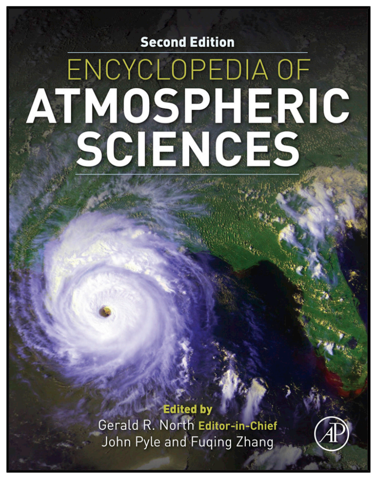

From McIntyre, M.E., 2015. Balanced Flow. In: Gerald R. North (editor-in-chief), John Pyle and Fuqing Zhang (editors). Encyclopedia of Atmospheric Sciences, 2nd edition, Vol 2, pp. 298–303.
ISBN: 9780123822253

Academic Press

Balanced Flow

ME McIntyre, University of Cambridge, Cambridge, UK

Synopsis

A balanced flow is one in which the three-dimensional velocity field is functionally related to the mass field, presumed hydrostatically related to the pressure field. Such a functional relation between the velocity and mass fields is called a balance relation, or filtering condition. The simplest but least accurate such relation is geostrophic balance. There are more accurate balance relations, of which the most accurate are fully nonlocal. That is, the velocity at a point depends on the mass field throughout the domain. There are ultimate limitations to accuracy, governed by the fuzziness of the slow quasimanifold.

Introduction

The concept of *balanced flow* is the counterpart, in atmosphere–ocean dynamics, to the well-known concept of nearly incompressible flow in classical aerodynamics. In aerodynamics, a key aspect of such flow – long recognized as central to understanding the behavior of subsonic aircraft – is that all the significant dynamical information is contained in the vorticity field. To that extent the flow has, in effect, fewer degrees of freedom than a fully general flow. We may think of it as being elastostatically balanced, in the sense that freely propagating sound waves can be neglected in the dynamics.

In atmosphere–ocean dynamics there is a corresponding statement with vorticity replaced by potential vorticity (PV), understood in a suitably generalized sense; see generalized PV field in the article **Dynamical Meteorology:** Potential Vorticity. For many cases of rotating, stably stratified fluid flow, with parameter values typical of the atmosphere and oceans, all the significant dynamical information is contained in the generalized PV field. One may invert this field at each instant to obtain the mass and velocity fields. The article on Potential Vorticity gives a more precise statement. All such flows may be characterized as balanced.

Again this means that the flow has, in effect, fewer degrees of freedom than a fully general flow. More precisely, balance and invertibility mean that not only sound waves but also freely propagating inertia–gravity waves can be neglected in, or filtered from, the dynamics. Thus balanced flows can be much simpler to understand than fully general flows, thanks to the relatively simple way in which the advective nonlinearity acts on the PV.

Cases of fluid flow describable as balanced come under headings such as Rossby waves, Rossby-wave breaking, vortex dynamics, vortical modes, vortical flow, vortex coherence, vortex resilience, eddy-transport barriers, blocking, cyclogenesis, baroclinic instability and barotropic instability (meaning the wavy shear instabilities), all of which are related to the fundamental Rossby-wave restoring mechanism or quasi-elasticity that exists whenever there are isentropic gradients of PV in the interior of the flow domain, or gradients of potential temperature on an upper or lower boundary. The concept of balanced flow is fundamental, also, to theories of wave–mean interaction and wave–vortex interaction, needed in order to understand, for instance, the gyroscopic pumping that drives global-scale stratospheric circulations and chemical transports (§6 of **Dynamical Meteorology:** Potential Vorticity). In these theories the mean or vortical flow is usually considered to be balanced, regardless of the wave types involved. Indeed, the concept of balance enters, implicitly or explicitly, into almost any discussion of meteorologically interesting fluid phenomena; and balance versus imbalance is part of the conceptual foundation that underpins data analysis, data assimilation, and weather prediction.

