Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 227 (2008) 3715–3730
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Some conservation issues for the dynamical cores of NWP and climate models

J. Thuburn

Mathematics Research Institute, School of Engineering, Computer Science and Mathematics, University of Exeter, Exeter EX4 4QE, UK

Received 11 May 2006; received in revised form 8 August 2006; accepted 16 August 2006
Available online 17 October 2006

Abstract

The rationale for designing atmospheric numerical model dynamical cores with certain conservation properties is reviewed. The conceptual difficulties associated with the multiscale nature of realistic atmospheric flow, and its lack of time-reversibility, are highlighted. A distinction is made between robust invariants, which are conserved or nearly conserved in the adiabatic and frictionless limit, and non-robust invariants, which are not conserved in the limit even though they are conserved by exactly adiabatic frictionless flow. For non-robust invariants, a further distinction is made between processes that directly transfer some quantity from large to small scales, and processes involving a cascade through a continuous range of scales; such cascades may either be explicitly parameterized, or handled implicitly by the dynamical core numerics, accepting the implied non-conservation. An attempt is made to estimate the relative importance of different conservation laws. It is argued that satisfactory model performance requires spurious sources of a conservable quantity to be much smaller than any true physical sources; for several conservable quantities the magnitudes of the physical sources are estimated in order to provide benchmarks against which any spurious sources may be measured.

Keywords: Weather prediction; Climate simulation; Conservation

1. Introduction

The formulation of a numerical model of the atmosphere is usually considered to be made up of a dynamical core, and some parameterizations. Roughly speaking, the dynamical core solves the governing fluid and thermodynamic equations on resolved scales, while the parameterizations represent subgrid scale processes and other processes not included in the dynamical core such as radiative transfer. Here, no attempt is made to give a precise definition of 'dynamical core' because, as discussed below, there are some open questions concerning exactly which terms and which processes should be included in a dynamical core.

It is usually taken for granted that it is beneficial for a dynamical core to possess discrete analogues of the conservation properties of the continuous adiabatic frictionless governing equations. However, these continuous equations possess an infinite number of conserved quantities (see Section 2), whereas a numerical model

E-mail address: j.thuburn@ex.ac.uk

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2006.08.016

can have only a small finite number of analogous properties. Therefore, in the design of a numerical scheme, some choice must be made as to which conservation properties are most desirable. In addition, other possible properties of a numerical scheme, such as good wave dispersion properties, smooth behaviour near the grid scale, or computational efficiency, might be incompatible with certain conservation properties, and some decision must be made about their relative importance. This paper reviews some of the issues related to conservation in order to gain some understanding of why conservation properties are important (Section 3), and hence to help decide which conservation properties are most important, and whether discrete conservation properties should take priority over other desirable properties. The focus is on the context of numerical weather prediction and climate models; the issues may be rather different for models simulating much smaller scales.

An important set of questions concerns how accurately a dynamical core needs to capture various conservation properties. Is global conservation sufficient or is satisfaction of a local conservation law essential? Is a conservative spatial discretization sufficient, or must the full space–time discretization be conservative? In Section 3 it is argued that we can expect accurate solutions provided the timescale for any spurious numerical sources is long compared to the timescale for true physical sources. In Sections 4–9 the physical sources for various conservable quantities are estimated in order to provide benchmarks against which spurious sources may be compared.

The starting point for the consideration of conservation issues is often the adiabatic frictionless governing equations, which have a number of conservation properties (see Section 2), and are time-reversible. However, realistic solutions of those equations lead to flows in which some non-conservation, often due to small-scale mixing, would be inevitable for any values of viscosity and diffusivity no matter how small. For example, layerwise two-dimensional turbulence involves a transfer to small scales and inevitable non-conservation of potential enstrophy, while vertical propagation of gravity waves leads to wave breaking and inevitable small-scale vertical mixing. Therefore, certain quantities that are conserved by truly adiabatic frictionless flow will not be conserved in the adiabatic frictionless *limit*. The conservation of such quantities is thus *non-robust*, and the adiabatic frictionless limit is not time-reversible. This suggests that dynamical cores need not, and perhaps should not, be time-reversible.

A related set of issues arises because of the strongly multiscale nature of realistic atmospheric flows, so that some flow features are inevitably below the resolution of a given numerical model. First, some conservable quantities, such as tracer variance (Section 6) and energy (Section 7), have contributions from unresolved scales; design of a numerical model to conserve just the resolved contribution therefore requires careful justification. Second, some decision must be made about exactly which equation set the dynamical core is supposed to solve. Large-eddy simulation, for example, is based on equations that explicitly average out unresolvable scales (e.g. [44]). Dynamical cores for large-scale atmospheric models should also be thought of as solving averaged equation sets, though this is rarely made explicit. Which conservation properties survive the averaging, if any, will depend on the type of averaging (Eulerian or Lagrangian, mass-weighted or not, etc.), but little work appears to have been done to address this last question.

One must also decide, particularly for non-robust invariants, whether only resolved scales are to be handled by the dynamical core, with all unresolved scales and downscale transfer processes handled by parameterizations or subgrid models, or whether some contributions from unresolved scales and downscale transfer processes are to be handled by the dynamical core. Here we make a distinction between two classes of downscale transfer process.

(i) Processes in which there is a transfer directly from large scales to small scales, missing out a range of intermediate scales that includes the model truncation scale. An example is boundary layer turbulence, which directly transfers energy from the large-scale flow to small-scale boundary layer eddies. Such processes must always be parameterized in practice.
(ii) Processes involving a downscale transfer or ‘‘cascade’’ across a continuous range of scales that includes the model truncation scale. An example is the downscale cascade of tracer variance due to advective straining by the large-scale flow. For these class (ii) processes it appears we have a choice either to parameterize or to allow the dynamical core numerics to handle the resolved-scale to unresolved-scale transfers (Section 6).

The use of scale-selective dissipation in numerical models is relevant to this last point. All numerical models of the atmosphere include some form of scale-selective dissipation, either explicitly specified or inherent in the numerical schemes used. It serves a variety of purposes, including cleaning up noise generated by dispersion errors, computational modes, and physical parameterizations, crudely representing subgrid Reynolds stresses, and soaking up tracer variance and potential enstrophy that cascade downscale to prevent spectral blocking (see Section 6). Even if the scale-selective dissipation is switchable it is usual to include it, even in nominally adiabatic frictionless integrations, since without it models behave badly. Thus it is reasonable to consider the scale-selective dissipation to be an essential part of the dynamical core [71]. Any claim about the conservation properties of a dynamical core should take into account the scale-selective dissipation needed to integrate the dynamical core in practice.

Finally, although it is well known that a numerical model can possess only a finite number of independent conserved quantities, it appears to be rather poorly understood exactly how many and what combinations of conserved quantities are possible. Some comments are made on this question in Section 10.

2. Quantities conserved by the adiabatic frictionless governing equations

For adiabatic frictionless flow, in the absence of condensation and evaporation and chemical sources and sinks, the equation sets normally used as the basis for atmospheric modelling have a number of local, flux-form conservation laws of the form

$$\frac{\partial A}{\partial t} + \nabla \cdot \mathbf{F} = 0, \tag{1}$$

where A is the density of the conserved quantity and $\mathbf{F}$ is the flux.

(1) Mass: $A = \rho$ and $\mathbf{F} = \rho\mathbf{u} = \mathbf{u}A$, where ρ is the air density and $\mathbf{u}$ is the three-dimensional air velocity.
(2) Angular momentum: $A = \rho\widehat{\mathbf{z}} \cdot [\mathbf{r} \times (\mathbf{u} + \mathbf{\Omega} \times \mathbf{r})]$ and $\mathbf{F} = \mathbf{u}A + p\widehat{\mathbf{z}} \times \mathbf{r}$, where $\mathbf{\Omega}$ is the Earth's rotation vector, $\widehat{\mathbf{z}}$ is a unit vector in the same direction, $\mathbf{r}$ is the position vector relative to the origin at the centre of the Earth, and p is the pressure.
(3) Energy: $A = \rho(\frac{1}{2}\mathbf{u}^2 + c_vT + \Phi)$ and $\mathbf{F} = \rho\mathbf{u}(\frac{1}{2}\mathbf{u}^2 + c_pT + \Phi) = \mathbf{u}(A + p)$, where T is temperature, p is pressure, c_v and c_p are the specific heat capacities at constant volume and constant pressure respectively, and Φ is the geopotential.
(4) Potential temperature: $A = \rho f(\theta)$ and $\mathbf{F} = \rho\mathbf{u}f(\theta) = \mathbf{u}A$, for any function f of the potential temperature θ. This includes the specific entropy $f(\theta) = \eta = c_p \ln\theta$.
(5) Potential vorticity (PV): $A = \rho g(Q)$ and $\mathbf{F} = \rho\mathbf{u}g(Q) = \mathbf{u}A$, for any function g of the potential vorticity $Q = \boldsymbol{\zeta} \cdot \nabla f(\theta)/\rho$, where $\boldsymbol{\zeta}$ is the absolute vorticity vector and again f is any function of the potential temperature. Of particular interest is $g(Q) = Q^2/2$; A is then the density of potential enstrophy.
(6) Tracer: $A = \rho h(\chi)$ and $\mathbf{F} = \rho\mathbf{u}h(\chi) = \mathbf{u}A$, where χ is the tracer mass mixing ratio and h is any function of χ. This applies in particular to water vapour in the absence of condensation and evaporation.

