Acta Numerica (2006), pp. 385–470
doi: 10.1017/S0962492906250013
© Cambridge University Press, 2006
Printed in the United Kingdom

Numerical modelling of ocean circulation

Robert L. Higdon
Department of Mathematics,
Oregon State University,
Corvallis, Oregon 97331-4605, USA
E-mail: higdon@math.oregonstate.edu

Computational simulations of ocean circulation rely on the numerical solution of partial differential equations of fluid dynamics, as applied to a relatively thin layer of stratified fluid on a rotating globe. This paper describes some of the physical and mathematical properties of the solutions being sought, some of the issues that are encountered when the governing equations are solved numerically, and some of the numerical methods that are being used in this area.

CONTENTS

1 Introduction 385
2 Physical characteristics of large-scale oceanic flows 388
3 The choice of vertical coordinate 394
4 Analytical properties of solutions 399
5 Time discretization 410
6 Spatial discretization and related issues 415
7 Governing equations in general coordinates 427
8 Summary 465
References 466

1. Introduction

The circulation of the world's oceans plays a major role in the global climate system. For example, the circulations of the atmosphere and ocean move large amounts of heat from tropical regions to higher latitudes, and this transport serves to moderate the temperature differences caused by unequal solar heating. In the case of the ocean, much of the transport is due to intense currents that are typically found along the western boundaries of ocean basins, and some is also due to turbulent mixing due to eddies.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

An important part of the picture is the buoyancy-driven 'thermohaline circulation', in which a portion of the warm water in the Gulf Stream flows to the far northern Atlantic, becomes colder and saltier due to atmospheric forcing, sinks due to the increased density, and then moves slowly southward along the ocean's bottom to become part of a global circulation with a period of roughly a millennium.

An illustration of the circulation of the ocean is given in Figure 1.1. The figure shows the sea-surface temperature in the western Atlantic, at a fixed time, as computed in a numerical simulation using the Miami Isopycnic Coordinate Ocean Model (Bleck, Rooth, Hu and Smith 1992, Bleck 2002). The black regions represent land masses, and the various shades of grey indicate water temperatures. (Bright regions do not indicate the highest temperatures. Instead, this greyscale plot was reproduced from a colour original, and in some cases different colours were mapped to similar shades of grey.) The warmest waters are located in the southernmost region, and these are seen to move northward along the eastern coast of the United States and then into the interior of the North Atlantic. The figure suggests the role of the ocean in the Earth's climate, and it also illustrates how ocean currents exhibit small-scale meanders and eddies in addition to their larger-scale patterns.

In order to assess various scenarios for the future evolution of the Earth's climate, numerous groups have used numerical simulations based on coupled climate models that include models of the atmosphere, ocean, sea ice, and effects of the land surface. These models are run for centuries of model time and are very computationally intensive. Extensive information about the usage of such models is included in a recent report by the Intergovernmental Panel on Climate Change (2001).

More localized aspects of ocean circulation are also of interest, for scientific and societal reasons. For example, in certain coastal regions (such as along the northwestern United States) the prevailing winds sometimes cause the surface waters to shift offshore, so that the relatively cooler deep waters well upward to the surface. This upwelling brings nutrients that support the base of the food chain. These coastal upwelling regions occupy at most a few per cent of the area of the world's oceans, but in biological terms they are extraordinarily productive (Gill 1982).

There are multiple reasons for obtaining a better understanding of the circulation of the ocean, and numerical simulation is an important tool for gaining that understanding. The design of numerical algorithms for this purpose is heavily influenced by the physical properties of the flows being modelled and by the mathematical properties of the solutions of the governing equations. A description of some of these properties is a major emphasis of this review.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press



Figure 1.1. Sea-surface temperature in the western Atlantic Ocean, as computed in a numerical simulation. (Figure provided by Rainer Bleck, NASA Goddard Institute for Space Studies.)

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Section 2 gives an overview of length, time, and velocity scales and some other physical properties of oceanic flows. Section 3 describes the problem of choosing a vertical coordinate for an ocean circulation model; the geometrical height z is a traditional choice, but there are thermodynamic and hybrid alternatives that are currently receiving serious attention. Section 4 summarizes some properties of solutions of the governing equations, including geostrophic balance, conservation of potential vorticity, Rossby waves, and the multiple time scales resulting from the contrast between fast external waves and the slower internal and advective motions. Section 5 discusses the numerical treatment of these multiple time scales, and it also describes some time-stepping schemes. Section 6 discusses grids and spatial discretizations, the numerical simulation of advection, and the numerical solution of the momentum equations.

Section 7 gives a detailed derivation of the partial differential equations that describe the conservation of mass, momentum, and tracers for a stratified and hydrostatic fluid that is in motion relative to a rotating spheroid. In this discussion the horizontal coordinates are arbitrary orthogonal curvilinear coordinates, and the vertical coordinate is a generalized coordinate that includes all of the cases discussed in Section 3. This derivation is placed at the end of the paper because of its length, and the preceding sections refer forward to it when needed.

2. Physical characteristics of large-scale oceanic flows

The physical motions of the ocean exhibit a wide range of space and time scales, and when viewed on a sufficiently large scale the circulation of the ocean is very nearly hydrostatic. In addition, the interior of the ocean is stratified, and over much of the ocean the relatively warm upper layers are nearly isolated from the relatively cold lower layers. These and other physical properties of the ocean are outlined in the present section.

2.1. Length scales and the hydrostatic assumption

Table 2.1 shows some features of oceanic flows that are resolved in simulations on a basin scale or global scale. The phrase 'horizontal gyres' refers to closed loops of circulation that are driven by the wind stress acting at the upper surface of the ocean. For example, a portion of the Gulf Stream returns laterally and southward to complete a closed loop, and the Kuroshio current plays an analogous role in the North Pacific.

The table indicates that the vertical length scale in the ocean is much smaller than the lateral dimensions of the various features listed there. The statement that the vertical length scale is small compared to the horizontal length scale is known as the shallow water condition, and it is one of the fundamental assumptions of large-scale ocean modelling. In colloquial terms,

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Table 2.1. Approximate length scales in the ocean. All quantities are horizontal dimensions, unless otherwise stated.

Ocean basins	$\sim$ 5000 to 10000 km
Horizontal gyres	basin scale
Western boundary currents	$\sim$ 100 km
Eddies	tens to hundreds of km
Depth of the ocean	$\sim$ 4 to 5 km, in the mid-ocean
Depth of surface currents	hundreds of metres

this condition can be phrased as follows: *The Pacific Ocean is shallow. Your coffee cup is deep.*

One consequence of the shallow water condition is that, for large-scale modelling, the ocean can be regarded as nearly hydrostatic, *i.e.*, the vertical acceleration of the fluid is negligible. This conclusion can be reached by a formal scaling analysis, as given, for example, by Holton (1992) or Pedlosky (1987). For a more informal argument, let L be a horizontal distance over which the horizontal velocity varies significantly. This variation in horizontal velocity can induce a horizontal divergence that causes the elevation of the free surface at the top of the fluid to rise or fall. However, this divergence acts over a vertical distance that is much smaller than L, and since the effect of this divergence is spread out over a large horizontal extent, the elevation of the free surface must change very slowly. For an analogy, suppose that a large swimming pool is being filled by a high-volume pump. An observer next to the pump may think that the water is flowing rapidly into the pool, but the water level changes slowly because the effect of the pumping is spread over a large horizontal area.

The hydrostatic assumption is not valid for motions whose horizontal scale is comparable to the vertical scale, or smaller. For example, in the case of the thermohaline circulation, the sinking of water in the far northern Atlantic takes place in fields of convective plumes; the fields are on the order of a hundred kilometres wide, and the width of each individual plume is on the order of a kilometre or less. The downdraft in such a plume is not hydrostatically balanced, and it also cannot be resolved in a climate model using present-day computers. Instead, this process must be represented by some kind of parametrization (Marshall and Schott 1999). In addition, limited-area models of near-shore processes can have grid spacings that are fine enough to resolve nonhydrostatic motions, and in that case the models must be configured accordingly. However, the emphasis in this review is on physical processes and numerical methods associated with large-scale circulation, and so the hydrostatic condition will assumed here except when stated otherwise.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

2.2. Stratification

One simple model of oceanic motions is provided by the shallow water equations (Section 4.2), which describe a hydrostatic fluid having constant density. This system of equations is sufficient, for example, to enable highly accurate modelling of global tides (Bennett 2002). However, the assumption of constant density is only an approximation. Within a given water column, the density can vary by a few per cent from the top of the fluid to the bottom, and additional variation is found over the lateral extent of the ocean (Gill 1982). It turns out that these variations in density play a crucial role in major features of the ocean's dynamics, such as the thermohaline circulation, and so they must be incorporated into models of the general circulation of the ocean.