The Elastic Pendulum

Balance has counterparts not only in aerodynamics but also in simple mechanical systems such as the elastic pendulum. This is a massive bob suspended from a pivot by a stiff elastic spring of negligible mass. Such a pendulum has slow, swinging modes of oscillation in which the relatively fast, compressional modes of the bob and spring are hardly excited: they can be neglected in the dynamics if the spring is stiff enough. The slow, swinging modes correspond to balanced flow, and the fast, compressional modes to sound and inertia–gravity waves. One may describe the swinging modes to a crude first approximation by making the spring strictly incompressible, i.e., by making its length strictly constant. There is a hierarchy of more accurate approximations that allow the spring to change its length in a quasi-static or elastostatic way, the spring being longest when the bob moves fastest and shortest when the bob is stationary.

In such a quasi-static description the length of the spring is functionally related to the speed of the bob. The functional relation holds at each instant t, i.e., it holds diagnostically. No derivatives or integrals with respect to t are involved, and values of t do not explicitly enter into the definition of the functional relation. The property of being diagnostic, in this sense, provides us with a useful mathematical and conceptual simplification.

Such approximations and their ultimate limitations can be studied mathematically via techniques ranging all the way from two-timing formalisms (method of multiple scales) and bounded-derivative theory to KAM (Kolmogorov–Arnol'd–Moser) theory and other dynamical-systems techniques; there is an enormous literature.

The error incurred in using the most accurate quasi-static descriptions becomes exponentially small as the fast–slow timescale separation increases. It may even be zero, or in some

 http://dx.doi.org/10.1016/B978-0-12-382225-3.00484-9

Encyclopedia of Atmospheric Sciences, Second Edition, 2015, 298–303

circumstances small but inherently nonzero (corresponding to KAM tori breaking into thin chaotic layers, also called fractal layers or stochastic layers).

Balance Relations

In atmosphere–ocean dynamics the defining property of balance is that an analogous functional relation holds – diagnostic in precisely the same sense. The functional relation between bob speed and spring length is replaced by a functional relation between the fluid's velocity and mass fields. More precisely, a flow is said to be balanced if the three-dimensional velocity field $\mathbf{u}(\mathbf{x},t)$ satisfies a functional relation of the form $\mathbf{u}(\mathbf{x},t) = \mathbf{u}^{\mathrm{B}}$ where $\mathbf{u}^{\mathrm{B}}$ depends only on the mass field or mass configuration, i.e., on the spatial distribution of mass throughout the fluid system, presumed to be hydrostatically related to the pressure field. (Knowledge of the mass field then implies knowledge of the pressure, temperature and potential temperature fields, given zero pressure at the top of the atmosphere.)

Such a functional relation $\mathbf{u}(\mathbf{x},t) = \mathbf{u}^{\mathrm{B}}$ between the velocity and mass fields is called a filtering or balance condition, or balance relation. It supplies just enough information to make the PV field invertible. The property of being diagnostic means that if one knows the mass field at some instant t, but knows nothing about its time dependence, nor the value of t itself, then the balance relation must nevertheless allow one to deduce the complete three-dimensional velocity field $\mathbf{u}$. It must allow the velocity field to be deduced from the mass field and from the mass field alone. (Such a diagnostic relation should not, however, be mistaken for a causal relation. To think that the mass or pressure field causes the velocity field is like thinking that the spring length causes the pendulum's motion.)

To the extent that a balance relation holds it excludes, or filters, freely propagating sound waves and inertia–gravity waves from the repertoire of possible fluid motions. The system then has too few degrees of freedom to describe such waves. The reduction in degrees of freedom is sometimes expressed by saying that some degrees of freedom are slaved to others, or that the possible states of the dynamical system have been confined to a so-called slow manifold within phase space, having lower dimensionality than the full phase space in which it is embedded. In this language we say that, in particular, the velocity field is slaved to the mass field. A more careful statement would be that in the actual flow the velocity field evolves as if it were slaved to the mass field, to some useful approximation at least. This is like saying that the swinging motion of the pendulum evolves as if the bob speed were slaved to the spring length, to some useful approximation, even though there is no actual mechanical linkage between the two variables.