By integrating the local flux-form conservation law (1) over the entire domain, with suitable boundary conditions, it follows that the global integral of each of these quantities is invariant. Note that the conservation laws for potential temperature, PV, and tracer each imply an infinite number of global invariants. The global invariants corresponding to the arbitrary functions of potential temperature and PV are known as Casimirs (e.g. [56]).

The conservation laws for potential temperature, PV, and tracer can be combined with conservation of mass to obtain Lagrangian conservation laws:

$$\frac{Df(\theta)}{Dt} = 0, \tag{2}$$

$$\frac{Dg(Q)}{Dt} = 0, \tag{3}$$

$$\frac{Dh(\chi)}{Dt} = 0. \tag{4}$$

The conservation laws for mass and potential temperature together imply that the global integral over any isentropic surface of the mass per unit θ,

$$\mathscr{F}(\theta) = \int \frac{\rho}{|\nabla\theta|}\,\mathrm{d}A, \tag{5}$$

is conserved, and the conservation laws for mass, potential temperature and PV together imply that $\int \rho g(Q)/|\nabla\theta|\,\mathrm{d}A$ over any isentropic surface is also conserved. (Note here $\mathrm{d}A$ is the element of area of the surface, not its horizontal projection.) Moreover, conservation of mass, potential temperature and PV imply that, for any material contour Γ lying on an isentropic surface bounding a region D, the mass per unit θ within the contour $\mathscr{M} = \int_D \rho/|\nabla\theta|\mathrm{d}A$ and the absolute circulation around the contour $\mathscr{C} = \oint_\Gamma \mathbf{u}_a \cdot \mathrm{d}\mathbf{r} = \int_D \rho Q/|\nabla\theta|\mathrm{d}A$, where $\mathbf{u}_a$ is the absolute velocity, are both conserved (e.g. [56]). In particular, since PV is materially conserved, Γ may be a PV contour. In this case the distributions of $\mathscr{M}(Q)$ and $\mathscr{C}(Q)$ are related by

$$\frac{\partial\mathscr{C}}{\partial Q} = Q\frac{\partial\mathscr{M}}{\partial Q}, \tag{6}$$

(irrespective of whether $\mathscr{M}(Q)$ and $\mathscr{C}(Q)$ are conserved [63]), so that in fact $\mathscr{M}(Q)$ and $\mathscr{C}(Q)$ are not independent.

For a Hamiltonian system, such as the equations describing adiabatic frictionless atmospheric dynamics, Noether's theorem relates conservation properties to symmetries of the Hamiltonian such as invariance under rotation or time translation (e.g. [56]). When considering disturbances to a basic state that has some symmetry (e.g. zonal symmetry or steadiness), Noether's theorem implies that certain measures of wave activity (e.g. pseudomomentum or pseudoenergy) satisfy conservation laws (e.g. [1,29,56]). The conservation laws can be explicitly constructed by combining the conservation laws for angular momentum or energy with those for the Casimirs (e.g. [47,29]); in particular, the pseudomomentum can be expressed partly in terms of $\mathscr{M}(Q)$ and $\mathscr{C}(Q)$ [63]. However, despite their obvious dynamical importance, such conservation laws do not appear to have been used as criteria in the development of numerical models.

The expressions for conserved quantities listed above are valid for the deep atmosphere compressible Euler equations (though with the approximation that Φ is spherically symmetric so that $\mathbf{r} \times \nabla\Phi$ can be taken as zero). When further approximations are made to the governing equations, it is nearly always considered desirable to retain analogous conservation properties in the approximate equations, though the expressions for the conserved quantities and their fluxes must be modified appropriately (e.g. [70]). Retaining conservation properties in approximate equation sets can be guaranteed by deriving them from a Hamiltonian that has been approximated in such a way as to retain the corresponding symmetry (e.g. [56]). However, it should be noted that retaining conservation properties in approximate equation sets might not always be appropriate. For example, if there were a significant systematic transfer of energy from balanced vortical motion to gravity wave motion then conserving energy in an approximate equation set that describes only the balanced component of the flow could be inappropriate (e.g. [46]).

3. Why is conservation desirable in a numerical model, and how well do we need to conserve?

There are a few general reasons why it may be desirable for a numerical model to possess discrete analogues of conservation laws:

(1) Solutions of the equations describing atmospheric dynamics are in general extremely complicated, and there are very few exact analytical statements we can make about them. We should, at least, try to capture those analytical properties that we do know for certain, such as the conservation properties.
(2) The accuracy of certain aspects of a numerical simulation can be closely tied to conservation properties. In the simple case of Burgers equation, for example, conservation of "momentum" is crucial to capture accurately the speed of any shock that forms (e.g. [38]). For full atmospheric models it has been argued that conservation of energy and enstrophy is important for accurately capturing nonlinear transfers to small scales (e.g. [54,2]). Also, it has been argued [17] that the accuracy of semi-Lagrangian schemes

could be improved if trajectories were constrained to be consistent with the mass continuity equation. These examples all argue for the satisfaction of a local flux-form conservation law, not merely global conservation.

(3) Some early attempts to model atmospheric dynamics suffered from problems of instability, particularly nonlinear instability, in which nonlinear interactions near the grid scale lead to aliasing and an amplification of the total energy (e.g. [2]). Ensuring conservation of energy, and perhaps potential enstrophy, can help to suppress such instability. (However, the energy-conserving centred-difference spatial discretization for Burgers equation in one dimension demonstrates that energy conservation alone does not prevent the build-up of grid-scale oscillations, even though it prevents the unbounded growth of energy associated with nonlinear instability, e.g. [21].)

(4) There is currently no existence proof for solutions to the compressible Euler equations for finite times, though such proofs do exist for some approximate equation sets valid in certain asymptotic limits. Cullen [15] argues that those properties of the equations that are crucial for the existence proofs are also crucial for controlling the dynamics, and are therefore the properties that need to be captured by numerical methods. Cullen discusses these ideas in detail and gives several examples, including mass conservation, and boundedness under advection of materially conserved quantities such as potential temperature and potential vorticity.

(5) Each conservation property that can be enforced reduces by one the dimension of the manifold that the solution is constrained to follow. This might seem like a relatively small benefit for a numerical weather prediction model with $O(10^7)$ degrees of freedom. However, if the constraints are physically important then the benefit is likely to be greater.

(6) Exact satisfaction of a discrete conservation law allows the diagnosis of a closed budget for the conserved quantity. This can be especially important when that particular budget is the subject of scientific interest.

(7) If a numerical scheme is formally exactly conserving then checking conservation in practice can be valuable for debugging.

Beyond these general considerations, we need some rationale for deciding which particular conservation properties are most important, and how accurately they need to be respected in models. One possible framework for addressing these issues is the following. In the real atmosphere all conservable quantities have sources and sinks, and the value of the conservable quantity will be determined by the balance between its sources and sinks. Hence, the errors that arise from imperfect conservation in a dynamical core will depend on the magnitude of the imperfection in conservation compared to the true physical sources and sinks. It is convenient to express the sources and sinks in terms of timescales to allow comparison between different conservable quantities. It is then reasonable to argue that the conservation properties with the longest physical source–sink timescales are the most important. Moreover, the physical source–sink timescale for any property provides a benchmark against which to assess the magnitude of any imperfection in conservation; satisfactory performance will require the timescale for spurious sources to be much longer than the physical source–sink timescale (or much longer than the model integration time—conservation properties will be often more important for climate simulation than for numerical weather prediction simply because of the longer timescale available for errors to accumulate and become significant). Sections 4–9 below discuss the roles of some conservation properties in the global circulation, and estimates are made for the physical source–sink timescales to begin to address these questions.