The density in the ocean is determined by temperature, salinity, and pressure, with most of the vertical variation within a water column being due to the pressure. However, the dependence on pressure is dynamically not very significant (Sun *et al.* 1999), and in physical oceanography it is common practice to use the concept of 'potential density', which is the density after adiabatic adjustment to a reference pressure. That is, imagine that the pressure acting on a fluid parcel is changed to some fixed reference pressure, without any heat flowing into or out of the parcel and without any changes in the salinity of the parcel. The reference pressure could be atmospheric pressure or the pressure at some specified depth. The density after this adjustment depends only on temperature and salinity, and it is referred to as the potential density. Similarly, the 'potential temperature' is defined to be the temperature of a fluid parcel after the same kind of adjustment to a reference pressure.

The stratification of the ocean is illustrated in Figure 2.1, which shows a contour plot of the zonal (east–west) average of potential density in the Pacific Ocean, with the potential density referenced to atmospheric pressure. This plot represents a climatological mean taken from the atlas of Levitus (1982). In this plot the vertical coordinate is the depth, in metres, and the horizontal coordinate is the latitude, in degrees, from 90°S to 90°N. The horizontal axis thus extends for approximately 20,000 kilometres, whereas the vertical axis covers 5.5 kilometres. One feature seen in the plot is that some of the surfaces of constant potential density outcrop to the surface, owing to the lateral variation in temperature, and some of the surfaces intersect the bottom topography. Another feature is that most of the variation of potential density is found near the upper surface, whereas the abyssal waters are relatively homogeneous.

One feature not shown in Figure 2.1 is the mixed layer at the top of the ocean. The uppermost part of the ocean tends to be vertically homogeneous, due in part to mixing caused by wind forcing. In circumstances where the

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

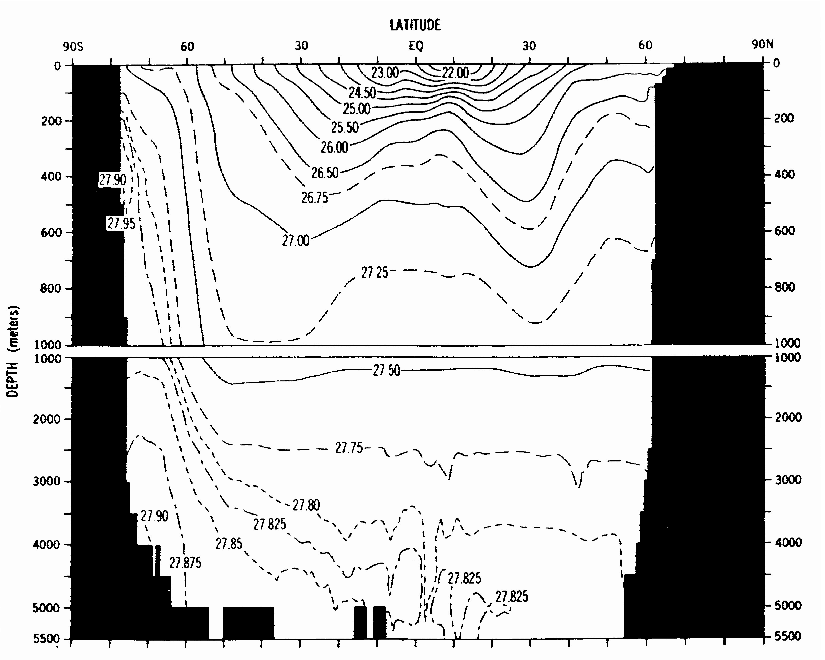


Figure 2.1. East–west average of potential density in the Pacific Ocean. The contour labels represent values of potential density minus 1000 kg/m^3; for example, the contour labelled 27.00 corresponds to potential density 1027 kg/m^3. (From Levitus (1982).)

atmosphere is colder than the ocean, mixing is also generated by convective overturning within the upper ocean. The thickness of the mixed layer is typically on the order of tens of metres to perhaps hundreds of metres. However, in deeply convective regions such as those in the far northern Atlantic, the mixed layer can sometimes be more than 2000 metres thick (Marshall and Schott 1999).

2.3. Time and velocity scales

Table 2.2 lists some approximate time and velocity scales for prominent motions in the ocean. The table includes the periods of some large-scale circulation patterns, and it also lists the velocities of currents and of two types of external waves and two types of internal waves: see, *e.g.*, Gill (1982). In the case of an external wave, all fluid layers thicken or thin by approximately the same proportion at a given horizontal location and time, as illustrated in Figure 2.2. For such a wave, the behaviour of the free surface at the top of the fluid reveals the nature of the wave motion throughout

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Table 2.2. Time and velocity scales.

Thermohaline circulation	$\sim$ millennium
Wind-driven gyres	decades
Currents	$\sim$ 1 m/sec for strong currents, elsewhere much less
External gravity waves	$\sim \sqrt{gH}$, *e.g.*, $\sim$ 200 m/sec
Internal gravity waves	a few m/sec or less
External Rossby waves	$\sim$ 20 m/sec
Internal Rossby waves	$\sim$ 0.02 to $\sim$ 1 m/sec, depending on the latitude

the interior, and the horizontal velocity field is essentially independent of vertical position. On the other hand, an internal wave consists of undulations of surfaces of constant density within the fluid, and the free surface remains nearly level.

In the case of external gravity waves the restoring force is due to the density contrast between water and air, whereas for internal gravity waves the restoring force is due to variations of density within the fluid. The restoring force is much weaker in the latter case, so internal gravity waves move much more slowly than external gravity waves. In the case of Rossby waves the restoring mechanism is based on vorticity instead of gravity, and the existence of such waves depends on the variation of the Coriolis parameter with latitude and/or variations in the topography at the bottom of the fluid domain. Some properties of the above motions are derived mathematically in Section 4.

In general, fluids can also admit sound waves, which travel much more rapidly than the waves listed in Table 2.2. However, the circulation of the ocean is not affected by the propagation of sound waves within the ocean,

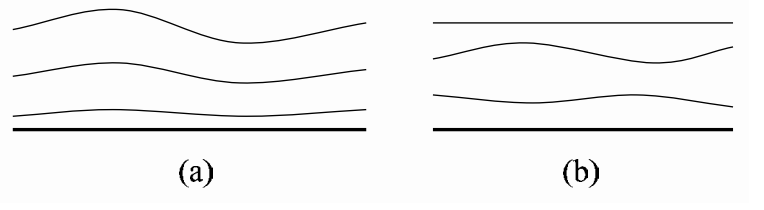


Figure 2.2. (a) External wave. (b) Internal wave. In each case, the uppermost curve represents the free surface at the top of the fluid, and the bottom line represents the bottom of the fluid domain. The intermediate curves represent surfaces of constant density within the fluid. The vertical displacements are exaggerated for the sake of visibility.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

and these waves can be filtered from the equations of motion by using either the hydrostatic approximation or the Boussinesq approximation (*e.g.*, Gill (1982), Griffies (2004)). With the latter approximation, the density of the fluid is assumed constant in the momentum equations, except in the buoyancy term in the vertical momentum equation.

In a numerical model of the general circulation of the ocean, the fastest motions are typically the external gravity waves. These travel much more rapidly than any of the other motions that are present in the system, and their presence can cause serious problems with the efficiency of a numerical algorithm. In the ocean modelling community it has therefore become common practice to split the external motions into a separate subsystem that is solved by different techniques from the remainder of the system. This issue is discussed in Section 5.1.

One implication of the space and time scales of the large-scale circulation of the ocean is that the Coriolis term and lateral pressure gradient are typically in an approximate balance, known as the geostrophic balance. The same property also applies to the circulation of the atmosphere, and it is one of the most prominent features of geophysical fluid dynamics. When a system is in a geostrophically balanced state, the fluid flows along curves of constant pressure. The concept of geostrophic balance is developed in Section 4.3.

