbf{v}' \cdot \nabla Q$ due to the difference between inverting $Q + Q'$ and inverting Q . The intercomparisons in section 6 reveal the combined, cumulative effects of the terms contributing to the departure.

8. Concluding remarks

Perhaps the most important result of this study—aside from confirming yet again the astonishing accuracy of high-order PBMs outside their regimes of asymptotic validity—is the point made at the end of section 6, namely, that on the present numerical evidence, which is quite extensive, MN00’s conjecture appears to be wrong. In other words, there appears to be no systematic trade-off between local mass conservation and accuracy. Therefore, accuracy is not a countervailing consideration against the theoretical advantages of the hyperbalance equations.

Of course MN00’s arguments against energy and momentum conservation still hold. The spontaneous-adjustment emission of inertia-gravity waves by unsteady vortical motions is a physically real, and ubiquitous, phenomenon. It must contribute to the energy and momentum budgets of PE evolution. However, the quantification of such budgets and their PBM counterparts remains a challenge for future work, and will surely require at least as much numerical delicacy as was required in the present study.

Other questions for future work include testing the idea of quasi-stochastic resonance hinted at by the curve crossovers in Figs. 5 and 6, when compared with

the third- and fourth-order imbalances shown in the left and middle panels of Fig. 7. It would be interesting to do the converse experiment of seeing whether adding weak random noise to an accurate PBM might in some circumstances increase its accuracy still further, mimicking the $\mathbf{v}_{\text{imb}} \cdot \nabla Q$ contribution to the PV evolution described by the PEs. It also remains, as always, a challenge to widen still further the parameter regime, especially on the side of larger Froude and Rossby numbers Fr, Ro in order to push the envelope of the balance concept and its ultimate limitations. The dependence of minimum error ϵ upon Fr and Ro is not obvious once we are outside the standard asymptotic regimes in which $Fr \rightarrow 0, Ro \rightarrow 0$.

Finally, as mentioned in section 3, we have carried out some experiments using the CASL instead of the SL algorithm. The CASL algorithm is able to describe realistically sharp-edged PV distributions and has advantages when dealing with nearly frictionless flows, such as occur in the real atmosphere and oceans, for which one would like to avoid the use of numerical models with artificial diffusion. There is nothing obviously pathological about the PV inversion operation itself when applied to sharp-edged PV distributions. However, as was also mentioned in section 3, the ability to handle sharp-edged PV distributions seems to exact a price in terms of noisiness in the way it handles mass rearrangement, in current implementations at least. Furthermore, it does so in ways that are not well understood at present, impacting large as well as small scales. The appendix gives a brief discussion. This story will no doubt continue, since the challenge to cope simultaneously with balance and imbalance in model at-