A standard example of a balance relation is the so-called *geostrophic relation*, which is simple to write and, for typical extratropical parameter values, qualitatively useful though quantitatively not very accurate:

$$\mathbf{u}(\mathbf{x},t) = \frac{1}{f}\left(-\frac{\partial\Phi(\mathbf{x},t)}{\partial y}, \frac{\partial\Phi(\mathbf{x},t)}{\partial x}, 0\right) \quad [1]$$

Here f is the Coriolis parameter, $\Phi(\mathbf{x},t)$ is the geopotential height (approximately geometric altitude times gravitational acceleration), and position $\mathbf{x}$ is specified using pressure altitude along with horizontal position x, y. Thus the horizontal spatial derivatives $\partial/\partial x$ and $\partial/\partial y$ are taken at constant pressure altitude rather than at constant geometric altitude. This qualifies as a balance relation because of the presumption that the hydrostatic relation also holds, as normally assumed when using pressure as the vertical coordinate. Knowing Φ on each constant-pressure (isobaric) surface is then equivalent to knowing the mass field. So eqn [1] is, as required, a diagnostic functional relation between the velocity field and the mass field. The vertical derivative of eqn [1] is the so-called thermal wind equation.

The horizontal coordinates x, y are orthogonal coordinates, and can be taken either as local curvilinear following the Earth's geometry, or as local Cartesian in a tangent-plane approximation. If we also take $f =$ constant, giving us the so-called f-plane approximation, then eqn [1] asserts not only that $\mathbf{u}$ is slaved to the mass field but also that it is two-dimensionally incompressible or nondivergent, with streamfunction $\Psi = \Phi/f$, so that

$$\mathbf{u}(\mathbf{x},t) = \left(-\frac{\partial\Psi}{\partial y}, \frac{\partial\Psi}{\partial x}, 0\right) \quad [2]$$

The geostrophic relation [1] – or relations, plural, if one prefers to think in components rather than vectors – can be motivated as an approximation to the horizontal momentum equation. The accuracy of that approximation depends on smallness of the Rossby number, or, more precisely, on being able to neglect relative particle (Lagrangian) accelerations against Coriolis accelerations, i.e., against f times either side of eqn [1]. The Rossby number Ro, measuring the advective contribution to the relative particle acceleration against the Coriolis acceleration, is usually of the same order as f^{-1} times a typical magnitude of the relative vertical vorticity $\partial v/\partial x - \partial u/\partial y$, the latter being equal to $\nabla_{\mathrm{H}}^2\Psi$ if eqn [2] holds. Here u and v are the horizontal velocity components corresponding to x and y, and ∇_{H}^2 is the horizontal Laplacian. Extratropical Rossby numbers have orders of magnitude typically $\sim 10^{-1}$ for synoptic-scale weather systems.

The geostrophic relation [1] was recognized long ago by weather observers as helpful in making sense of synoptic-scale surface pressure patterns, important for instance to ships threatened by cyclonic storms. The history is sometimes discussed under headings such as Buys Ballot's law and cyclonic development theory. Buys Ballot's law is a qualitative counterpart of eqn [1] with surface pressure in place of Φ, 'wind in your back means low on your left' – low surface pressure, that is, at sea level in the Northern Hemisphere.

Today's concept of balance recognizes that, like the rigid-pendulum approximation, eqn [1] is merely the first in a hierarchy of more accurate balance relations. Next in the hierarchy is the relation first studied by Bert Bolin and Jule G. Charney in the 1950s, in connection with efforts to develop practical numerical weather prediction. The Bolin–Charney balance relation retains eqn [2] even if f varies with latitude, but redefines Ψ to satisfy

$$\nabla_{\mathrm{H}}\cdot(f\nabla_{\mathrm{H}}\Psi) = \nabla_{\mathrm{H}}^2\Phi + \nabla_{\mathrm{H}}\cdot(\mathbf{u}\cdot\nabla_{\mathrm{H}}\mathbf{u}) \quad [3]$$

where, as before, ∇_{H} is horizontally two-dimensional. Equation [3] is an approximation to the divergence equation, the latter being the result of taking the horizontal divergence of the horizontal momentum equation. The relative particle acceleration is now retained. Its advective part gives rise to the last term of eqn [3] while the remaining, $\partial/\partial t$ part is annihilated when the horizontal divergence is taken, because of eqn [2]. It is only because there are no $\partial/\partial t$ terms that the relation [3], with [2], qualifies as a balance relation. In the special case of an isolated circular vortex on an *f*-plane, eqn [3] reduces to so-called gradient-wind balance, namely eqn [1] corrected to include the centrifugal force of the relative motion.