4. Mass

Conservation of mass is arguably the most fundamental conservation property. Mass is a robust invariant, conserved irrespective of diabatic or frictional processes. The true physical sources of dry mass are minute—its timescale is, for practical purposes, infinite—so any imperfection in conservation of mass will eventually be noticeable in a long enough integration. The mass distribution is intimately tied to the pressure distribution through the equation of state, and spurious mass sources will lead to spurious motions. A systematic drift in mass will imply a systematic drift in mean surface pressure, with numerous secondary effects. Moreover, many other conservation laws depend crucially on conservation of mass; failure to conserve mass will preclude many other conservation properties. Conservation of mass of long-lived tracers (or families of tracers, such as total reactive nitrogen or chlorine) is crucial for modelling atmospheric chemistry and composition. Moist processes are strongly nonlinear and are likely to be particularly sensitive to imperfections in conservation of water. Thus there is a very strong argument for requiring a dynamical core to conserve mass of air, water, and long-lived tracers, particularly for climate simulation.

Currently most if not all atmospheric models fail to make proper allowance for the change in mass of an air parcel when water vapour condenses and precipitates out. A typical formulation in terms of virtual temperature implicitly replaces the condensed water vapour by an equal volume of dry air. This approximation can lead to noticeable forecast errors in surface pressure during heavy precipitation, for example. However, the approximation will not lead to a systematic long term drift in the atmospheric mass in climate simulations provided there is no long term drift in the mean water content of the atmosphere.

5. Momentum and angular momentum

Angular momentum is a non-robust invariant: there is a finite exchange across the lower boundary even in the large Reynolds number limit. However, unlike tracer variance and potential enstrophy (Section 6), it is non-robust because of a spatial transfer to a boundary rather than a downscale transfer to some dissipation scale.

The momentum equation is fundamental to atmospheric dynamics. However, on synoptic scales, its vertical component is dominated by hydrostatic balance. Thus, spurious sources of vertical momentum are likely to be important to the extent that they disrupt hydrostatic balance. The timescale for adjustment towards hydrostatic balance is of the order of the inverse of the buoyancy frequency N (a few tens of seconds, e.g. [4]) or less. Therefore spurious sources of vertical momentum are likely to be significant if their timescale is comparable to this or shorter. Similarly, in the extratropics the horizontal components of the momentum equation are dominated locally by approximate geostrophic balance. Spurious sources of horizontal momentum are likely to be significant if their timescale is comparable to or shorter than the inertial timescale $1/f$ (a few tens of hours), which is the timescale for adjustment towards geostrophic balance.

On a global scale, however, the above argument does not apply to zonal mean angular momentum, and the timescale for zonal mean angular momentum is much greater. Conservation of angular momentum is important for the zonal wind strength, particularly midlatitude jets and trade winds as well as other features such as the stratospheric winter vortex and Quasi-Biennial Oscillation. Through the Earth's rotation, the angular momentum budget is closely tied to the mean meridional circulation and its transport effects. It has been suggested that the angular momentum budget might provide a crucial feedback helping to maintain climate stability [6]. The total angular momentum of the Earth's atmosphere varies between about plus and minus 0.4×10^{26} kg m^2 s^{-1} on a seasonal timescale. The total surface torque (friction plus mountain) is close to zero but has a large eastward component in the tropics and a large westward component in the midlatitudes, both of order 0.5×10^{20} kg m^2 s^{-2} [51]. Thus the global time scale for atmospheric angular momentum is about 10 days. If imperfections in angular momentum conservation in a numerical model acted on timescales comparable to this then they would have a significant effect on the global angular momentum budget. Note, however, that in some regions of the atmosphere the angular momentum timescale is locally much longer. It is of order two years in the tropical lower stratosphere, associated with weak wave-induced angular momentum flux convergence. Thus, even quite small errors in the angular momentum budget could compromise the simulation of the Quasi-Biennial Oscillation, and the difficulty of obtaining a realistic simulation of the Quasi-Biennial Oscillation in climate models reflects this.

6. Tracer variance and potential enstrophy

Correct tracer variance and potential enstrophy budgets are important for maintaining correct variability in models. Potential enstrophy budgets for the atmosphere do not appear to have been computed. However, enstrophy budgets based on analyses [35,61] should give useful estimates for its sink timescale. These suggest a globally averaged enstrophy of order 10^{-9} s^{-2} and a globally averaged downscale enstrophy transfer of order

$10^{-15}\,\mathrm{s}^{-3}$, implying an enstrophy sink timescale of order 10 days. This timescale is comparable to an eddy turnover timescale, as might be expected from naive scaling arguments. A similar timescale might be expected for the downscale transfer of tracer variance. Some support for this is given by calculations of "mixdown time" for atmospheric tracers [65,30], which do indeed suggest a timescale of order 10–20 days for the lower stratosphere.

In the rest of this section two related arguments are made against attempting to conserve tracer variance in a dynamical core. The first is that the quantity diagnosed as variance in a numerical model is not the total variance but only the resolved contribution. The second is that variance inevitably cascades to small scales and must eventually be dissipated in any realistic flow. Similar arguments apply to other moments of the tracer distribution and to potential enstrophy on horizontal scales smaller than the Rossby radius.

Let the total mass of a tracer be

$$M = \int \rho\chi\,\mathrm{d}\mathbf{x}, \tag{7}$$

and also consider

$$R = \int \rho\chi^2\,\mathrm{d}\mathbf{x}. \tag{8}$$

(R is equal to the variance plus a constant when M and mass of air are conserved). In the absence of sources and sinks both M and R will be conserved. In a finite volume discretization the natural prognostic variables are the volume-weighted cell-average density

$$\bar{\rho}_i = \frac{\int_{\mathrm{cell}\,i}\rho\,\mathrm{d}\mathbf{x}}{\int_{\mathrm{cell}\,i}\mathrm{d}\mathbf{x}} = \frac{\mu_i}{v_i}, \tag{9}$$

(where i is the cell index) and the mass-weighted cell-average tracer

$$\widetilde{\chi}_i = \frac{\int_{\mathrm{cell}\,i}\rho\chi\,\mathrm{d}\mathbf{x}}{\int_{\mathrm{cell}\,i}\rho\,\mathrm{d}\mathbf{x}} = \frac{m_i}{\mu_i}. \tag{10}$$

Then the contribution from cell i to the total mass of tracer can be expressed simply in terms of prognostic variables

$$m_i = \int_{\mathrm{cell}\,i}\rho\chi\,\mathrm{d}\mathbf{x} = \bar{\rho}_i\widetilde{\chi}_i v_i, \tag{11}$$

and it is clear what constraint the prognostic variables must satisfy in order to conserve the total mass of tracer. On the other hand, the contribution to R from cell i is

$$r_i = \int_{\mathrm{cell}\,i}\rho\chi^2\,\mathrm{d}\mathbf{x}. \tag{12}$$

Write χ in terms of the cell-average value and a departure from that cell-average value: $\chi = \widetilde{\chi}_i + \chi'$. Then

$$r_i = \bar{\rho}_i\widetilde{\chi}_i^2 v_i + \int_{\mathrm{cell}\,i}\rho\chi'^2\,\mathrm{d}\mathbf{x}. \tag{13}$$

The first term on the right is the resolved contribution to r_i while the second term is the unresolved contribution. There is no justification for demanding that only the resolved contribution $\sum_i\bar{\rho}_i\widetilde{\chi}_i^2 v_i$ should be conserved. Instead, the resolved contribution should be allowed to change, and such changes should be interpreted as exchanges between resolved and unresolved contributions. The above argument easily generalizes for the conservation of any quantity that is a nonlinear function of the prognostic variables (but note that we can consider momentum $\sum_i\bar{\rho}_i\widetilde{u}_i$ and the internal contribution to energy $c_v\sum_i\bar{\rho}_i\widetilde{T}_i$ as analogous to tracer mass, so that these quantities can be entirely resolved).

In any realistically complex flow there is a systematic downscale transfer of tracer variance and potential enstrophy: in an initial value problem their spectra shift systematically towards larger wavenumbers as advection by straining flow draws out tracer and PV features into long thin streamers. This transfer implies that these conserved quantities are not *robust*; in any real physical situation the variance or potential enstrophy must inevitably be transferred to scales at which dissipation will become important, and then they will no longer be conserved. Moreover, the downscale transfers of variance and potential enstrophy act across a continuous range of scales, i.e. they are cascades, or class (ii) transfers in the terminology of Section 1.