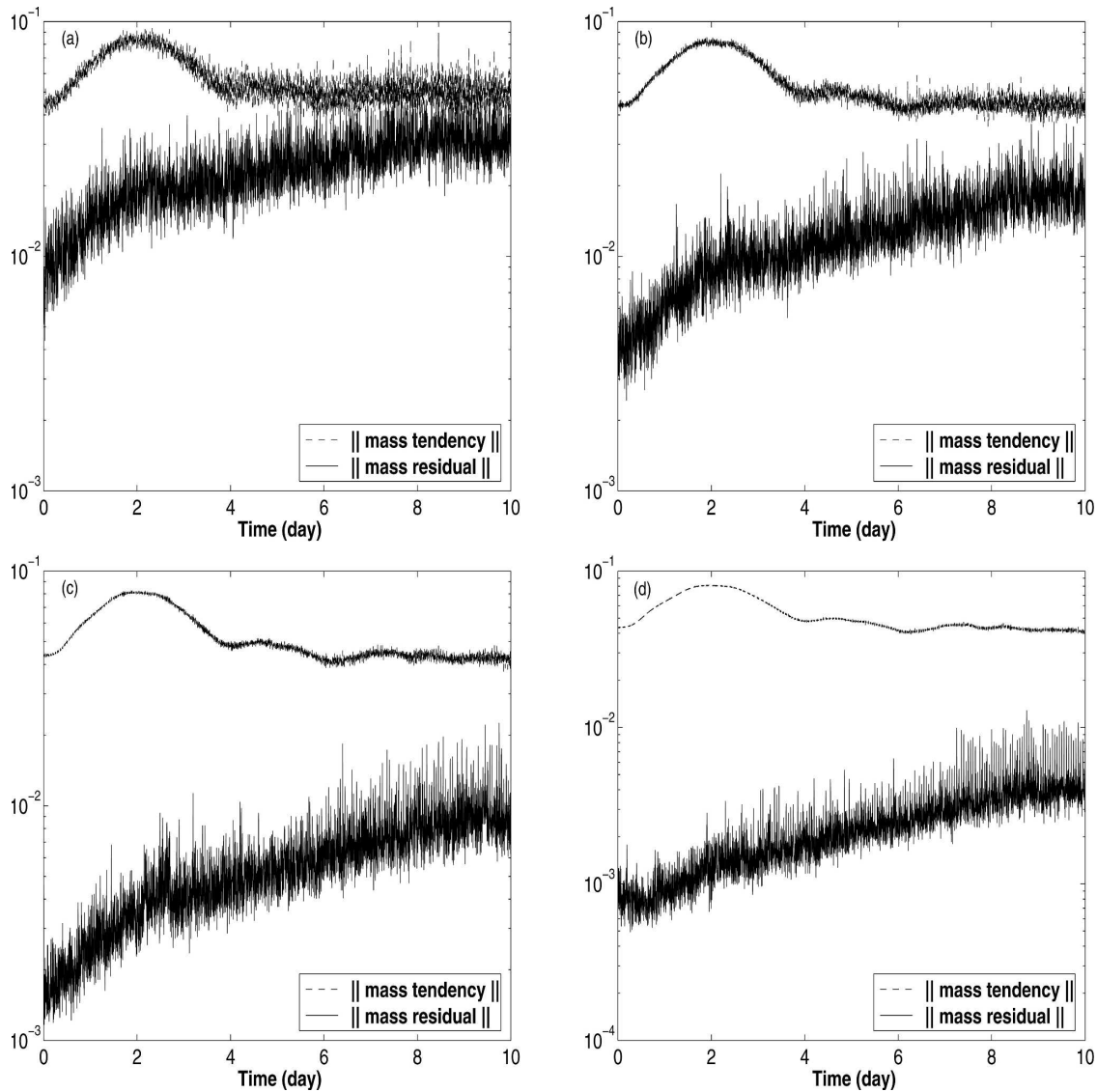


FIG. A1. The L_2 norms of mass tendency and mass residual, $\|\Delta h/\Delta t\|_2$ and $\|m_{\text{res}}\|_2$, vs time for (a) $n_g = 128$, (b) $n_g = 256$, (c) $n_g = 512$, and (d) $n_g = 1024$, where n_g is the number of grid points in each direction. The results are for the Bolin–Charney CASL PBM.

mospheres and oceans will continue to be important, and CASL-type methods will remain advantageous for handling the nearly balanced flows with sharp-edged PV distributions.

Acknowledgments. We thank D. G. Dritschel for helpful discussions. A. R. M. thanks the Iran Ministry of Science, Research and Technology for support in the form of a research scholarship, the Universities of Tehran and St Andrews, and the U.K. Natural Environment Research Council for a Research Fellowship. M. E. M. thanks the Engineering and Physical Sciences Research Council for crucial support in the form of a Senior Research Fellowship.

APPENDIX

CASL and Local Mass Conservation

We briefly explain why the CASL algorithm, despite its other virtues, proved unsuitable for assessing the accuracy of local mass conservation. The reason is noise in the mass residual m_{res} defined in Eq. (4.1). The noise has properties that are not all a priori obvious, as is now shown.

Figure A1 presents the L_2 norms of mass tendency and mass residual, $\|\Delta h/\Delta t\|_2$ (upper curves) and $\|m_{\text{res}}\|_2$ (lower curves), versus time for the Bolin–Charney PBM with the CASL algorithm for advecting PV. The

results for four spatial resolutions determined by the number of grid points in each direction $n_g = 128, 256, 512, 1024$ are shown. The noticeable features are (i) the growth of mass residual with time, (ii) a substantial high-frequency noise in the mass residual, and (iii) the slow rate of decrease of mass residual with resolution. These features are all, in one way or another, responses of the grid-based part of the algorithm for the PBM to the growth in complexity and sharpness of gradients of PV in the CASL simulations. Given that we are dealing with a balanced model here, the second feature, that is, the presence of high frequency noise, may come as a surprise. In fact what it shows is that a substantial rearrangement of mass takes place in the CASL PBM, merely to cope with and maintain sharp gradients of PV. Evidence is provided in Fig. A2 that, surprising as it may seem, this rearrangement is predominantly a large-scale effect in the CASL PBMs. Indeed, the L_2 norm of mass tendency is rather insensitive to resolution. Recall that in the continuum limit, the Bolin–Charney model has zero mass residual. The results for mass residual obtained for the third-order plain- $\delta\delta$ CASL PBM (not shown) tell us that even the high resolution of $n_g = 512$ is not sufficient to obtain a mass residual noticeably higher than that for the Bolin–Charney CASL PBM. More generally, for the CASL algorithm, prohibitively high resolutions are needed in order to detect levels of mass residual in the plain PBMs that are noticeably higher than in their counterpart hyperbalance PBMs. It should be noted that a smoother version of the contour to grid conversion part of the CASL has recently been implemented by D. G. Dritschel (2005, personal communication). However, the latter, smoother, version still leads to similar results for the mass residual, though with some shift to smaller values.