2.4. Subgrid-scale processes

If a numerical model of ocean circulation has a horizontal grid spacing of 0.1° latitude and longitude, for example, then the grid spacing near the equator is roughly 10 km. For a global ocean model this is presently considered high resolution, and many climate-scale studies are currently run with grid spacings that are much coarser, given the long time intervals that are involved (Intergovernmental Panel on Climate Change 2001). However, the dynamics of the ocean are influenced by the cumulative effects of turbulent processes that occur on scales that are too small to be resolved explicitly in a numerical model. For example, these scales can extend downward as far as centimetres or less (Gill 1982, Bleck 2005). It is then necessary to parametrize the large-scale effects of these subgrid-scale motions in terms of the dependent variables that are used in the model.

These processes affect both the transport of momentum and the transport of tracers. The term 'tracer' refers to a scalar-valued quantity that is transported with the flow, such as potential temperature, salinity, or the concentrations of various chemical or biological species. The first two of these affect the density of the fluid and thus its dynamics, whereas the last do not. These classes of tracers are labelled as active tracers and passive tracers, respectively.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

On the scales of interest here, the mixing of momentum and tracers is highly anisotropic, and it occurs primarily along directions that are approximately horizontal. This phenomenon has been described in terms of directions of neutral buoyancy, which at any point in the fluid are defined locally as lying in the plane of constant buoyancy through that point; this plane is tangent to the surface of constant potential density through that point, where in this case the potential density is referenced to the local pressure (McDougall 1987). If a fluid parcel leaves such a neutral plane, then the buoyancy contrast with the surrounding fluid tends to force the parcel back towards that plane. On the other hand, adiabatic motions within neutral planes encounter no such impediment, so mixing within neutral planes is far more efficient than mixing across such planes. This picture is a local description, and due to a lack of integrability it can be extended globally to a concept of 'neutral surface' in only an approximate manner (Eden and Willebrand 1999). The preceding issues are closely related to the problem of choosing a thermodynamic variable that can be used as a vertical coordinate for an isopycnic model, which is discussed in Section 3.1.

Empirical observations indicate that, in the interior of the ocean, the rate of transport along the principal directions of mixing can be as much as 10^8 times the rate of mixing in orthogonal directions (Ledwell, Watson and Law 1993, Davis 1994). However, this discrepancy is not as large in boundary layers, especially in the mixed layer at the top of the ocean.

Proper choices of subgrid-scale parametrizations have long been an area of active research. Irreversible mixing processes tend to be modelled by some kind of downgradient diffusion. For another example, Gent and McWilliams (1990) and Gent, Willebrand, McDougall and McWilliams (1995) introduced a parametrization of unresolved eddies that can be used, for instance, to represent the transfer of potential energy into eddy kinetic energy when slanted density surfaces are flattened. Extensive reviews of the problem of subgrid-scale parametrization are given by Griffies (2004) and by Haidvogel and Beckmann (1999).

3. The choice of vertical coordinate

A crucial decision in the construction of an ocean model is the choice of vertical coordinate, and an obvious choice is the geometrical height z. This coordinate is used, for example, in the Bryan–Cox class of ocean models. The development of the original Bryan–Cox model began in the 1960s (Bryan 1969), and the various descendants of this model constitute what is currently the most widely used class of ocean circulation models; a genealogy of this class is given by Semtner (1997). Examples include the Modular Ocean Model (Griffies, Harrison, Pacanowski and Rosati 2004), the Parallel Ocean Program (Dukowicz and Smith 1994), and the Parallel

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Ocean Climate Model (Semtner and Chervin 1992). Another example of a z-coordinate model is the MIT general circulation model (Marshall, Hill, Perelman and Adcroft 1997), which can represent both large-scale hydrostatic and small-scale nonhydrostatic processes.

Models that use z as the vertical coordinate are known as level models. However, other possibilities for the vertical coordinate are also used, and the purpose of this section is to describe and compare these alternatives.

3.1. Isopycnic coordinates

One alternative to z is isopycnic coordinates, for which the vertical coordinate is the reciprocal of potential density or some other related quantity. The term 'isopycnic' means 'equal density', which is an approximate statement about the surfaces of constant vertical coordinate in this setting. In the following discussion, the term 'isopycnal' refers to a surface of constant potential density (or related quantity), and 'diapycnal' refers to transport across such a surface.

In the setting of isopycnic coordinates, one seeks a quantity s that is (approximately) conserved along particle paths. In other words, one wants $Ds/Dt \approx 0$, where D/Dt denotes the material derivative (see Section 4.2). One also wants surfaces of constant s to be approximate neutral surfaces. For the moment, assume $Ds/Dt = 0$ exactly, everywhere in the fluid. In that event, a surface of constant s is a material surface; if a fluid parcel lies initially in such a surface, then as time evolves the fluid parcel will retain the same value of s and thus remain in the same surface. The elevation of such a surface can vary with horizontal position and time, but any two such surfaces will always enclose the same mass of fluid. If the fluid domain is discretized with respect to s, for purposes of numerical solution, then the fluid is divided into physical layers that do not mix. Water masses with distinct physical properties are then distinguished automatically by the choice of coordinate system. For this reason, isopycnic models are also referred to as 'layered' models.

If s is chosen to be the reciprocal of potential density or a related quantity, then the preceding statements are approximately true, except in the vertically homogeneous mixed layer at the top of the ocean. If such a quantity s is used as the vertical coordinate, then a graph such as Figure 2.1 provides a plot of the model's vertical coordinate system.

However, potential density, as such, is not necessarily an optimal choice for a vertical coordinate. For example, potential density is not exactly a neutral variable, and under certain conditions potential density can be non-monotonic as a function of z. In addition, if potential density is used as the vertical coordinate, then the form of the lateral pressure forcing in the momentum equations allows the possibility of numerical inaccuracy

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

and even numerical instability. Several investigators have recently explored these issues and have developed variations on potential density that are more suitable for use in an ocean circulation model: see, *e.g.*, Section 7.14 and the work of Sun *et al.* (1999), de Szoeke (2000), de Szoeke, Springer and Oxilia (2000), and Hallberg (2005).

As noted in Section 7, the equation for conservation of mass in isopycnic coordinates amounts to a statement about the thicknesses of the coordinate layers. The vertical diffusion of heat and/or salt causes a vertical movement of isopycnals relative to the fluid, and in the context of isopycnic coordinates this movement can be viewed as an advection of fluid across a coordinate surface.

The development of isopycnic coordinate ocean models began in earnest in approximately the 1980s, in analogy to the usage of isentropic coordinates in atmospheric models (Hsu and Arakawa 1990). Examples of isopycnic ocean models are the Miami Isopycnic Coordinate Ocean Model (Bleck and Smith 1990, Bleck 2002) and the Hallberg Isopycnal Model (Hallberg 1997).

3.2. Comparison of z and isopycnic coordinates

Over most of the real ocean, the warm upper layers are nearly isolated from the colder deep layers, and the diapycnal transports between layers are typically subtle. However, it is important to represent these transports accurately. For example, the thermohaline circulation involves warm water flowing to the far northern Atlantic, becoming colder and saltier, and then sinking; the dynamics of this circulation could be misrepresented in a numerical model if the model allows inaccurate or spurious diffusion of heat or salt between the upper and lower regions.

In the case of isopycnic coordinates, the representation of the diapycnal transport is under the explicit control of the modeller. In contrast, a level (z-coordinate) model can implicitly allow spurious, non-physical diapycnal transports. The transports of heat and salt are typically represented by advection-diffusion equations, and computational algorithms for representing advection typically exhibit diffusion that is purely numerical. For the terms involving horizontal advection, this numerical diffusion acts along surfaces of constant z. However, surfaces of constant z can intersect isopycnals, as suggested by Figure 2.1, and this numerical diffusion can then cause artificial transport between different water masses (Griffies, Pacanowski and Hallberg 2000*b*).

The diffusion terms in transport equations are another potential source of spurious transport in level models. However, it is possible to rotate the diffusion operator to yield a component tangent to isopycnals and a component perpendicular to isopycnals, with the rate of diapycnal diffusion

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

being orders of magnitude smaller than the rate of diffusion tangent to isopycnals. This process is reviewed in detail by Griffies (2004).