The downscale cascades of tracer variance and potential enstrophy imply systematic transfers from resolved scales to unresolved scales. The implication is that a dynamical core should not attempt to conserve only the resolved contribution, but should allow the resolved contribution to decrease, and this decrease should be interpreted as a transfer to unresolved scales, with the understanding that it would ultimately be dissipated.

Similar arguments apply to the PV-contour mass and circulation integrals $\mathscr{M}(Q)$ and $\mathscr{C}(Q)$. They too are non-robust; they are changed by the generation and eventual mixing of small-scale PV features. For similar reasons, pseudomomentum and pseudoenergy will be non-robust when waves become strongly nonlinear and break, leading to irreversible mixing.

It should be clear from the above argument that a scheme that conserves the resolved variance or potential enstrophy must behave badly in realistic flows. On well-resolved scales there will be a transfer to smaller scales, but near the grid scale this transfer would be halted and variance would build up near the grid scale, leading to a noisy solution, a process known as "spectral blocking".

If we accept that a dynamical core should allow the resolved tracer variance and potential enstrophy to decrease, to represent the cascade to unresolved scales, two strategies are then possible. The first is to use a numerical scheme that conserves the resolved variance or potential enstrophy, but supplement it by an explicit scale-selective dissipation (see Section 1). One difficulty of this approach is that the cascade operates across a continuous range of scales, and there is no 'spectral gap' between scales to be resolved and scales to be parameterized. Another is that the scale-selective dissipation must be tuned, depending on the grid scale and on the flow regime, to ensure that the dissipation length scale is not smaller than the grid scale (to prevent the build-up of noise) but not much larger than the grid scale (to prevent excessive smoothing). See e.g. [60,67] for some related discussion. The second strategy is to use an advection scheme that is accurate but inherently sufficiently dissipative (i.e. variance reducing) near the grid scale to prevent the build-up of grid-scale noise. This strategy is possible provided the cascade is driven by well-resolved scales while small scales are essentially passive. Odd-order finite volume schemes, supplemented by some form of "limiter" to prevent spurious oscillations, seem to work well in this respect [62], and have the advantage that they automatically ensure a near-optimal dissipation scale without flow-dependent tuning. Similar arguments should apply to semi-Lagrangian schemes with even-order interpolation (i.e. such that the advection scheme overall is odd-order and the leading truncation error is dissipative rather than dispersive, e.g. [45]). Whichever strategy is adopted, some representation of subgrid processes is then built into the dynamical core.

Note that the emphasis in this discussion is on representing the transfer of tracer variance from resolved to unresolved scales, rather than representing the ultimate molecular diffusive dissipation of tracer variance; of course the rates of the two processes should be equal when averaged over time and space, provided no other process changes the sub-grid scale tracer variance. Note also that, although the ultimate dissipation of tracer variance is a real physical process, we do not insist that it must be represented by a corresponding physical parameterization. An explicit scale-selective dissipation could be interpreted as such a parameterization (or, rather, a parameterization of the resolved to unresolved scale transfer). On the other hand, a suitably formulated dynamical core appears to be able to represent the resolved to unresolved scale transfer at least as well.

One final remark is appropriate. Although theories of two-dimensional and geostrophic turbulence, and discussions of conservative numerical schemes, pay great attention to conservation of potential enstrophy, there is virtually no discussion of why potential enstrophy, rather than any of the other, infinitely many, $g(Q)$ should be accorded such special status.

7. Energy

The energy of the atmosphere is made up of total potential energy (internal plus potential) and kinetic energy contributions. Total potential energy (about $3\times10^{9}\ \mathrm{J\,m^{-2}}$) is about 2000 times greater than kinetic energy (about $1.5\times10^{6}\ \mathrm{J\,m^{-2}}$). At first glance this suggests that the conservation question should focus primarily on the total potential energy component. Certainly the total potential energy is important for the global mean temperature, and hence for other basic features such as atmospheric water content and its consequences. However, correct handling of the kinetic energy budget is crucial for capturing the dynamics, including the strength of weather systems, midlatitude jets, and other features of meteorological interest, and cannot be ignored. In addition, the total potential energy can be divided into available and unavailable components. For any atmospheric state, a corresponding reference state can be defined by minimizing the potential energy under adiabatic air parcel rearrangements (i.e. rearrangements that conserve all $\mathscr{F}(\theta)$). The difference between the actual potential energy and the reference state potential energy is called the available potential energy, and is an upper bound on the potential energy that could be converted to kinetic energy by adiabatic dynamics. (In a more rigorous treatment additional constraints should be included, namely that PV must also be rearranged or mixed but not unmixed [16]; however, the calculation is then much more difficult and has not been done yet for realistic atmospheric states.) The unavailable part of the potential energy is typically 500 times greater than the available part (e.g. [51]). Moreover, the unavailable part is a function only of the $\mathscr{F}(\theta)$ and so is conserved separately from the total energy provided the $\mathscr{F}(\theta)$ are conserved.

The total energy throughput of the climate system is about 240 W m^{-2}, not counting reflected solar radiation, implying a timescale for total energy of about 150 days. Errors in total energy conservation significant compared to 240 W m^{-2} could lead to significant errors in global mean temperature.

The throughput of available energy (available potential plus kinetic) is of order 1.8 W m^{-2} [51], implying a timescale for available energy of about 20 days. GCM experiments [5] show a baroclinic eddy spin-down timescale of about 15–20 days after radiative forcing is switched off, consistent with this estimate. Errors in the available energy budget significant compared to 1.8 W m^{-2} could lead to significant errors in the strength of weather systems, midlatitude jets, etc.

To what extent do the arguments of Section 6 apply also to energy? In nearly all model formulations energy itself is not a model prognostic variable but is given by a nonlinear combination of prognostic variables. Therefore the total energy, like tracer variance, will have resolved and unresolved contributions. However, if temperature is a prognostic variable and a height-based vertical coordinate is used then only the relatively small kinetic component of the total energy has unresolved contributions. There can be a transfer of kinetic energy to small and unresolved scales, with frictional heating eventually reintroducing the energy as internal energy on resolved scales. The key question then is: "What fraction of the energy throughput is involved in a downscale cascade to unresolvable scales?" If the cascade is insignificant then attempts to conserve the resolved energy are justified. Many model development efforts have assumed that it is justified (e.g. [10,3,59,55]). If, however, the cascade is significant compared to the available energy budget then it might be appropriate to treat available energy in a similar way to potential enstrophy, with the dynamical core providing a sink to represent the cascade to unresolved scales. If the cascade to unresolved scales is significant compared to the total energy budget then the cascade rate should be diagnosed and fed back into the internal energy on resolved scales.

It will be argued in Section 8 that $\mathscr{F}(\theta)$, and hence the unavailable potential energy, have no significant downscale cascade. Hence there is a strong argument for attempting to conserve the unavailable potential energy in a numerical model, even if the available energy is not conserved.

Almost all of the available energy throughput is eventually dissipated by molecular viscosity after a transfer to small scales. Therefore the available energy is non-robust. However, by far the largest part of this dissipation, of order 1–2 W m^{-2}, occurs in boundary layer turbulence. The energy is transferred directly from the large-scale barotropic flow to the scale of the boundary layer eddies, i.e. the transfer is a class (i) process in the terminology of Section 1. Such a process must be parameterized in numerical models, and the models, in effect, see the dissipation occur on resolved scales rather than via a cascade. (In a large-eddy model of the boundary layer, however, this would be a class (ii) process.) In the free atmosphere there will be dissipation in three-dimensional turbulence generated by shear instability and gravity wave breaking, for example in fronts and elsewhere; again these are class (i) processes and should be parameterized in models (though they can only be parameterized if the need for the dissipation can be identified from resolved scale flow). Finally, there may be a contribution in the free atmosphere associated with a downscale cascade of energy by layerwise two-dimensional vortex interactions and nonlinear wave interactions. This final contribution should be handled by the dynamical core, either explicitly by a scale-selective dissipation term or implicitly by the numerics. Some estimates for its magnitude are given below.