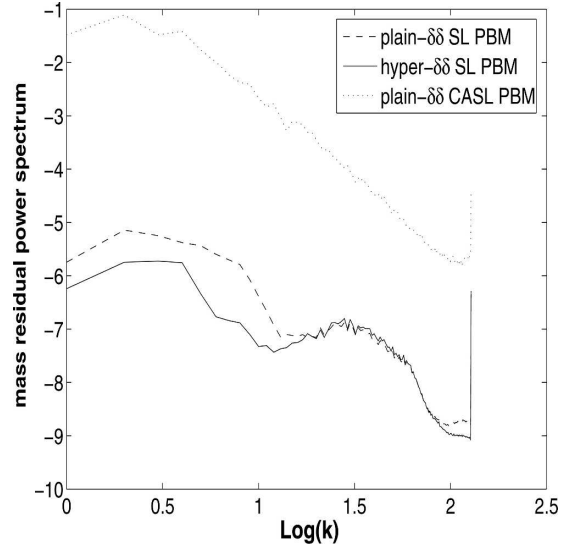


FIG. A2. The power spectrum of mass residual at $t = 10$ for the third-order plain- $\delta\delta$ (dashed), and hyper- $\delta\delta$ (solid) SL PBMs. Also shown is the power spectrum for the third-order plain- $\delta\delta$ CASL PBM (dotted). Both scales are logarithmic.

We note that the high rates of numerical mass rearrangement in the CASL PBMs is the counterpart of the phenomenon of the false generation of imbalance in the CASL for the primitive shallow-water equations in (Q, h, δ) representation (Mohebalhojeh and Dritschel 2000). As demonstrated in Dritschel et al. (1999) and Mohebalhojeh and Dritschel (2000), this false generation has little effect on the PV itself and can be substantially decreased if the primitive equations are instead solved in representations like (Q, δ, γ) or $(Q, \delta, \partial\delta/\partial t)$, depending on the regime of flow. Here $\gamma \equiv f\zeta - g\nabla^2 h$ is the ageostrophic vorticity multiplied by f . The

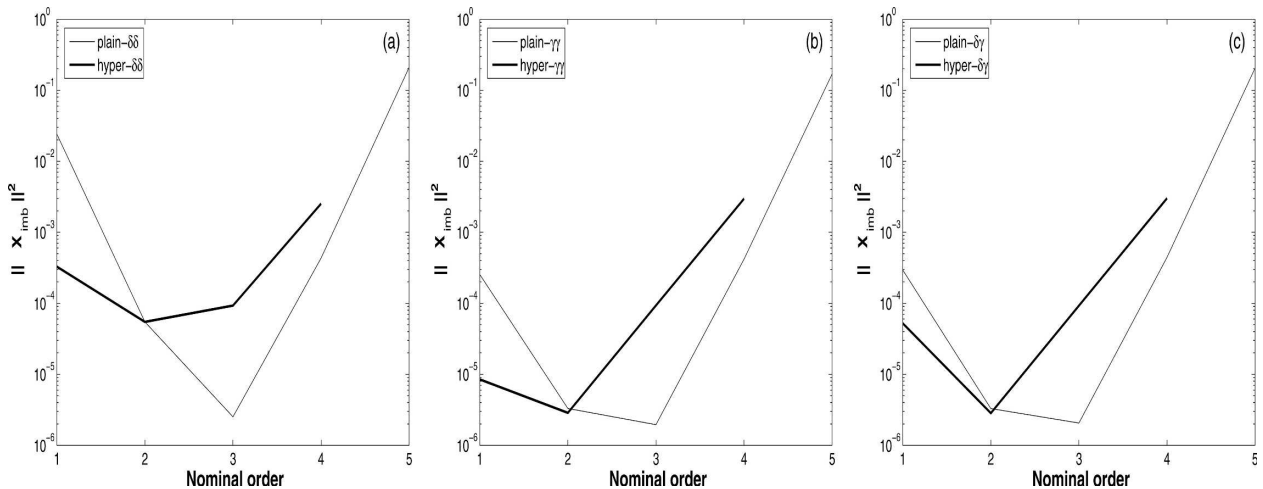


FIG. A3. As in Fig. 7 but the initial PV is that obtained from the third-order plain- $\delta\delta$ CASL PBM.

end result is that the CASL algorithm can achieve higher accuracy than the SL algorithm on both the vortical and the gravity wave parts of the flow. On the same basis, we can expect higher accuracy of the CASL PBMs over the SL PBMs as well.