Again because of eqn [2], the right-hand side of eqn [3] can be rewritten using a Jacobian in u and v, as $\nabla_{\mathrm{H}}^2\Phi - 2\partial(u,v)/\partial(x,y)$, or equivalently using a Hessian in Ψ so that eqn [3] becomes

$$\nabla_{\mathrm{H}}\cdot(f\nabla_{\mathrm{H}}\Psi) = \nabla_{\mathrm{H}}^2\Phi - 2\left\{\frac{\partial^2\Psi}{\partial x^2}\frac{\partial^2\Psi}{\partial y^2} - \left(\frac{\partial^2\Psi}{\partial x\partial y}\right)^2\right\} \qquad [4]$$

Regarded as an equation for Ψ when the mass field Φ is given, eqn [4] is of Monge–Ampère type, with an extensive mathematical theory. Iterative methods are needed to solve it numerically because of the nonlinear terms on the right. The problem of finding Ψ becomes ill-posed for certain mass fields Φ, adumbrating, for one thing, that there exist mass fields not even approximately balanceable by any velocity field.

A simple thought experiment to make this last point clear would be to pile up the whole of the Earth's atmosphere into a narrow column above the North Pole, leaving a vacuum elsewhere. It is obvious that no velocity field **u** can be in balance with such a mass field. The free evolution at subsequent times, in any such thought experiment, would start with the column collapsing downward and outward and would involve sound and inertia–gravity waves of enormous amplitude. That is, it would involve gross imbalance as well as, almost certainly, violent wave-breaking and turbulence.

Balance relations are useful in practice only because naturally occurring mass fields, or at least smoothed versions of them are, by contrast, often balanceable to good approximation. In most such cases, eqn [4] with suitable boundary conditions is a well-posed nonlinear elliptic boundary-value problem in the flow domain, the primary exception being flows near the equator, where Rossby numbers are not small and eqn [4] may fail to be elliptic, as can be verified from the theory of Monge–Ampère equations. Again the failure of ellipticity adumbrates a physical reality (though not in a way that is quantitatively precise), namely the fact that balance is liable to break down spontaneously through 'inertial' and 'symmetric' instabilities near the equator, where *f* changes sign. There are other varieties of spontaneous imbalance, some only recently clarified. Again these are usually unimportant when Rossby numbers are small.

Balance relations still more accurate than eqn [4] can be defined if one is prepared to deal with more complicated sets of equations. The next relation in the hierarchy – to be referred to here as the generalized Bolin–Charney balance relation – is the first in the hierarchy to yield a nonvanishing vertical component of **u**. It was implicit in the pioneering work of Charney published in 1962, in a famous paper entitled 'Integration of the primitive and balance equations'. It starts with eqns [2] and [4] but then adds to the resulting **u** field a horizontally irrotational, divergent correction field governed by another elliptic boundary-value problem in the flow domain, a generalization of the omega equation previously developed by Norman A. Phillips and others. The corrected **u** field is an asymptotically consistent improvement on eqn [1], for small Rossby number Ro, in the sense that it is one order more accurate in powers of Ro. The elliptic boundary-value problem is derived by taking $\partial/\partial t$ of eqn [4], then eliminating all the resulting time derivatives using the exact mass conservation and vorticity equations and the inverse Laplacian of the vorticity equation. The vorticity equation expresses $\nabla_{\mathrm{H}}^2(\partial\Psi/\partial t)$ in terms of diagnostically known, or knowable, quantities such as the corrected **u** field; so the inverse Laplacian is needed in order to eliminate $\partial\Psi/\partial t$ from $\partial/\partial t$ of eqn [4].

This process of eliminating all the time derivatives including others such as $\partial\Phi/\partial t$ has to be possible, in principle at least, if the end result is to be a balance relation. By definition, a balance relation may not contain any time derivatives. When the elimination is carried out explicitly, a rather complicated set of integro-differential equations results, containing Green's function integrals whose details depend on the geometry of the flow domain. It may therefore be notationally and computationally more convenient to work with a set of equations from which $\partial\Psi/\partial t$, $\partial\Phi/\partial t$, etc., have not been eliminated, but have been allowed to remain as unknowns that can, in principle, be eliminated.