Aircraft observations [49,13,14] have been used to construct kinetic energy spectra over wavelengths from about 1 km to 10,000 km. They show a spectral slope close to k^{-3} on synoptic scales with a smooth transition to a $k^{-5/3}$ spectrum on scales less than 100 km, where k is the horizontal wavenumber. The k^{-3} part of the spectrum has been interpreted as evidence for a potential enstrophy cascade (though it does not appear to be in an inertial range—see below). Various explanations for the $k^{-5/3}$ part of the spectrum have been put forward, including a downscale energy cascade associated with interacting gravity waves [19], an upscale energy cascade associated with layerwise two-dimensional turbulence forced by small-scale processes such as convective anvil outflows [40], and a downscale energy cascade associated with geostrophic turbulence [68]. Each of these theories relates the slope of the energy spectrum $E(k)$ to the spectral energy flux ε by a formula

$$E(k) = C\varepsilon^{2/3}k^{-5/3}, \tag{14}$$

where C is a constant of order unity (possibly different for the different theories). Fitting the aircraft data to (14) suggests a spectral energy flux of order $10^{-5}\ \mathrm{m^{-2}\ s^{-3}}$ (about 0.1 $\mathrm{W\ m^{-2}}$ in a vertical integral), though the amplitude of the spectrum does not determine the direction of the flux.

Global analyses can be used to diagnose not just the kinetic energy spectrum but also the spectral kinetic energy and enstrophy fluxes, and, as residuals, the spectral distribution of sources of kinetic energy and enstrophy [35,61]. They show a sink of kinetic energy on large scales ($n \leqslant 10$ where n is the global wavenumber, due to surface drag), a source of kinetic energy over a broad range of intermediate scales ($10 \leqslant n \leqslant 40$–50, due to baroclinic instability), and a sink of kinetic energy at small scales ($n \geqslant 40$–50). There is no range of scales over which the sources and sinks are negligible, implying that a true inertial range does not exist. The analyses also show a relatively large upscale kinetic energy flux (of order $5 \times 10^{-5}\ \mathrm{m^{-2}\ s^{-3}}$ or 0.5 $\mathrm{W\ m^{-2}}$) peaking around $n = 5$–10 and a smaller downscale kinetic energy flux (of order 0.5–$1 \times 10^{-5}\ \mathrm{m^{-2}\ s^{-3}}$ or 0.05–0.1 $\mathrm{W\ m^{-2}}$) at wavenumbers around $n = 40$–60. Note, however, that these analyses are limited by their finite resolution (T60 or T106), and the details of the energy spectrum and spectral fluxes at high wavenumbers are sensitive to the truncation of the data. Also, the analyses show no sign of the $k^{-5/3}$ part of the spectrum seen in aircraft data. Note also that these analyses examine only the rotational contribution to the kinetic energy and its flux. Very high resolution GCM simulations [36] suggest that the divergent and rotational contributions become comparable for n greater than about 100 (subject to the usual caveats about the ability of a GCM to avoid excessive generation or excessive damping of small-scale unbalanced motions).

The magnitude and direction of the spectral energy flux on the mesoscale can be estimated from aircraft data by computing third order velocity structure functions such as $\langle \delta\mathbf{u}\,\delta\mathbf{u}\,\delta\mathbf{u}\rangle$, where $\delta\mathbf{u}$ is the velocity difference over a displacement $\mathbf{r}$ [12,42]. Data in the lower stratosphere strongly suggest a downscale energy cascade on length scales between 10 and 100 km, of magnitude about $6 \times 10^{-5}\ \mathrm{m^{-2}\ s^{-3}}$ (about 0.6 $\mathrm{W\ m^{-2}}$ in a vertical integral if the same value is applicable throughout the troposphere). The tropospheric data are more difficult to interpret. Also, the data analysis does not clearly distinguish whether the flow on these scales is predominantly turbulent or wavelike. Other data, using aircraft [34] or Doppler radar [11], suggest that local energy dissipation rates in frontal regions can be two to three orders of magnitude greater than the above estimate.

In summary, there is some evidence for a downscale energy cascade across a wide range of scales shorter than a few hundred kilometres, of order $10^{-5}\ \mathrm{m^{-2}\ s^{-3}}$ (about 0.1 $\mathrm{W\ m^{-2}}$ in a vertical integral) or a few times this. This amounts to about 5–10% of the available energy throughput. In principle this argues against attempting to conserve exactly the resolved available energy. In practice, however, typical dissipation rates in climate models, due to explicit or inherent scale-selective dissipation are 1–2 $\mathrm{W\ m^{-2}}$ [69], which is an order of magnitude greater than the estimated downscale cascade in the real atmosphere. Moreover, it is not usual in most models to diagnose this energy sink and return the energy to the resolved scales in the form of heat. Thus there appears to be considerable scope to improve the handling of available energy in climate models.

8. Entropy

For adiabatic frictionless flow, there are infinitely many conservation laws of the form (1) with $A = \rho f(\theta)$, of which entropy $f(\theta) = \eta = C_p \ln\theta$ is just one. These imply an infinite set of global invariants $\int \rho f(\theta)\,\mathrm{d}\mathbf{x}$. An equivalent infinite set of invariants includes the mass per unit θ in each isentropic layer $\mathscr{F}(\theta) = \int \rho/|\nabla\theta|\mathrm{d}A$. Unlike energy, entropy (or any other $f(\theta)$) is transported only by advection and is therefore conserved separately in every air mass; this is why there are infinitely many entropy-related conserved quantities but only one conserved energy.

For adiabatic frictionless flow, entropy itself has no special status compared with any of the other possible $f(\theta)$ that might be considered. This can easily be seen by noticing that the mass per unit $f(\theta)$ in any isentropic layer is just a constant times the mass per unit θ in the same layer

$$\int \frac{\rho}{|\nabla f(\theta)|}\,\mathrm{d}A = \frac{1}{f'(\theta)} \int \frac{\rho}{|\nabla \theta|}\,\mathrm{d}A, \tag{15}$$

so that any $f(\theta)$ is as good as any other. However, when diabatic processes, including mixing, are introduced then entropy does have special status—see below.

In large parts of the atmosphere vertical mixing is weak. This is because the free atmosphere is strongly stably stratified and, moreover, θ itself (or any $f(\theta)$) determines the stratification. This, in turn, inhibits vertical mixing between isentropic layers (much more strongly than PV gradients inhibit horizontal mixing). In these regions the $\mathscr{F}(\theta)$ are robust invariants. In other regions, small-scale mixing by the boundary layer, shear instability, gravity wave breaking, etc., can modify the $\mathscr{F}(\theta)$ even in the adiabatic frictionless limit. However, these mixing processes are class (i), in the terminology of Section 1, involving very small three-dimensional eddies; the $\mathscr{F}(\theta)$ do not cascade from large to small scales. This argues for attempting to conserve $\mathscr{F}(\theta)$ in dynamical cores.

There is a small source for $\mathscr{F}(\theta)$, which would still exist in the adiabatic frictionless limit, associated with the small downscale cascade of available energy (Section 7), of order $10^{-5}\ \mathrm{m}^{-2}\ \mathrm{s}^{-3}$, and its ultimate dissipation and conversion to heat. However, the implied θ tendency is of the order $10^{-3}\ \mathrm{K\ day}^{-1}$, which is negligibly small compared to realistic diabatic heating.

Conservation of entropy-related quantities is physically important in at least two ways. First, the distribution of mass as a function of θ determines the unavailable component of the total potential energy, as discussed in Section 7. The unavailable part of the potential energy is typically 500 times greater than the available part (e.g. [51]). Moreover, the unavailable part is a function only of the $\mathscr{F}(\theta)$ and so is conserved separately from the available energy, and is not involved in a downscale cascade, as argued above. Hence there is a strong argument for attempting to conserve the unavailable potential energy in a numerical model, even if the available energy is not conserved.

Second, any failure to conserve entropy in a numerical model is likely to be in the form of a systematic spurious entropy production, because of the tendency of numerical schemes to mix rather than unmix [32,22], though numerical schemes can unmix too [64,73]. It has been suggested [32] that, in order to balance both the energy and entropy budgets in the presence of spurious entropy sources, a numerical climate model will adjust to a state that has a systematic cold bias, particularly in the lower tropical troposphere and near the high-latitude tropopause.

Some recent work [73] found that in an "adiabatic" baroclinic wave life cycle, simulated using a standard sigma-coordinate spectral model, there were large local spurious sources and sinks of entropy with a small positive residual of around $0.4\ \mathrm{mW\ m}^{-2}\ \mathrm{K}^{-1}$. They also found that the net numerical generation of entropy depended only weakly on model resolution and on the timescale specified for scale-selective dissipation. A plausible interpretation is that to a large extent the entropy production is inevitable and is driven by the large-scale flow, and occurs when modelled features collapse to the model's dissipation scale, whatever that scale is. In reality the dissipation and entropy production mechanisms would be different and would occur on smaller scales, but the overall entropy production might be similar.