As a comparison of the structure of mass residual between the CASL and SL PBMs, on the one hand, and between the plain and hyperbalance PBMs, on the other, we present the power spectrum of mass residual at time $t = 10$ for the third-order plain- $\delta\delta$ SL and CASL PBMs and the third-order hyper- $\delta\delta$ SL PBM (Fig. A2). The spectrum for the CASL is clearly steeper than that for the SL. That is, the response to sharper gradients of PV present in CASL compared with SL generate mass rearrangements with larger values and, as already noted, significantly larger [sic] scales. For the SL algorithm, the success of hyperbalance PBM in decreasing $\|m_{\text{res}}\|_2$ shows itself mainly at large scales. The conversion from the plain- $\delta\delta$ PBM to the hyper- $\delta\delta$ PBM leaves the small scales of mass residual unchanged. This insensitivity is another artifact of the numerics arising from the rounding and truncation errors. The sharp spike at the small-scale end is due to the two-grid-interval noise resulting from the aliasing effect of the pseudospectral computation of the derivatives. Since the noise has almost no effect on the time evolution of PV, no attempt has been made to dealias the PBM codes, as already mentioned.

The counterpart to Fig. 7 may be of some interest. It is shown as Fig. A3. The only difference between this and Fig. 7 lies in the initial conditions at $t = t_i = 5$. Here, a sharp-edged PV distribution is used at $t = 5$; the remainder of the initialization experiment is performed using SL integration. So the rather large differences between Figs. 7 and A3 are entirely due to the initial PV distribution.

REFERENCES

- Bleck, R., 1974: Short range prediction in isentropic coordinates with filtered and unfiltered numerical models. *Mon. Wea. Rev.*, **102**, 813–829.
- Dritschel, D. G., and M. H. P. Ambaum, 1997: A contour-advective semi-Lagrangian algorithm for the simulation of fine-scale conservative fields. *Quart. J. Roy. Meteor. Soc.*, **123**, 1097–1130.
- , and A. R. Mohebalhojeh, 2000: The contour-advective semi-Lagrangian algorithm: Keeping the balance. *Proc. Developments in Numerical Methods for Very High Resolution Global Models*, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 119–136. [Available online at http://www.ecmwf.int/publications/library/ecpublications/_pdf/workshop/2000/HighResolution/ws_highres_dritschel.pdf.]
- , L. M. Polvani, and A. R. Mohebalhojeh, 1999: The contour-advective semi-Lagrangian algorithm for the shallow water equations. *Mon. Wea. Rev.*, **127**, 1551–1565.
- Hasselmann, K., 1976: Stochastic climate models. Part I. Theory. *Tellus*, **28**, 473–485.
- McIntyre, M. E., and W. A. Norton, 2000: Potential vorticity inversion on a hemisphere. *J. Atmos. Sci.*, **57**, 1214–1235; Corrigendum, **58**, 949.
- Mohebalhojeh, A. R., 2002: On shallow-water potential-vorticity inversion by Rossby number expansions. *Quart. J. Roy. Meteor. Soc.*, **128**, 679–694.
- , and D. G. Dritschel, 2000: On the representation of gravity waves in numerical models of the shallow water equations. *Quart. J. Roy. Meteor. Soc.*, **126**, 669–688.
- , and —, 2001: Hierarchies of balance conditions for the f -plane shallow water equations. *J. Atmos. Sci.*, **58**, 2411–2426.
- , and —, 2004: Contour-advective semi-Lagrangian algorithms for many-layer primitive equation models. *Quart. J. Roy. Meteor. Soc.*, **130**, 347–364.
- , and M. E. McIntyre, 2007: Local mass conservation and velocity splitting in PV-based balanced models. Part I: The hyperbalance equations. *J. Atmos. Sci.*, **64**, 1782–1793.
- Rossby, C. G., 1936: Dynamics of steady ocean currents in the light of experimental fluid mechanics. *Pap. Phys. Oceanogr. Meteor.*, **5**, 1–43.
- Temperton, C., and A. Staniforth, 1987: An efficient two-time-level semi-Lagrangian semi-implicit integrating scheme. *Quart. J. Roy. Meteor. Soc.*, **113**, 1025–1040.
- Warn, T., O. Bokhove, T. G. Shepherd, and G. K. Vallis, 1995: Rossby number expansions, slaving principles, and balance dynamics. *Quart. J. Roy. Meteor. Soc.*, **121**, 723–739.