The above difficulties are avoided if isopycnic coordinates are used, but isopycnic coordinates have their own limitations. As seen in Figure 2.1, isopycnals can intersect the upper or lower boundaries of the fluid, and this implies a loss of vertical resolution, especially at the higher latitudes. A loss of resolution is especially pronounced in the mixed layer at the upper boundary of the ocean, as this layer is vertically homogeneous. In addition, when an isopycnal intersects the top (bottom) of the fluid domain, the layer above (below) the isopycnal approaches zero thickness. The vanishing of layers introduces algorithmic complications that are described in Section 6.

3.3. Sigma-coordinates

Another vertical coordinate is the terrain-following coordinate σ, which is chosen to vary linearly from $\sigma = 0$ at the upper boundary of the fluid to $\sigma = -1$ at the bottom boundary. This coordinate allows for an accurate representation of the topography of the ocean's bottom, and it facilitates the modelling of the bottom boundary layer.

However, σ-coordinate surfaces can intersect isopycnals, so the use of σ allows the possibility of spurious diapycnal transport, in analogy to the case of z-coordinates discussed above. Another limitation concerns the accurate computation of the lateral pressure gradient. When the pressure is written in σ-coordinates, the lateral pressure gradient $\nabla_\sigma p = (\partial p/\partial x, \partial p/\partial y)$ actually represents differentiation along surfaces of constant σ, not constant z. Conversion to a pressure gradient $\nabla_z p$ that is truly horizontal requires a correction term involving the hydrostatic balance and the slope of the σ-surface (Griffies 2004). In some regions this correction can have a magnitude comparable to that of $\nabla_\sigma p$. In such situations $\nabla_z p$ can be the difference of two large quantities of opposite signs, and errors in those quantities can dominate the result.

The use of σ as a vertical coordinate has been especially prominent in limited-area regional and coastal modelling, as in the Princeton Ocean Model (Blumberg and Mellor 1987) and the Regional Oceanic Modeling System (Shchepetkin and McWilliams 2005).

3.4. Hybrid coordinates

As noted in Section 3.2, isopycnic coordinates suffer the disadvantage of losing vertical resolution when isopycnals intersect the upper or lower boundaries of the fluid domain, and in addition they provide little resolution in vertically homogeneous regions such as the mixed layer at the top of the ocean. An attempt to overcome these limitations is represented by the

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

class of hybrid coordinates, which combine features of isopycnic coordinates, z-coordinates, and σ-coordinates. Examples of hybrid coordinate models include the Hybrid Coordinate Ocean Model (Bleck 2002), which is derived from the Miami Isopycnic Coordinate Ocean Model; HYPOP (Dukowicz 2004); and Poseidon (Schopf and Loughe 1995).

In the general framework used in these models, each coordinate layer is assigned a target density. (For convenience, the present subsection uses the word 'density' to refer to a quantity related to potential density that would be used in an isopycnic framework, as discussed in Section 3.1.) If a given layer's target density lies within the range that is represented in the fluid at a given horizontal location and time, then the layer is assigned that density, and the layer at that location is isopycnic. However, if a given layer's target density is not present in the fluid, then in a purely isopycnic framework the layer would reduce to zero thickness; this would be the case, for example, for a layer of relatively low density at a high latitude. With a hybrid coordinate, such a layer is inflated in order to have positive thickness and thus contribute to the vertical resolution of the model. There is considerable freedom in the choice of methods for carrying out this process, but the basic idea is to impose a minimum thickness and then provide a recipe for the transition between the isopycnic state and the state where the layers are purely geometric (Bleck 2005). This use of a moving mesh resembles the Arbitrary Lagrangian–Eulerian (ALE) method of Hirt, Amsden and Cook (1974), although the ALE technique does not involve restoration to a target density.

Bleck (2002) recommends that any isopycnic layers that intersect bottom topography not be inflated, as the computation of horizontal pressure gradients in that situation could be inaccurate for the same reasons discussed in Section 3.3 for the case of σ-coordinates. Instead, the hybridization process should focus on layers near the top of the fluid. He also recommends that the minimum layer thickness be constant in the ocean's interior in order to avoid spurious dynamics resulting from the stretching of fluid columns within a layer. However, in relatively shallow regions near coastlines, the nonzero coordinate layers can be scaled with depth as with the terrain-following σ-coordinate, since vortex columns are likely to extend over the entire depth of the fluid in that case.

One difficulty with hybrid coordinates, as stated by Bleck (2002), is related to the seasonal cycle of the surface mixed layer. During winter, this layer is relatively thick, due to strong mixing caused by storms and by convective overturning. However, during the summer this forcing is much weaker, and consequently much of the fluid in this layer reverts to a stratified state. In a hybrid model, isopycnic layers are thus alternately destroyed and then re-created near the top of the fluid as part of the annual cycle. If the geometrically constrained, non-isopycnic layers are evenly distributed

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

throughout a deep mixed layer during winter, for the sake of resolution, then they may need to migrate a long distance upward in order to reach their target densities in the spring or summer. This long migration implies large transports of fluid between coordinate layers; numerical errors in representing such transports can then cause a vertical dispersal of water properties, such as the concentrations of tracers in various layers. Such a dispersal could be alleviated by clustering the non-isopycnic layers near the top during the winter, but this would imply an uneven vertical resolution in that case.

The use of hybrid coordinates is relatively new and is a topic of current research. Issues currently debated include the details of transport and interpolation algorithms for carrying out the re-gridding from one state to the next, and whether the re-gridding should be performed at every time step or less frequently.

4. Analytical properties of solutions

We next describe some basic properties of solutions of the governing equations for the physical system considered here. This is done as a prelude to discussing methods for solving those equations numerically.

4.1. Governing equations

Section 7 gives a detailed derivation of partial differential equations that describe the conservation of mass, momentum, and tracers for a hydrostatic and stratified fluid that is in motion relative to a rotating spheroid such as the Earth. These equations are often called the 'primitive equations'; the term 'primitive' refers to the fact that the momentum equation is expressed in terms of velocity or momentum. In contrast, certain simplified models used for analytical or numerical studies employ dependent variables that are derived from velocity, such as vorticity, divergence, and streamfunction (*e.g.*, Pedlosky (1996), Miller (2006)). The general circulation models quoted in this review largely employ the primitive equations, with the exception that nonhydrostatic effects are sometimes considered.

In Section 7 it is assumed that the horizontal position on the spheroid is represented with general orthogonal curvilinear coordinates and that the vertical coordinate is a generalized coordinate s that includes all of the cases discussed in Section 3. However, in Section 7 the representation of the diffusion of momentum and tracers assumes that the vertical coordinate is an isopycnic coordinate, and for other vertical coordinates the diffusion terms would have to be modified appropriately.

In Section 7 the equations are first derived for the continuous case. In anticipation of solving these equations numerically, the equations are also

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

integrated vertically between surfaces of constant s to yield equations involving mass-weighted vertical averages. Such equations can be used in a vertically discrete system obtained by partitioning the fluid into coordinate layers.

The conservation of mass for the fluid is expressed in equations (7.32)–(7.33) for the continuous case and in equation (7.38) for the case of a discrete layer. In the latter case the dependent variable is essentially the mass per unit horizontal area in the layer, and its evolution is determined by lateral mass transport and by the vertical motions of layer interfaces relative to the fluid. The conservation of mass for tracers in the fluid is described by equations (7.106) and (7.107) for the continuous and vertically discrete cases, respectively.

The conservation of horizontal momentum is expressed by the equations (7.92) for the continuous case and by equations (7.101) and (7.104) for the vertically discrete case. In the latter equations the dependent variables are momentum density (velocity times the mass per unit horizontal area) and the transport terms are written in flux form, which facilitates the usage of (nearly) non-oscillatory advection schemes for those terms. The momentum and mass equations can also be combined to yield equations for which the unknowns are components of velocity instead of momentum density. Also needed for the momentum equation is the relation (7.105), which is a discretization of the hydrostatic condition in generalized coordinates. This relation provides a means for communicating pressure effects between layers.