The global mean irreversible entropy sources in the real atmosphere [50,25] are estimated to be of order $20\ \mathrm{mW\ m}^{-2}\ \mathrm{K}^{-1}$ for moist processes and of order 5–10 $\mathrm{mW\ m}^{-2}\ \mathrm{K}^{-1}$ for turbulent and dissipative processes. (It is difficult to re-express these values as timescales because it is not obvious what to take as the reference value for entropy.) However, in the absence of sources and sinks entropy is conserved separately in every air mass, so it is relevant to discuss local sources. These can be very much larger, or very much smaller, than the global mean values. Conservation of entropy, and spurious numerical sources of entropy, are likely to be particularly important where the true source is very small, e.g. in the tropical lower stratosphere where the radiative timescale is very long and vertical mixing is very weak.

Finally, although entropy has no special status compared to other $f(\theta)$ for purely adiabatic flow, it does have special status when diabatic processes, including mixing, act. First, the relation

$$T\,\mathrm{d}\eta = \mathrm{d}q, \tag{16}$$

where q is diabatic heating and η is specific entropy, is simpler than any analogous relation involving $f(\theta)$ rather than η. More importantly, the entropy statement of the second law of thermodynamics (e.g. [18]) states that the total entropy of a thermally isolated system can never decrease. For example, mixing processes in the atmosphere, or radiative exchange between different parts of the atmosphere at different temperature, will always increase the total entropy. There is no similar expression of the second law of thermodynamics in terms of other $f(\theta)$ except those trivially proportional to the specific entropy.

9. Potential vorticity, and Lagrangian conservation

Because of its dynamically important Lagrangian conservation and invertibility properties, there are advantages to diagnosing atmospheric flow in terms of PV (e.g. [31]). It is reasonable to expect improved model performance from an accurate numerical treatment of PV (e.g. [7]).

In the literature most emphasis is on the Lagrangian conservation property (3) rather than the flux-form (1). Accurate Lagrangian conservation of PV is crucial for correct simulation of the strength of weather systems such as fronts and cyclones. Even after PV features have been strained out into thin sheets and streamers they can still be meteorologically active. Spurious generation of PV of the wrong sign could lead to spurious instabilities.

The timescale for physical processes to change an air parcel's PV is typically a few days in the free atmosphere. Spurious numerical modification of PV needs to be slower than this to avoid damaging numerical solutions. The simulation of a baroclinic wave life cycle has been studied [72] in a sigma-coordinate spectral model with a standard vorticity-divergence formulation. A significant spurious amplification of PV maxima was found in the upper troposphere, accompanied by a significant spurious increase in total potential enstrophy on the relevant isentropic level. Such spurious amplification was absent in an isentropic-coordinate model that directly predicted PV using a non-oscillatory advection scheme.

Accurate numerical treatment of some variable is most easily achieved by making that variable one of the predicted variables, though there have been few attempts to do this in three-dimensional atmospheric models with PV [27,39,48] except in balanced models. On the other hand, improved treatment of PV can be obtained by careful treatment of the other variables in formulations based on wind components [3,41].

Accurate Lagrangian conservation of PV and other variables is obviously possible with Lagrangian numerical schemes such as contour advection (e.g. [20]). In Eulerian models Lagrangian conservation cannot be captured exactly: there is unavoidable information loss as the fluid moves relative to the finite resolution grid. However, certain properties implied by Lagrangian conservation, such as positivity and local boundedness under advection, can be enforced, even while maintaining flux-form conservation. Lagrangian conservation of long-lived tracers can be improved by the use of a Lagrangian or isentropic vertical coordinate [74,43]. With Eulerian advection schemes Lagrangian conservation tends to improve with increasing order of accuracy of the scheme (taking into account any explicit scale-selective dissipation needed with even order and spectral advection schemes). Third order advection seems to be the minimum acceptable for weather and climate modelling. Even then the inherent dissipation may be noticeable (e.g. [72]; a similar experience was found with the cubic semi-Lagrangian scheme of the current operational version of the Met Office Unified Model UM5). Where the Lagrangian timescale is very long, as in the stratospheric "tape recorder", a fifth or even seventh order advection scheme might be needed for accurate simulation [28].

Certain thermodynamic quantities, such as equivalent potential temperature, are materially conserved, or nearly conserved, even when phase changes of water take place. This Lagrangian conservation can be poorly simulated in a numerical model if the conserved quantity is diagnosed from other prognostic variables rather than directly predicted itself, especially if the other variables are not treated consistently—the example of equivalent potential temperature has been discussed in depth [33]. Capturing the Lagrangian conservation property in a numerical model can be improved by directly predicting the conserved variable in place of temperature or potential temperature (e.g. [66]). However, retrieving the temperature from the predicted variables then requires iterative solution of a transcendental equation, adding to the computational cost. Also, in practice, approximate versions of the conserved variables are usually used, and care must be taken to ensure that the approximation is sufficiently accurate (e.g. [52,9]). This type of approach is commonly used in cloud modelling, but has received little attention in the context of weather prediction and climate modelling.

A further complication is that, even when predicting a materially conserved quantity and using a non-oscillatory scheme for advection, interaction between advection and condensation can lead to spurious oscillations in temperature and moisture fields. The problem can be reduced by using specially formulated ‘‘limiters’’ that account for the interaction between advection and condensation [26].

10. Which combinations of quantities can be conserved?

Given that a numerical scheme cannot possibly conserve all of the quantities conserved by the continuous governing equations, an important question is ‘‘What combinations of quantities could a numerical model conserve?’’ There seems to be no theory to provide a general answer to this question, but some comments are offered in the following.

A straightforward way to ensure the conservation of some quantity is to make it a prognostic variable in a flux-form finite volume discretization. However, atmospheric models use only a small number of prognostic variables (though different choices are possible, for example temperature versus entropy, or momentum versus vorticity and divergence). Any additional conservation properties must be obtained through some special mathematical property of the discretization. For example, spatial discretizations can be found that conserve (global integrals of) mass, energy and potential enstrophy as well as preserving an initial constant potential vorticity [3]. Local mass conservation of long-lived tracers can be obtained via a flux-form finite volume discretization and at the same time boundedness of the tracer (implied by the Lagrangian conservation law) can be guaranteed by carefully ‘‘limiting’’ the fluxes. A mass-conserving isentropic-coordinate model could conserve discrete analogues of $\mathscr{F}(\theta)$ independently in each model layer; these in turn would imply conservation of a discrete approximation to the unavailable part of the potential energy.

As noted in Section 2, for the continuous equations conservation properties can be related to continuous symmetry properties of the Hamiltonian. Unfortunately, an Eulerian discretization destroys the continuous symmetry, so there does not appear to be a way to exploit symmetry properties to develop conservative discretizations [57]. However, conservation properties can also be derived via a Hamiltonian Poisson-bracket approach, and this approach can be used to obtain a systematic method of deriving conservative spatial discretizations [57]. There is certainly scope for further exploration of this approach.

For Hamiltonian systems another interesting approach to obtaining desirable conservation properties is through the use of symplectic integration schemes. Such schemes are designed to preserve the symplectic structure in phase space, which provides strong constraints on the behaviour of the system. Symplectic schemes for ordinary differential equations can be shown to give exponentially accurate conservation of energy over exponentially long times [53]. The ideas can be extended to multi-symplectic schemes for partial differential equations. For linear Hamiltonian systems of a certain form, a multi-symplectic box scheme has been described [8] that gives exact local conservation of energy and momentum. The authors claim that conservation is also excellent (though not exact) for nonlinear systems.

Finally, it might be possible to obtain a greater number of conservation properties by moving to a fully Lagrangian, rather than Eulerian, formulation. For example, a particle-mesh method for the shallow water equations has been described [23,24]. By combining a Hamiltonian spatial discretization with a symplectic time stepping algorithm the authors claim exact conservation of mass and excellent conservation of energy and potential vorticity. If it is possible to overcome the traditional difficulties of fully Lagrangian methods, associated for example with spatial irregularity and inhomogeneity of the particle distribution, then this approach might prove attractive.

11. Conclusions

Conservation issues for atmospheric model dynamical cores should be addressed in the context of realistic atmospheric flows. The relative importance of different conservation properties depends on the roles they play in realistic circulations, and on the timescales for their corresponding physical sources and sinks. For this reason, results of idealized test cases should be interpreted with care.