Then '$\partial\Psi/\partial t$', in scare-quotes, so to speak, must be regarded not as the rate of change of Ψ but, rather, as an auxiliary variable – better described as a diagnostic estimate of, as distinct from the actual, rate of change. Such a diagnostic estimate must be expected to differ, in general, from the actual rate of change of Ψ, for the reasons explained under *ultimate limitations* below. To avoid confusion over this point a special notation is sometimes used, such as Ψ_1 for a diagnostic estimate of $\partial\Psi/\partial t$ and Ψ_2 for $\partial^2\Psi/\partial t^2$, and so on.

The general form of the functional dependence defining a balance relation, assuming a balanceable mass field represented by $\Phi(\mathbf{x},t)$, can be written symbolically as

$$\mathbf{u}(\mathbf{x},t) = \mathbf{u}^{\mathrm{B}}(\mathbf{x};\Phi(\cdot,t)) \qquad [5]$$

where it is again emphasized that no derivatives or integrals with respect to t may appear. It must be possible, in principle at least, to eliminate them all. Time t enters solely via the second argument $\Phi(\cdot,t)$ of $\mathbf{u}^{\mathrm{B}}$. The notation $\Phi(\cdot,t)$ follows mathematical convention and signifies nonlocal spatial dependence. In other words, the second argument of $\mathbf{u}^{\mathrm{B}}$ is the whole *function*, Φ of $\mathbf{x}$, over the whole flow domain at the given instant t – not merely the *value* of Φ at the single value of $\mathbf{x}$ to which the left-hand side of eqn [5] refers. Such nonlocal functions are sometimes called functionals.

Even the geostrophic relation [1] is enough to illustrate the point, though it involves nothing more than the behavior of Φ in the immediate neighborhood of $\mathbf{x}$ – more precisely, it involves enough about that behavior to permit the evaluation of the two horizontal derivatives. The Bolin–Charney balance relations, generalized or not, are fully nonlocal, as is plain from the occurrence of elliptic partial differential operators like ∇_{H}^2 and, implicitly or explicitly, the associated Green's function integrals. To find **u** from Φ one has to solve elliptic partial differential equations in the flow domain, as already emphasized, implying for instance that the value of **u** at some position **x** will depend on values $\Phi(\mathbf{x}',t)$ at other positions $\mathbf{x}'$ well outside the neighborhood of **x**.

The generalized Bolin–Charney balance relation is often accurate enough for practical purposes, such as observational data analysis and assimilation, and the initialization of the full dynamics for numerical weather prediction. Of fundamental interest, however, from a theoretical viewpoint, is the fact that the pattern of elimination of time derivatives can be extended systematically to higher derivatives, often resulting in balance relations that are still more accurate.

The ideas involved seem to have been first explored by Karl Hinkelmann in the 1960s, in connection with the initialization problem in numerical weather prediction. They were later approached from another direction via the normal modes of the Laplace tidal equations, by Bennert Machenhauer, Ferdinand Baer, Joseph Tribbia and others. The history then went full circle, successively under headings such as nonlinear normal-mode initialization, bounded-derivative method, implicit normal-mode initialization, non-normal-mode initialization, and non-normal-mode filtering, of which the last four represent a rediscovery or further development of Hinkelmann's ideas. The ideas were first applied to accurate PV inversion by Warwick A. Norton in the late 1980s. An ingenious numerical approach bypassing the explicit consideration of diagnostic estimates like Ψ_1, Ψ_2, … or their normal-mode counterparts was developed by Álvaro Viúdez and David G. Dritschel in 2004.

The Ultimate Limitations

The most accurate balance relations can, in some circumstances, be far more accurate than values of parameters like the Rossby number Ro might ever suggest; and this accuracy extends over a far wider range of parameter values than could reasonably have been expected a priori – with Ro values of order unity, and even greater, in some cases. This astonishing fact – first discovered by Norton through numerical experiments on hemispherical shallow-water flows, for which $\mathrm{Ro} = \infty$ at the equator – cannot be deduced by inspection or scaling analysis of the momentum equation or other forms of the equations of motion. It involves great mathematical subtlety, and full understanding has yet to be achieved. Norton's most accurate results used the nonlinear normal-mode technique.