The equations for mass, tracers, and momentum are supplemented with a nonlinear equation of state that relates density, salinity, pressure, and temperature (or potential temperature). Discussions of the equation of state are given, for example, by Gill (1982) and Griffies (2004).

In the derivations in Section 7, the horizontal coordinates are denoted by x_1 and x_2. These quantities could be latitude and longitude or some other suitable parameters. Increments Δx_1 and Δx_2 in these coordinates correspond to spatial displacements $m_1 \Delta x_1$ and $m_2 \Delta x_2$, respectively, where m_1 and m_2 are metric coefficients developed in Section 7. Also appearing in that section is a quantity $G = m_1 m_2$, which relates 'area' in the space of the parameters x_1 and x_2 to physical area on the surface of the rotating spheroid. The special case of Cartesian coordinates on a tangent plane is then obtained by setting $m_1 = m_2 = G = 1$. This case is assumed in the present section, for the sake of simplicity.

4.2. Shallow water equations

For the present discussion it is also useful to refer to the shallow water equations, which describe the motions of a hydrostatic fluid of constant density (*e.g.*, Pedlosky (1987)). If the system described in Section 7 is

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

discretized in the vertical by integrating over each of a set of coordinate layers, then the result can be regarded as a stack of shallow water models, with mechanisms for transferring mass, momentum, and pressure effects between adjacent layers.

In order to describe the shallow water system for a single-layer fluid, let x and y denote the horizontal coordinates; $u(x, y, t)$ and $v(x, y, t)$ denote the x- and y-components of velocity, respectively, defined relative to a rotating reference frame; $z_{\text{top}}(x, y, t)$ denote the elevation of the free surface at the top of the fluid; and $h(x, y, t) = z_{\text{top}}(x, y, t) - z_{\text{bot}}(x, y)$ denote the thickness of the fluid layer, where $z_{\text{bot}}(x, y)$ denotes the elevation of the bottom topography. Then

$$\begin{aligned} \frac{Du}{Dt} - fv &= -g\frac{\partial z_{\text{top}}}{\partial x}, \\ \frac{Dv}{Dt} + fu &= -g\frac{\partial z_{\text{top}}}{\partial y}, \\ \frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(hu) + \frac{\partial}{\partial y}(hv) &= 0, \end{aligned} \tag{4.1}$$

where g is the magnitude of the acceleration due to gravity. The material derivative D/Dt is defined by $D/Dt = \partial/\partial t + u\partial/\partial x + v\partial/\partial y$, and it represents the time derivative as seen by an observer moving with the fluid. Since the velocity components u and v are defined relative to a rotating reference frame, the material derivatives Du/Dt and Dv/Dt do not represent acceleration relative to an inertial frame, and the Coriolis terms $-fv$ and fu are included to yield such an acceleration. In the case of a rotating spheroid, the Coriolis parameter f is given by equation (7.94), $f = 2\Omega \sin\theta$, where Ω is the angular rate of rotation and θ is the latitude.

The first two equations in (4.1) describe the conservation of momentum, and in particular they state that the lateral acceleration of the fluid is due to variations in the free-surface elevation, which correspond to lateral variations of pressure within the fluid. The third equation in (4.1) describes the conservation of mass, and for this constant-density fluid the equation relates time variations in the layer thickness h to lateral variations in the volume flux (hu, hv).

An alternate formulation of the pressure forcing is given by the following. Define the Montgomery potential M by

$$M(x, y, z, t) = \alpha p(x, y, z, t) + gz, \tag{4.2}$$

where α is the specific volume (reciprocal of density ρ) of the fluid, and $p(x, y, z, t)$ is the pressure (Montgomery 1937). The hydrostatic condition $\partial p/\partial z = -\rho g$ (see (7.8)) implies that M is independent of z in a layer of constant density. Assuming that the pressure at the top of the fluid has a constant value, such as atmospheric pressure, then $\partial M/\partial x = g\partial z_{\text{top}}/\partial x$

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

and $\partial M/\partial y = g\partial z_{\text{top}}/\partial y$. The shallow water system (4.1) can then be written as

$$\begin{aligned} \frac{D\mathbf{u}}{Dt} + f\mathbf{u}^{\perp} &= -\nabla M, \\ \frac{\partial h}{\partial t} + \nabla \cdot \big(h\mathbf{u}\big) &= 0, \end{aligned} \tag{4.3}$$

where $\mathbf{u} = (u, v)$, $\mathbf{u}^{\perp} = (-v, u)$, and $\nabla = (\partial/\partial x, \partial/\partial y)$. In addition, $-\nabla M = -\frac{1}{\rho}\nabla p$, which is a common form for representing the pressure term in fluid dynamics.

If one were to model a stratified fluid as a stack of constant-density immiscible shallow water models, then the equations (4.3) could be applied in each layer. At an interface between layers, the pressure p and elevation z are continuous, so equation (4.2) implies $\Delta M = p\Delta\alpha$ at the interface. Here, ΔM and $\Delta\alpha$ are the jumps in M and α, respectively, across that interface. The relation $\Delta M = p\Delta\alpha$ is an analogue of the discretization (7.105) of the hydrostatic condition in generalized coordinates.

If a continuously stratified fluid is modelled with a generalized vertical coordinate, then the Montgomery potential enables gradients along slanting coordinate surfaces to represent pressure forcing that is truly horizontal. This idea is developed in Section 7.11.

4.3. Geostrophic balance

A striking feature of large-scale flows in the ocean and atmosphere is the approximate balance between the Coriolis and pressure terms, which causes the fluid to flow along curves of constant pressure. Here this 'geostrophic balance' is derived, in the context of the shallow water equations, by using an approximate scale analysis.

Let T, L, and U denote numbers that represent typical scales for time, horizontal distance, and horizontal velocity, respectively. A representative scale for the nonlinear terms in the momentum equation in (4.3) (*i.e.*, $uu_x + vu_y$ and $uv_x + vv_y$) is then UU/L, and a representative value for the Coriolis term is fU. The ratio of these scales is the Rossby number $Ro = U/(fL)$.

As noted in Section 7.15, the Coriolis parameter is zero at the equator, and at the middle and high latitudes of the Earth this parameter is on the order of 10^{-4} sec^{-1}. A velocity scale $U = 1$ m/sec represents a strong current, and a length scale $L = 100$ km $= 10^5$ m represents the approximate width of a western boundary current. With these scales, and with a Coriolis parameter equal to 10^{-4} sec^{-1}, the Rossby number is 0.1. However, for basin-scale flows, U is typically smaller and L is larger, so the Rossby number is much smaller in that case. Similarly, the time derivative in the momentum equation in (4.1) has the scale U/T, and the ratio of this scale to the scale of the Coriolis term is $1/(fT)$. If the time, space, and length

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

scales are related by $U = L/T$, then the parameter $1/(fT)$ is equal to the Rossby number. If a process has a time scale of a day (86400 seconds), then $1/(fT) \approx 0.1$ for the middle and high latitudes; for a time scale of 100 days, $1/(fT) \approx 10^{-3}$. It then follows that, for large-scale flows, the time derivative and nonlinear terms in the momentum equation are much smaller than the Coriolis term, except near the equator.

A remaining consideration is the size of a possible diffusion term. In geophysical fluid dynamics it is common practice to include some sort of diffusion term in the momentum equation to represent the large-scale effect of subgrid-scale mixing. For the system (4.3), such a term could take the form $A_H \nabla^2 \mathbf{u}$, where ∇^2 denotes the Laplace operator and A_H is an eddy viscosity coefficient. (Also see Section 7.12.) A scale for the diffusion term is then $A_H U/L^2$, and the ratio of the scale of the diffusion term to the scale of the Coriolis term is given by the Ekman number, $E = A_H/(fL^2)$. A physically appropriate choice of A_H involves considerable uncertainty; for example, values in the range 10 m^2/sec to 10^4 m^2/sec have been proposed (Pedlosky 1987). A horizontal length scale $L = 100$ km $= 10^5$ m and an eddy viscosity $A_H = 10^4$ m^2/sec leads to an Ekman number on the order of 10^{-2}. Larger length scales and/or smaller viscosities produce smaller Ekman numbers. For large-scale flows, the diffusion term would therefore be much smaller than the Coriolis term, except near the equator.