Table 1
Approximate source timescales for some conservable quantities, and indications of whether they are robust and whether they cascade to small scales

Quantity	Robust	Cascade	Approx. timescale
Mass	Yes		Infinite
Momentum			Minutes to hours
Angular momentum			10 days (locally longer)
Potential enstrophy		Yes	10 days
Tracer variance		Yes	10 days
Unavailable energy	Almost		150 days
Available energy		Yes (5–10%)	20 days
Entropy	Almost		Variable

We caution against dogmatic attempts to obtain exact conservation of everything possible. Some conservation properties of the original equation set might no longer hold for the averaged equations that must be integrated in practice. Also, for non-robust invariants that cascade downscale, numerical models must allow for a transfer from resolved to unresolved scales, and conservation of only the resolved contribution might not be appropriate. We do not insist that every non-conservative physical process be handled by a specific corresponding physical parameterization. For example, the dissipation (or, rather, resolved to unresolved scale transfer) of tracer variance can be handled adequately by a suitably formulated dynamical core.

Table 1 summarizes some of the discussion of the preceding sections. There is a very strong argument for attempting to conserve mass; it is a truly robust invariant with, for practical purposes, an infinite source timescale. Strong arguments can also be made for attempting to conserve angular momentum, unavailable energy, and $\mathscr{F}(\theta)$; they have moderately long source timescales, at least locally, and, although not truly robust, are not dissipated via downscale class (ii) cascades. Potential enstrophy and tracer variance, on the other hand, have significant sinks via downscale cascades, and so, as noted above, a dynamical core must allow for the transfer from resolved to unresolved scales.

Currently one of the most challenging conservation issues appears to be the handling of the available energy budget. Around 5–10% of the available energy throughput cascades downscale (Section 7), and, as argued above, could in principle be handled by the dynamical core. However, typical dissipation rates in climate models, due to explicit or inherent scale-selective dissipation are an order of magnitude greater [69]. A fundamental difficulty appears to be that the types of scale-selective dissipation most commonly used, whether explicitly of the form ∇^{2n} or inherent in the numerical schemes (e.g. [45,62]), are essentially linear, with the damping at any scale proportional to the amplitude of the damped quantity at that scale; then the ratio of energy dissipation to enstrophy dissipation must scale roughly like the square of the grid length, if the dissipation is dominated by near-grid scales. For current climate resolutions that ratio is of order $10^{-4}\,\mathrm{m}^2\,\mathrm{s}^{-3}/10^{-15}\,\mathrm{s}^{-3} = 10^{11}\,\mathrm{m}^2$, which is an order of magnitude too large. Higher order dissipation operators (e.g. [37]) can help ensure that the dissipation is indeed dominated by near-grid scales, but cannot overcome the scaling law. The scaling argument implies that the problem should be reduced as finer resolutions become affordable. Until then, it might be desirable to consider more sophisticated schemes that reduce the ratio of energy dissipation to enstrophy dissipation by nonlinearly coupling different scales [55], or reintroducing some energy on resolved scales through some representation of the relevant spectral interactions or ‘‘backscatter’’ [35,58].

Acknowledgements

The author gratefully acknowledges discussions with Mike Cullen, Nigel Wood and Tim Woollings, which helped to clarify some of the ideas presented here.

References

[1] D.G. Andrews, M.E. McIntyre, On wave action and its relatives, J. Fluid Mech. 89 (1978) 647–664.
[2] A. Arakawa, A personal perspective on the early years of general circulation modeling at UCLA, in: D.A. Randall (Ed.), General Circulation Model Development, Past, Present, and Future, Academic Press, 2000, pp. 1–65.

[3] A. Arakawa, V.R. Lamb, A potential entrophy and energy conserving scheme for the shallow water equations, Mon. Weather Rev. 109 (1981) 18–36.
[4] P.R. Bannon, Hydrostatic adjustment: Lamb's problem, J. Atmos. Sci. 52 (1995) 1743–1752.
[5] L. Barry, G.C. Craig, J. Thuburn, A GCM investigation into the nature of baroclinic adjustment, J. Atmos. Sci. 57 (2000) 1141–1155.
[6] J.R. Bates, A dynamical stabilizer in the climate system: a mechanism suggested by a simple model, Tellus 51A (1999) 349–372.
[7] J.R. Bates, Y. Li, A. Brandt, S.F. McCormick, J. Ruge, A global shallow-water numerical model based on the semi-Lagrangian advection of potential vorticity, Quart. J. Roy. Meteorol. Soc. 121 (1995) 1981–2005.
[8] T.J. Bridges, S. Reich, Multi-symplectic integrators: numerical schemes for Hamiltonian PDEs that conserve simplecticity, Phys. Lett. A. 284 (2001) 184–193.
[9] G.H. Bryan, J.M. Fritsch, A reevaluation of ice–liquid water potential temperature, Mon. Weather Rev. 132 (2004) 2421–2431.
[10] D.M. Burridge, J. Haseler, A model for medium range weather forecasting—Adiabatic formulation. ECMWF Tech. Rep. 4, Reading, United Kingdom, 1977, 46pp.
[11] D. Chapman, K.A. Browning, Measurements of dissipation rate in frontal zones, Quart. J. Roy. Meteorol. Soc. 127 (2001) 1939–1959.
[12] J.-Y. Cho, E. Lindborg, Horizontal velocity structure functions in the upper troposphere and lower stratosphere 1. Observations, J. Geophys. Res. 106 (2001) 10,223–10,232.
[13] J.-Y. Cho et al., Horizontal wavenumber spectra of winds, temperature and trace gases during the Pacific Exploratory Missions. Part I: Climatology, J. Geophys. Res. 104 (1999) 5697–5716.
[14] J.-Y. Cho, R.E. Newell, J.D. Barrick, Horizontal wavenumber spectra of winds, temperature and trace gases during the Pacific Exploratory Missions. Part II: Gravity waves, quasi-two-dimensional turbulence, and vortical modes, J. Geophys. Res. 104 (1999) 16,297–16,308.
[15] M.J.P. Cullen, Modelling atmospheric and oceanic flows. Acta Numerica, in preparation.
[16] M.J.P. Cullen, R.J. Douglas, Large-amplitude nonlinear stability results for atmospheric circulations, Quart. J. Roy. Meteorol. Soc. 129 (2003) 1969–1988.
[17] M.J.P. Cullen, D. Salmond, P.K. Smolarkiewicz, Key numerical issues for the future development of the ECMWF model. in: Proceedings of the ECMWF Workshop on Developments in Numerical Methods for Very High Resolution Global Models, June, 2000, pp. 183–206.
[18] J.A. Curry, P.J. Webster, Thermodynamics of atmospheres and oceans, Academic Press, 1999, 471pp.
[19] E.M. Dewan, Stratospheric wave spectra resembling turbulence, Science 204 (1979) 832–835.
[20] D.G. Dritschel, M.H.P. Ambaum, A contour-advective semi-Lagrangian numerical algorithm for simulating fine-scale conservative dynamical fields, Quart. J. Roy. Meteorol. Soc. 123 (1997) 1097–1130.
[21] D.R. Durran, Numerical Methods for Wave Equations in Geophysical Fluid Dynamics, Springer, 1999, 465pp.
[22] J. Egger, Numerical generation of entropies, Mon. Weather Rev. 127 (1999) 2211–2216.
[23] J. Frank, S. Reich, Conservation properties of smoothed particle hydrodynamics applied to the shallow water equations, BIT 43 (2003) 40–54.
[24] J. Frank, S. Reich, The Hamiltonian particle-mesh method for the spherical shallow water equations, Atmos. Sci. Lett. 5 (2004) 89–95.
[25] R. Goody, Sources and sinks of climate entropy, Quart. J. Roy. Meteorol. Soc. 126 (2000) 1953–1970.
[26] W.W. Grabowski, P.K. Smolarkiewicz, Monotone finite-difference approximations to the advection–condensation problem, Mon. Weather Rev. 118 (1990) 2082–2097.
[27] A.R. Gregory, Numerical simulations of winter stratosphere dynamics, Ph.D. Thesis, University of Reading, 1999.
[28] A.R. Gregory, V. West, The sensitivity of a model's stratospheric tape recorder to the choice of advection scheme, Quart. J. Roy. Meteorol. Soc. 128 (2002) 1827–1846.
[29] P.H. Haynes, Forced, dissipative generalizations of finite-amplitude wave activity conservation relations for zonal and non-zonal basic flows, J. Atmos. Sci. 45 (1988) 2352–2362.
[30] P.H. Haynes, J. Anglade, The vertical-scale cascade in atmospheric tracers due to large-scale differential advection, J. Atmos. Sci. 54 (1997) 1121–1136.
[31] B.J. Hoskins, M.E. McIntyre, A.W. Robinson, On the use and significance of isentropic potential vorticity maps, Quart. J. Roy. Meteorol. Soc. 111 (1985) 877–946.
[32] D.R. Johnson, "General coldness of climate models" and the second law: Implications for modeling the Earth system, J. Clim. 10 (1997) 2826–2846.
[33] D.R. Johnson, A.J. Lenzen, T.H. Zapotocny, T.K. Schaak, Numerical uncertainties in the simulation of reversible isentropic processes and entropy conservation, J. Clim. 13 (2000) 3860–3884.
[34] P.J. Kennedy, M.A. Shapiro, The energy budget in a clear air turbulence zone as observed by aircraft, Mon. Weather Rev. 103 (1975) 650–654.
[35] J.N. Koshyk, G.J. Boer, Parametrization of dynamical subgrid-scale processes in a spectral GCM, J. Atmos. Sci. 52 (1995) 965–976.
[36] J.N. Koshyk, K. Hamilton, J.D. Mahlman, Simulation of the $k^{-5/3}$ mesoscale spectral regime in the GFDL SKYHI general circulation model, Geophys. Res. Lett. 26 (1999) 843–846.
[37] L. Laursen, E. Eliasen, On the effects of the damping mechanisms in an atmospheric general circulation model, Tellus 41A (1989) 385–400.
[38] R.J. LeVeque, Numerical Methods for Conservation Laws, Birkhauser, 1992, 214pp.