Some insight into the ultimate limitations on accuracy has come from classical aerodynamics. There have been many theoretical, experimental and numerical studies of the weak aerodynamic sound generation or *Lighthill radiation* named after M. James Lighthill's celebrated pioneering work of 1952. This is a simple form of spontaneous imbalance. It is now known that in continuously stratified flows there are further forms of spontaneous imbalance, neither instability-related nor Lighthill-like. Recent work at the cutting edge of this problem can be found in papers by D. Muraki, R. Plougonven, C. Snyder, A. Viúdez, and F. Zhang, appearing in the literature from about 2007.

The spontaneous-imbalance literature gives us a clear answer, in the negative, to a classic question posed in 1980 by Edward N. Lorenz. Could there be an exact balance relation? Could there be unsteady stratified, rotating flows that evolve in such a way that freely propagating sound and inertia–gravity waves are completely absent? More precisely, is there a slow manifold within the full phase space that is an invariant manifold of the full dynamics? Evolution on such a manifold would be such that spontaneous imbalance vanishes exactly.

Lighthill's arguments are enough to show that such a situation is overwhelmingly improbable. Though falling short of rigorous proof, they amount to a very strong heuristic. They show that, whatever else is going on, unsteady vortical flows are practically certain to emit sound and inertia–gravity waves, albeit sometimes very weakly; and practically all the flows of interest are unsteady. This means that spontaneous imbalance is generically nonzero, even though it may often be very close to zero, implying in turn that the so-called 'slow manifold' within the full phase space must be an infinite-dimensional chaotic layer. Though astonishingly thin in places – over a far wider range of parameter values than could reasonably have been expected a priori, as shown by Norton's work, in some cases at least – it is not a manifold, which by definition is infinitesimally thin. Though sometimes astonishingly accurate, the concept of balance is inherently and fundamentally approximate. The layer is sometimes referred to, therefore, as the slow quasimanifold.

(Arguably, a self-contradictory term like 'fuzzy manifold' is best avoided. By its mathematical definition a *manifold* is a perfectly sharp, smooth hypersurface and not at all fuzzy. Thus 'fuzzy manifold' would add yet another item to the list of self-contradictory terms like 'variable solar constant' and 'asymmetric symmetric baroclinic instability' – which of course we inevitably have to live with but, perhaps, need not add to.)

The fact that spontaneous imbalance can take a variety forms beyond those described by Lighthill's arguments does not change the conclusion that the slow quasimanifold is generically a chaotic layer. Adding to the repertoire of possible imbalance mechanisms can only reinforce that conclusion.

Balanced Models

As already indicated, the swinging modes of the elastic pendulum can be described in a simplified yet in some cases accurate manner by imposing a functional relation between bob speed and spring length, suitably chosen. This reduces the dimensionality of the dynamical system's phase space. Similarly, vortical flows can be described by simplified balanced models or balance models, so called. These are constructed by imposing a balance relation from the start, thereby forcing a true slow manifold into existence. The phase space of the original equations – usually taken as the hydrostatic 'primitive equations', so called – is collapsed into a smaller phase space, though still infinite-dimensional.

The initialization of a balanced model requires only a single scalar field to be specified, such as the mass field, or the PV field in the generalized sense. This scalar field is sometimes called the master field or master variable of the balanced model, to which all other dependent variables are slaved diagnostically. The model has only one prognostic equation, involving only one true time derivative, the rate of change of the master variable. This rate of change is to be sharply distinguished from the diagnostic estimates of time derivatives that may be hidden inside the definition of the balance relation [5], such as the diagnostic estimates Ψ_1, Ψ_2,… already mentioned.

Among those qualifying as balanced models in this standard sense are the models labeled quasigeostrophic theory, semigeostrophic theory, and the Bolin–Charney model, also called 'the' balance equations, in isentropic coordinates or in shallow water. In the Bolin–Charney model the master variable can equally well be taken as the mass or as the PV. Both are advected by the same velocity field, a velocity field that satisfies the generalized Bolin–Charney balance relation. Here, as implicitly above, the PV is the exact (Rossby–Ertel) PV and is to be evaluated with the same velocity field.

This property of having a single velocity field is unusual in balanced models. Unlike the Bolin–Charney model, most balanced models have different velocity fields to do different jobs. Semigeostrophic theory is a well-studied example. It has three separate velocity fields. The first advects PV and mass, the second evaluates energy and momentum, and the third evaluates the PV taken as the exact, Rossby–Ertel PV. This last fact is hidden from view in most expositions of the theory. Traditionally the model's PV is written in terms of the second velocity field, complicating the appearance of the formula for PV and disguising its origin. The three velocity fields differ from each other by fractional amounts $O(\mathrm{Ro})$ where Ro is again the Rossby number.