The only remaining term in the momentum equation is the pressure term $-g\nabla z_{\text{top}} = -\nabla M$. Since the other terms are much smaller than the Coriolis term, the pressure term and the Coriolis term must be approximately equal, *i.e.*, the flow is approximately in a geostrophic balance. If this balance is exact, then $f\mathbf{u}^{\perp} = -\nabla M$, or

$$\begin{aligned} -fv &= -g\frac{\partial z_{\text{top}}}{\partial x}, \\ fu &= -g\frac{\partial z_{\text{top}}}{\partial y}. \end{aligned} \tag{4.4}$$

In this case the horizontal velocity vector $\mathbf{u} = (u, v)$ is orthogonal to ∇z_{top}. If the fluid is viewed from the top, then the fluid is seen to flow along contours of constant free-surface elevation z_{top}, which coincide with contours of constant pressure within the fluid. The geostrophic balance is also seen in weather maps, which show upper-level winds blowing along isobars. In general, a geophysical fluid flow is not in a state of exact balance, and the evolution of the flow is driven by small departures from a balanced state.

The geostrophic balance may seem counter-intuitive, as everyday experience suggests that fluid should flow from a region of high pressure to a region of low pressure, not along curves that separate the two regions. However, the key point is that the flows considered here are dominated by the effects of rotation. For a simple analogy, try walking in a straight line on a rotating

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

merry-go-round; this 'straight' path is curved relative to the ground, and a force from the side is required to keep you on that straight/curved path.

4.4. Potential vorticity and Rossby waves

The vorticity of a fluid flow is the curl of the velocity field. If a fluid is stationary relative to a rotating spheroid, then the vorticity of the fluid relative to an inertial reference frame is $2\mathbf{\Omega}$, where $\Omega = |\mathbf{\Omega}|$ is the angular rate of rotation, the direction of the vector $\mathbf{\Omega}$ aligns with the axis of rotation, and the direction of $\mathbf{\Omega}$ and the direction of rotation are related by the right-hand rule. If a fluid is in motion relative to the spheroid, then the relative velocity gives rise to a vorticity relative to the rotating reference frame. In the case of the shallow water equations (4.1) and (4.3) on a tangent plane, the relative vorticity is $\zeta\mathbf{k}$, where $\mathbf{k}$ is the unit vector in the (local) upward direction, and

$$\zeta(x, y, t) = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}.$$

However, as noted in Section 7.15, the Coriolis parameter f is the local vertical component of the planetary vorticity $2\mathbf{\Omega}$. Therefore $\zeta + f$ is the local vertical component of the absolute (relative plus planetary) vorticity. In the present subsection we develop a conservation property associated with this quantity.

For the sake of simplicity, this property is developed in the context of the shallow water system (4.1), (4.3). The momentum equations in that system can be written in the form

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left[\frac{1}{2}\big(u^2 + v^2\big)\right] - (\zeta + f)v &= -\frac{\partial M}{\partial x}, \\ \frac{\partial v}{\partial t} + \frac{\partial}{\partial y}\left[\frac{1}{2}\big(u^2 + v^2\big)\right] + (\zeta + f)u &= -\frac{\partial M}{\partial y}; \end{aligned} \tag{4.5}$$

that is, the nonlinear terms are written in terms of the relative vorticity ζ and the kinetic energy per unit mass, $\frac{1}{2}(u^2 + v^2)$. Now compute the x-derivative of the second equation minus the y-derivative of the first equation. The result is

$$\frac{\partial \zeta}{\partial t} + u\frac{\partial}{\partial x}\big(\zeta + f\big) + v\frac{\partial}{\partial y}\big(\zeta + f\big) + \big(\zeta + f\big)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0.$$

A comparison with the mass equation in (4.1), coupled with the fact that f is independent of t, yields

$$\frac{D}{Dt}\big(\zeta + f\big) - \big(\zeta + f\big)\frac{1}{h}\frac{Dh}{Dt} = 0,$$

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

or

$$\frac{D}{Dt}\left(\frac{\zeta + f}{h}\right) = 0. \tag{4.6}$$

The quantity $(\zeta + f)/h$ is known as the potential vorticity, and it is the ratio of the absolute vorticity to the layer thickness.

Equation (4.6) states that the potential vorticity is constant, when seen by an observer moving with the fluid. For example, if a column of water moves laterally and experiences a change in h, then the absolute vorticity $\zeta + f$ changes, due to vortex stretching.

Of particular interest is the effect of variations in the Coriolis parameter on the relative vorticity. Suppose that ζ is initially zero, and consider a region where the bottom topography z_{bot} is constant, so that the layer thickness h is nearly constant. If a water column moves northward, for example, then the resulting increase in f causes a decrease in ζ, which corresponds to a clockwise rotation when seen from above. This local rotation causes water columns to the east and west to move southward and northward, respectively. The resulting changes in relative vorticity for those columns tend to force the first column to return towards its original position. The resulting wave motion is known as a Rossby wave, and for this wave the restoring mechanism is based on vorticity instead of gravity. Variations in bottom topography have an analogous effect, and a similar discussion of topographic Rossby waves is given by Pedlosky (1987). The existence of Rossby waves depends fundamentally on variations in bottom topography and/or the variation of the Coriolis parameter with latitude.

Detailed analyses of the dynamics of Rossby waves are given in several references, *e.g.*, Gill (1982) and Pedlosky (1987, 2003). These waves propagate with low frequencies, and the resulting time derivatives in the system (4.1) are so small that a Rossby wave is in an approximate geostrophic balance; as the bumps and depressions in the free surface propagate, the velocity field is tangent to contours of constant elevation.

The phase velocities of Rossby waves are always westward. Relatively long Rossby waves are nearly nondispersive, but relatively short Rossby waves are strongly dispersive with eastward group velocity. Long Rossby waves then propagate westward from the eastern boundaries of ocean basins, and short Rossby waves propagate eastward (in the sense of group velocity) from western boundaries. The latter propagate more slowly, and the resulting accumulation of short waves near the western boundaries helps to explain the east–west asymmetry in ocean circulation patterns (Gill 1982).

4.5. External and internal modes

A stratified fluid can admit both external and internal wave motions. External waves and internal waves propagate at very different speeds, and the

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

separation of time scales has a major impact on the process of solving the governing equations numerically. Internal waves cannot be modelled with the shallow water equations, since those equations apply to a fluid of constant density, so instead the present discussion refers to a simple linearized version of the governing equations that are developed in Section 7 for a stratified fluid. The following analysis is similar to one given by Higdon and Bennett (1996).

For the sake of simplicity, assume that the vertical coordinate s is the specific volume α (reciprocal of density) over the range $\alpha_{\rm bot} \le \alpha \le \alpha_{\rm top}$ and that $\dot{s} = \dot{\alpha} = 0$, *i.e.*, α is constant in time for each fluid parcel; an analysis of external and internal modes with a more general equation of state is given by Dukowicz (2006). In order to obtain a linearized model, assume that the bottom of the fluid domain is level and that the flow is a small perturbation of a static state having a level free surface at the top of the fluid and level surfaces of constant density within the fluid. Let $\tilde{p}(\alpha)$ denote the pressure in the fluid at the stationary state, and let $\tilde{M}(\alpha)$ denote the Montgomery potential at that state. Then let $p(x, y, \alpha, t)$ and $M(x, y, \alpha, t)$ denote the perturbations in those quantities from the static values. The pressure and Montgomery potential at an arbitrary state are then $\tilde{p}(\alpha) + p(x, y, \alpha, t)$ and $\tilde{M}(\alpha) + M(x, y, \alpha, t)$, respectively, with the present notation. Also assume that the viscosity and applied stresses are zero. Under these assumptions the governing equations from Section 7 for a continuously stratified fluid simplify to

$$\begin{aligned} u_t - fv &= -M_x, \\ v_t + fu &= -M_y, \\ p_{\alpha t} + \tilde{p}_\alpha(\alpha)(u_x + v_y) &= 0, \\ M_\alpha &= p. \end{aligned} \tag{4.7}$$

The subscripts denote partial derivatives. The first two equations in (4.7) are the momentum equations (7.92) as applied to the present case, the third equation is the equation (7.32) for conservation of mass, and the fourth equation is the hydrostatic condition (7.65).