[39] Y. Li, J. Ruge, J.R. Bates, A. Brandt, A proposed adiabatic formulation of 3-dimensional global atmospheric models based on potential vorticity, Tellus 52 (2000) 129–139.
[40] D.K. Lilly, Stratified turbulence and the mesoscale variability of the atmosphere, J. Atmos. Sci. 40 (1983) 749–761.
[41] S.-J. Lin, R.B. Rood, An explicit flux-form semi-Lagrangian shallow-water model on the sphere, Quart. J. Roy. Meteorol. Soc. 123 (1997) 2477–2498.
[42] E. Lindborg, J.-Y. Cho, Horizontal velocity structure functions in the upper troposphere and lower stratosphere. 2. Theoretical considerations, J. Geophys. Res. 106 (2001) 10,233–10,241.
[43] N.M. Mahowald, R.A. Plumb, P.J. Rasch, J. del Corral, F. Sassi, W. Heres, Stratospheric transport in a three-dimensional isentropic coordinate model, J. Geophys. Res. 107 (2002). Art. no. 4254.
[44] P.J. Mason, A.R. Brown, On subgrid models and filter operations in large eddy simulations, J. Atmos. Sci. 56 (1999) 2101–2114.
[45] J.D. McCalpin, A quantitative analysis of the dissipation inherent in semi-Lagrangian advection, Mon. Weather Rev. 116 (1988) 2330–2336.
[46] M.E. McIntyre, W.A. Norton, Potential vorticity inversion on a hemisphere, J. Atmos. Sci. 57 (2000) 1214–1235.
[47] M.E. McIntyre, T.G. Shepherd, An exact local conservation theorem for finite-amplitude disturbances to non-parallel shear flows, with remarks on Hamiltonian structure and on Arnold's stability theorems, J. Fluid Mech. 181 (1987) 527–565.
[48] A.R. Mohebalhojeh, D.G. Dritschel, The contour-advective semi-Lagrangian algorithms for many-layer primitive-equation models, Quart. J. Roy. Meteorol. Soc. 130 (2004) 347–364.
[49] G.D. Nastrom, K.S. Gage, A climatology of atmospheric wavenumber spectra observed by commercial aircraft, J. Atmos. Sci. 42 (1985) 950–960.
[50] J.P. Peixoto, A.H. Oort, M. de Almeida, A. Tomé, Entropy budget of the atmosphere, J. Geophys. Res. 96 (1991) 10,981–10,988.
[51] J.P. Peixoto, A.H. Oort, Physics of Climate, American Institute of Physics, 1992.
[52] Y. Pointin, Wet equivalent potential temperature and enthalpy as prognostic variables in cloud modeling, J. Atmos. Sci. 41 (1984) 651–660.
[53] S. Reich, Backward error analysis for numerical integrators, SIAM J. Numer. Anal. 36 (1999) 1549–1570.
[54] R. Sadourny, The dynamics of finite-difference models of the shallow-water equations, J. Atmos. Sci. 32 (1974) 680–689.
[55] R. Sadourny, C. Basdevant, Parameterization of subgrid scale barotropic and baroclinic eddies in quasi-geostrophic models: anticipated potential vorticity method, J. Atmos. Sci. 42 (1985) 1353–1363.
[56] R. Salmon, Lectures on Geophysical Fluid Dynamics, Oxford University Press, 1998, 378pp.
[57] R. Salmon, Poisson-bracket approach to the construction of energy- and potential-enstrophy-conserving algorithms for the shallow-water equations, J. Atmos. Sci. 61 (2004) 2016–2036.
[58] G. Shutts, A kinetic energy backscatter algorithm for use in ensemble prediction systems, Quart. J. Roy. Meteorol. Soc. 131 (2005) 3079–3102.
[59] A.J. Simmons, D.M. Burridge, An energy and angular-momentum conserving vertical finite-difference scheme and hybrid vertical coordinates, Mon. Weather Rev. 109 (1981) 758–766.
[60] K.S. Smith, Comments on "the k^{-3} and $k^{-5/3}$ energy spectrum of atmospheric turbulence: quasigeostrophic two-level model simulation", J. Atmos. Sci. 61 (2004) 937–942.
[61] D.M. Straus, P. Ditlevsen, Two-dimensional turbulence properties of the ECMWF reanalyses, Tellus 51A (1999) 749–772.
[62] J. Thuburn, Dissipation and cascades to small scales in numerical models using a shape-preserving advection scheme, Mon. Weather Rev. 123 (1995) 1888–1903.
[63] J. Thuburn, V. Lagneau, Eulerian mean, contour integral, and finite-amplitude wave activity diagnostics applied to a single layer model of the winter stratosphere, J. Atmos. Sci. 56 (1999) 689–710.
[64] J. Thuburn, M.E. McIntyre, Numerical advection schemes, cross-isentropic random walks, and correlations between chemical species, J. Geophys. Res. 102 (1997) 6775–6797.
[65] J. Thuburn, D.G.-H. Tan, A parameterization of mixdown time for atmospheric chemicals, J. Geophys. Res. 102 (1997) 13,037–13,049.
[66] G.J. Tripoli, W.R. Cotton, The use of ice–liquid water potential temperature as a thermodynamic variable in deep atmosphere models, Mon. Weather Rev. 109 (1981) 1094–1102.
[67] K.K. Tung, Reply to comments on "the k^{-3} and $k^{-5/3}$ energy spectrum of atmospheric turbulence: quasigeostrophic two-level model simulation", J. Atmos. Sci. 61 (2004) 943–948.
[68] K.K. Tung, W.T. Welch Orlando, The k^{-3} and $k^{-5/3}$ energy spectrum of atmospheric turbulence: quasigeostrophic two-level model simulation, J. Atmos. Sci. 60 (2003) 824–835.
[69] WGNE, WMO Atmospheric Research and Environment Programme, CAS/JSC Working Group on Numerical Experimentation, Report No. 18, 2003.
[70] A.A. White, A view of the equations of meteorological dynamics and various approximations, in: J. Norbury, I. Roulstone (Eds.), Large-Scale Atmosphere–Ocean Dynamics I: Analytical Methods and Numerical Models, Cambridge University Press, 2002, pp. 1–100.
[71] D.L. Williamson, J.B. Drake, J.J. Hack, R. Jakob, P.N. Swarztrauber, A standard test set for numerical approximations to the shallow water equations in spherical geometry, J. Comput. Phys. 102 (1992) 211–224.
[72] T.J. Woollings, Entropy and potential vorticity in atmospheric dynamical core models, PhD Thesis, University of Reading, 2004.
[73] T.J. Woollings, J. Thuburn, Entropy sources in a dynamical core atmospheric model, Quart. J. Roy. Meteorol. Soc. 132 (2006) 43–59.
[74] T.H. Zapotocny, A.J. Lenzen, D.R. Johnson, F.M. Reames, T.K. Schaak, A comparison of inert trace constituent transport between the University of Wisconsin isentropic-sigma model and the NCAR community climate model, Mon. Weather Rev. 125 (1997) 120–142.