Semigeostrophic theory has remarkable mathematical properties but is comparable to quasigeostrophic theory in having $O(\mathrm{Ro})$ errors relative to the primitive equations. (Semigeostrophic theory is, however, superior in some respects, such as describing frontogenesis, albeit inferior in others such as describing mesoscale vortices.) The property of having more than one velocity field – for want of a better term we may call it 'velocity splitting' – was thought until recently to be a property of all balanced models with the sole exception of the Bolin–Charney model. In all other cases, refining the accuracy was thought to reduce greatly, but not to eliminate, disparities between the velocity fields of a model.

All this was indeed known to be true not only of Norton's and similar highly accurate balanced models, but true also of another important subclass of balanced models, namely all the Hamiltonian models that can be constructed by Salmon's method. Semigeostrophic theory is one of these. In the 1980s Rick Salmon showed how to construct balanced from unbalanced models in a systematic way within the Hamiltonian framework. Within that framework one imposes a balance relation as a constraint on the full dynamics, preserving the symplectic geometry of phase space. The constraint is imposed not only on dynamical trajectories but also on functional variations about those trajectories. The resulting balanced models are thus guaranteed to inherit Hamiltonian structure, as well as being accurate to the same formal order in Ro as the imposed balance relation.

A reason for using the Hamiltonian framework is the control it provides over conservation principles. The framework, properly applied, guarantees that the balanced model will exactly conserve mass, momentum, and energy as well as PV materially. However, there is a fundamental tension between balance relations and conservation principles. A balanced model tries to mimic vortical flows that in reality produce Lighthill radiation or other forms of spontaneous imbalance. The spontaneous imbalance must give rise to wave-induced fluxes of energy and momentum, none of which can be exactly described by the balanced model.

So if one forces a true slow manifold into existence by imposing a balance relation, while insisting that all conservation relations still hold, something has to give way. What gives way, as it turns out, is the concept of a unique velocity field. All balanced models constructed by Salmon's method exhibit velocity splitting, usually into two separate velocity fields but sometimes, as with semigeostrophic theory, into three. For more detail see the Further Reading list.

Even if we abandon energy and momentum conservation, there remains a possible tension between balance and local mass conservation. This is because spontaneous imbalance involves local adjustments in the mass field. Until recently, it was thought that this explained why the most accurate non-Hamiltonian balanced models then known still exhibit velocity splitting in one form or another, albeit by tiny amounts. So it was a further surprise when, thanks to recent work by A.R. Mohebalhojeh, a class of highly accurate balanced models was discovered that are entirely free of velocity splitting, yet as far as we know pay no systematic price in terms of accuracy, within shallow-water dynamics at least. Each such model has a unique velocity field, just as does the far-less-accurate Bolin–Charney model. The unique velocity field advects and evaluates the exact PV, advects mass, and evaluates energy and momentum which latter, however, are not conserved. One consequence, though, is that the models possess exact Casimir invariants (*see* **Dynamical Meteorology:** Potential Vorticity).

These new balanced models have been collectively designated *hyperbalance equations*. In order to write these equations one has to use functional derivatives, which are nonlocal, as well as ordinary partial derivatives. This may perhaps explain why the hyperbalance equations were not discovered long ago.

It is still an open question whether there will prove to be a tradeoff between accuracy and local mass conservation if we go beyond shallow-water dynamics, toward multi-layer models and continuously stratified reality.

Note on Terminology

The reader is warned that the terms geostrophic balance and its shorthand form, geostrophy, are sometimes used in the literature to mean balance more accurate than geostrophic, i.e., more accurate than eqn [1]. A common example is the self-contradictory phrase 'geostrophic adjustment', which refers to the mutual adjustment of the mass and velocity fields to approach balance or to stay close to balance – and balance, of course, in real fluid flow, nearly always means not geostrophy, eqn [1], but a more accurate balance within the generic class [5]. The example of a circular vortex adjusting toward ageostrophic, gradient-wind balance while radiating inertia–gravity waves is enough to illustrate the point. As already mentioned, gradient-wind balance is the particular case of Bolin–Charney balance that applies to a steady circular vortex, equivalent to eqn [1] plus a correction term representing relative centrifugal force. Thus by implication we have another piece of self-contradictory terminology, 'ageostrophic geostrophic adjustment', unfortunately well established.