Upper and lower boundary conditions for the system (4.7) can be formulated as follows. At the top of the fluid the pressure is equal to the atmospheric pressure, which is assumed here to be a constant. The perturbation in pressure is then zero, so in the present notation

$$M_\alpha = p = 0 \qquad \text{if } \alpha = \alpha_{\rm top}. \tag{4.8}$$

At the bottom of the fluid the perturbation in elevation is zero, so the perturbation in Montgomery potential satisfies

$$M = \alpha p = \alpha M_\alpha \qquad \text{if } \alpha = \alpha_{\rm bot}. \tag{4.9}$$

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Special solutions of the system (4.7)–(4.9) can be constructed by separating the vertical dependence from the time and horizontal dependences. Let $u(x,y,\alpha,t) = \hat{u}(x,y,t)\phi(\alpha)$, $v(x,y,\alpha,t) = \hat{v}(x,y,t)\phi(\alpha)$, and $M(x,y,\alpha,t) = \hat{M}(x,y,t)\phi(\alpha)$. Then $p(x,y,\alpha,t) = \hat{M}(x,y,t)\phi_\alpha(\alpha)$, and the system (4.7) becomes

$$\begin{aligned} \hat{u}_t - f\hat{v} &= -\hat{M}_x, \\ \hat{v}_t + f\hat{u} &= -\hat{M}_y, \\ \tilde{M}_t + (1/\lambda)(\hat{u}_x + \hat{v}_y) &= 0, \end{aligned} \tag{4.10}$$

where λ and ϕ satisfy

$$\begin{aligned} \phi_{\alpha\alpha} &= \lambda\tilde{p}_\alpha(\alpha)\phi(\alpha) \quad && \text{if } \alpha_{\text{bot}} < \alpha < \alpha_{\text{top}}, \\ \phi_\alpha &= 0 && \text{if } \alpha = \alpha_{\text{top}}, \\ \phi &= \alpha\phi_\alpha && \text{if } \alpha = \alpha_{\text{bot}}. \end{aligned} \tag{4.11}$$

The Sturm–Liouville problem (4.11) admits a countable set of eigenvalues and a complete set of eigenfunctions. The eigenvalues are all positive, since $\tilde{p}_\alpha < 0$, and they can be denoted by $0 < \lambda_0 < \lambda_1 < \lambda_2 < \dots$. In the general solution of the system (4.7), u can be represented as

$$u(x,y,\alpha,t) = \sum_{j=1}^{\infty} \hat{u}^{(j)}(x,y,t)\phi^{(j)}(\alpha), \tag{4.12}$$

where $\phi^{(j)}$ is an eigenfunction corresponding to eigenvalue λ_j, and $\hat{u}^{(j)}(x,y,t)$ is obtained from the reduced system (4.10) with $\lambda = \lambda_j$. Orthogonality of the eigenfunctions with respect to the weight function $\tilde{p}_\alpha$ implies

$$\hat{u}^{(j)}(x,y,t) = \frac{\int_{\alpha_{\text{bot}}}^{\alpha_{\text{top}}} u(x,y,\alpha,t)\ \phi^{(j)}(\alpha)\ \tilde{p}_\alpha(\alpha)\,\mathrm{d}\alpha}{\int_{\alpha_{\text{bot}}}^{\alpha_{\text{top}}} \phi^{(j)}(\alpha)^2\ \tilde{p}_\alpha(\alpha)\,\mathrm{d}\alpha}. \tag{4.13}$$

Similar expansions apply to v, M, and $p = M_\alpha$.

The reduced system (4.10), which describes the dependence with respect to (x,y,t), has the form of the linearization of the shallow water equations (4.1). In the case where the Coriolis parameter f is constant, this system admits gravity wave solutions with speed $c = \sqrt{1/\lambda}$. In the case where f varies with latitude, the gravity waves are still admitted, and in addition Rossby waves can also be present (Gill 1982, Pedlosky 1987).

The vertical structures of the modal solutions are determined by the eigenfunctions $\phi^{(j)}$, which can be visualized as indicated in Figure 4.1. Assume that the α-axis is vertical and that the ϕ-axis is horizontal. The upper boundary condition $\phi_\alpha(\alpha_{\text{top}}) = 0$ states that the tangent line to the graph of ϕ is vertical when $\alpha = \alpha_{\text{top}}$. The lower boundary condition

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

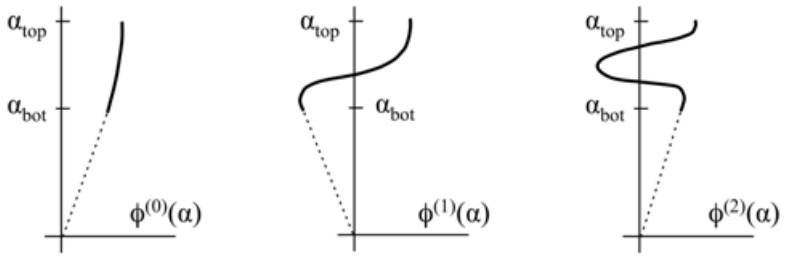


Figure 4.1. Graphs of eigenfunctions which give the vertical dependences in modal solutions of the linearized system (4.7). The eigenfunction $\phi^{(0)}$ corresponds to the external mode, and $\phi^{(1)}$ and $\phi^{(2)}$ correspond to the first two internal modes. The dotted lines illustrate the lower boundary condition $\phi_\alpha(\alpha_{\rm bot}) = \phi(\alpha_{\rm bot})/\alpha_{\rm bot}$. In these plots, the α-axes are not to scale, as typically $(\alpha_{\rm top} - \alpha_{\rm bot})/\alpha_{\rm bot} \approx 10^{-2}$. On the interval $\alpha_{\rm bot} \le \alpha \le \alpha_{\rm top}$, the relative variation in $\phi^{(0)}$ is bounded by the relative variation in α.

$\phi_\alpha(\alpha_{\rm bot}) = \phi(\alpha_{\rm bot})/\alpha_{\rm bot}$ states that the tangent line to the graph of ϕ at $\alpha = \alpha_{\rm bot}$ passes through the origin. Without loss of generality, assume $\phi(\alpha_{\rm top}) = 1$, and consider the variation of $\phi(\alpha)$ as α varies from $\alpha_{\rm top}$ downward towards $\alpha_{\rm bot}$. If $\lambda > 0$ then $\phi_{\alpha\alpha}$ and ϕ have opposite signs, since $\tilde{p}_\alpha < 0$, so the graph of ϕ always bends so as to return toward the α-axis. For a given value of $\phi(\alpha)$, a larger value of λ yields a larger value for the concavity $\phi_{\alpha\alpha}$.

In the case of the smallest eigenvalue λ_0, the corresponding eigenfunction $\phi^{(0)}$ is monotone and maintains constant sign, and its derivative is largest when $\alpha = \alpha_{\rm bot}$. This condition on the derivative, coupled with the mean value theorem, implies

$$\left|\phi^{(0)}(\alpha_{\rm top}) - \phi^{(0)}(\alpha_{\rm bot})\right| < \frac{\phi^{(0)}(\alpha_{\rm bot})}{\alpha_{\rm bot}} \left|\alpha_{\rm top} - \alpha_{\rm bot}\right|. \tag{4.14}$$

It follows that the relative variation of the eigenfunction $\phi^{(0)}$ over the interval $\alpha_{\rm bot} \le \alpha \le \alpha_{\rm top}$ is less than the relative variation of α over that interval. In the ocean, α typically varies by at most a few per cent over the vertical extent of a water column. With such behaviour of α, the eigenfunction $\phi^{(0)}$ is nearly constant.

For the eigenvalue λ_1, the corresponding eigenfunction $\phi^{(1)}$ changes sign once; in general, the eigenfunction $\phi^{(j)}$, corresponding to the eigenvalue λ_j, changes sign j times. The orthogonality of eigenfunctions with respect

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

to $\tilde{p}_\alpha$, coupled with the observation that $\phi^{(0)} \approx 1$, implies

$$\int_{\alpha_{\rm bot}}^{\alpha_{\rm top}} \phi^{(j)}(\alpha)\ \tilde{p}_\alpha(\alpha)\,\mathrm{d}\alpha \approx \int_{\alpha_{\rm bot}}^{\alpha_{\rm top}} \phi^{(j)}(\alpha)\ \phi^{(0)}(\alpha)\ \tilde{p}_\alpha(\alpha)\,\mathrm{d}\alpha = 0 \qquad (4.15)$$

for $j \geq 1$. The mass-weighted vertical average of $\phi^{(j)}$ is thus nearly zero, if $j \geq 1$.