It may also be noted that the term adjustment is itself used in two distinct senses that are sometimes confused with each other. The first is Rossby or initial-condition adjustment, the mutual adjustment of the mass and velocity fields toward balance that occurs primarily because a system is started in an unbalanced state. The second is spontaneous adjustment, the continual mutual adjustment of the mass and velocity fields to stay close to balance in unsteady vortical flow, even after initial conditions are forgotten. This second process is far more subtle and sets the ultimate limitations of the balance concept itself, the degree of fuzziness of the slow quasimanifold.

For all the foregoing reasons, some authors are beginning to avoid the term geostrophic adjustment, instead using the terms Rossby adjustment or initial-condition adjustment in the first case, and spontaneous adjustment or spontaneous imbalance in the second.

The term semigeostrophic theory is used here in its standard sense, referring to the balanced model originally introduced by Brian J. Hoskins in 1975. The reader is warned that in Salmon's papers the same term, semigeostrophic theory, is used in a different, more generic sense.

See also: **Data Assimilation and Predictability:** Data Assimilation. **Dynamical Meteorology:** Coriolis Force; Hamiltonian Dynamics; Inertial Instability; Kelvin Waves; Kelvin–Helmholtz Instability; Lagrangian Dynamics; Potential Vorticity; Primitive Equations; Quasigeostrophic Theory; Symmetric Stability; Wave Mean-Flow Interaction. **Gravity Waves:** Buoyancy and Buoyancy Waves: Theory. **Mountain Meteorology:** Lee Waves and Mountain Waves.

Further Reading

Hinkelmann, K.H., 1969. Primitive equations. WMO No. 297. In: Bykov, V.V. (Ed.), Lectures on Numerical Methods of Short-Range Weather Forecasting (Regional Training Seminar of the World Meteorological Organization). Hydrometeoizdat, Leningrad, pp. 306–375 (This paper is of historical as well as fundamental interest. Hinkelmann was the first to show how to construct highly accurate balance relations, for the purpose of filtering from initial conditions the so-called meteorological noise represented by sound and inertia–gravity waves. This noise, i.e. imbalance, was known to have been the main reason for a magnificent scientific failure reported in 1922—the failure of Lewis Fry Richardson's pioneering attempt at a numerical weather forecast using the primitive equations without initialization.).

Lighthill, M.J., 1952. On sound generated aerodynamically. I. General theory. Proceedings of the Royal Society of London A 211, 564–587 (This famous, lucidly-argued and penetrating classic was the first to reveal the surprising properties of acoustic imbalance for non-rotating, nearly incompressible flow in three dimensions.).

McIntyre, M.E., Norton, W.A., 2000. Potential-vorticity inversion on a hemisphere. Journal of the Atmospheric Sciences 57, 1214–1235. Corrigendum 58, 949. (Section 7 describes the only available investigation of a fundamental issue neglected above—how to make (5) Galilean invariant as well as highly accurate. Section 8 points to the possible tension between accuracy and local mass conservation, now surprisingly discounted by A.R. Mohebalhojeh's recent work on the hyperbalance equations.).

McIntyre, M.E., 2009. Spontaneous imbalance and hybrid vortex–gravity structures. Journal of the Atmospheric Sciences 66, 1315–1326 (This review brings into one convenient place the disparate recent works of Muraki, Plougonven, Snyder, Viúdez, and Zhang that together have clarified when the Lighthill paradigm does and does not apply.).

Norbury, J., Roulstone, I. (Eds.), 2002. Large-Scale Atmosphere–Ocean Dynamics. Geometric Methods and Models, vol. II. University Press, Cambridge (This volume covers many of the deeper mathematical aspects of balanced models, especially Hamiltonian balanced models, including a thorough discussion of Hamiltonian velocity splitting by Roulstone and the author and a wide-ranging, in-depth review of the elastic-pendulum problem by P. Lynch. In particular, Section 5.3 gives a brief but careful discussion of relation between slow manifolds and unbroken KAM tori in the pendulum problem, following the work of O. Bokhove and T.G. Shepherd. In this regard the pendulum problem is rather different from the fluid-dynamical problem.).