A modal solution for $j = 0$ is an external mode, whereas the modes for $j \geq 1$ are internal modes. To see this, partition the interval $\alpha_{\rm bot} \leq \alpha \leq \alpha_{\rm top}$ into layers separated by surfaces of constant α. For the case $j = 0$, the horizontal velocity divergence $u_x + v_y = \phi^{(0)}(\alpha)(\hat{u}_x + \hat{v}_y)$ is nearly independent of depth. If the third equation in the linearized system (4.7) is integrated with respect to α over one of those layers, the result is that $\partial/\partial t(\Delta p/\Delta\tilde{p})$ is approximately independent of depth. Here, Δp and $\Delta\tilde{p}$ represent the vertical differences of p and $\tilde{p}$, respectively, across the layer. It follows that if $j = 0$ then all layers are thickened or thinned by approximately the same proportion, for given (x, y, t). The behaviour of the free surface at the top of the fluid then indicates the nature of the wave motion throughout the interior.

On the other hand, if $j \geq 1$ then the horizontal divergence $\phi^{(j)}(\alpha)(\hat{u}_x+\hat{v}_y)$ varies in sign with depth, so some layers are thickened and some are thinned, for given (x, y, t). The wave motion is then manifested by undulations of surfaces of constant density within the fluid. Furthermore, a vertical integration of the third equation in the linearized system (4.7) over the entire depth of the fluid, coupled with the zero-integral condition (4.15), implies that the elevation of the free surface remains nearly unperturbed if $j \geq 1$.

Estimates of wave speeds can be obtained as follows. Vertical integration of the equation $\phi^{(0)}_{\alpha\alpha} = \lambda_0\tilde{p}_\alpha(\alpha)\phi^{(0)}(\alpha)$ over the interval $\alpha_{\rm bot} \leq \alpha \leq \alpha_{\rm top}$, combined with the boundary conditions in (4.11) and the condition $\phi^{(0)} \approx 1$, implies that the gravity-wave speed for the external mode is $c_0 = \sqrt{1/\lambda_0} \approx \sqrt{\alpha_{\rm bot}\big(\tilde{p}_{\rm bot} - \tilde{p}_{\rm top}\big)}$. However, the hydrostatic condition $\partial p/\partial z = -\rho g$ (see (7.8)) implies $\alpha_{\rm bot}\big(\tilde{p}_{\rm bot} - \tilde{p}_{\rm top}\big) \approx gH$, where H is the total depth of the fluid. The speed of gravity waves as seen in the linearized shallow water equations is $\sqrt{gH}$, which is approximately the speed seen in the external mode. If, for example, $H = 4000$ m, then $c_0 \approx \sqrt{gH} \approx 200$ m/sec. On the other hand, the speeds of internal gravity waves in the ocean are typically at most a few metres per second (Gill 1982).

The external mode is essentially a two-dimensional phenomenon, in the sense that it can be approximately described by functions of (x, y, t), and the dynamics of this mode are very similar to those seen in the linearized shallow water equations. Furthermore, the velocity field for the external

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

mode is obtained by the projection (4.13) for the case $j = 0$. Since $\phi^{(0)} \approx 1$, equation (4.13) implies

$$\hat{u}^{(0)}(x, y, t) \approx \frac{1}{\tilde{p}_{\text{bot}} - \tilde{p}_{\text{top}}} \int_{\alpha_{\text{bot}}}^{\alpha_{\text{top}}} u(x, y, \alpha, t) \, \big(-\tilde{p}_\alpha(\alpha)\big) \, \mathrm{d}\alpha, \tag{4.16}$$

and an analogous equation holds for $\hat{v}^{(0)}$. The velocity field in the external mode is thus approximately given by the mass-weighted vertical average of the horizontal velocity field for the three-dimensional system. These observations form the basis for techniques for isolating the fast dynamics into a relatively simple two-dimensional subsystem, as described in Section 5.1.

The preceding analysis assumes that the fluid is continuously stratified. One can also consider a vertical discretization of the system (4.7), or equivalently, a multi-layer stack of shallow water models. In that event, the system admits an external mode and finitely many internal modes. For example, Higdon (2005) developed a test problem involving linearized flow in a two-layer fluid in a straight channel having a level bottom, with the Coriolis parameter varying linearly in the cross-channel direction y. In this case the system admits both external and internal waves, and within each class are Rossby waves and gravity waves. The gravity waves include Poincaré waves and Kelvin waves; the former are approximately sinusoidal in y, whereas the latter decay exponentially from the one or another of the solid boundaries. The domain is discretized in space, and after a Fourier transform in time and in the along-channel direction x, the time frequencies and y-dependences are obtained via numerical solution of a matrix eigenvalue problem. The resulting modal solutions can then be used to test numerical methods for the time splittings that are discussed in Section 5.1. In particular, the external and internal modal solutions make it possible to test the fast and slow subsystems independently.

5. Time discretization

The present section outlines some aspects of time discretization for numerical models of ocean circulation. Included is a description of the process of splitting the fast and slow dynamics into separate subsystems and a discussion of some time-stepping methods.

5.1. Barotropic–baroclinic splitting

As noted in Section 2.3, the fastest motions in a numerical model of ocean circulation are typically the external gravity waves, and these travel much more rapidly than any of the other motions that are present in the system. If an explicit time discretization is used to solve a system of partial differential

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

equations numerically, then the time step Δt is limited in terms of the fastest motions that can be present. However, in the present case the fastest motions are essentially two-dimensional, as demonstrated in Section 4.5. The numerical solution of a complex three-dimensional system is therefore severely constrained by a special set of two-dimensional motions.

It has therefore become common practice to split the dynamics of the system into two subsystems, one a relatively simple two-dimensional system that represents the fast motions, and the other a relatively complex three-dimensional system that represents the remaining (slow) processes. The latter system is solved explicitly with a relatively long time step that is appropriate for resolving the slow motions. The fast system is either solved implicitly with the same time step or explicitly with short steps. The resulting algorithm is much more efficient than solving the entire unsplit system explicitly with short steps.

Such a splitting is traditionally referred to as a barotropic–baroclinic splitting. In classical fluid dynamics, a flow is labelled barotropic if the density is a function of pressure only (Pedlosky 1987), and it is labelled baroclinic otherwise. In the case of a stratified fluid, a barotropic state implies that the surfaces of constant density coincide with the surfaces of constant pressure. If a fluid exhibits only an external wave, then all fluid layers thicken or thin by approximately the same proportion, at a given horizontal location and time (see Figure 2.2). The surfaces of constant density within the fluid thus have essentially the same shape, with the amplitudes of variation increasing from the bottom of the fluid domain to the top. Each such surface remains approximately the same distance below the free surface at the top of the fluid, at all horizontal locations and times, so the flow is approximately barotropic. On the other hand, this is not the case with an internal wave, and in that case the fluid is in a baroclinic state. In the following, the fast and slow subsystems are referred to as the barotropic and baroclinic equations, respectively.

The early versions of the z-coordinate Bryan–Cox class of ocean models used a rigid-lid boundary condition $w = 0$ at the top of the fluid, where w is the vertical component of fluid velocity. This assumption has the effect of replacing the fast speed of external gravity waves with an infinite speed, and a two-dimensional elliptic partial differential equation for this mode must then be solved at each (long) time step. More recent efforts have replaced the rigid lid with a free-surface boundary condition, which restores the finite speed of propagation for the barotropic subsystem. This system is obtained by a vertical integration of the three-dimensional momentum and mass equations, and it has a structure similar to that of the shallow water equations. Killworth, Stainforth, Webb and Paterson (1991) solved the barotropic equations explicitly with short steps, whereas Dukowicz and Smith (1994) solved the barotropic equations implicitly with same

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

(long) time step as for the baroclinic equations. A hybrid-coordinate update of the latter model solves the barotropic equations explicitly with short steps (Dukowicz 2004). The current version of the Modular Ocean Model also solves the barotropic equations explicitly with short steps (Griffies *et al.* 2004).

A barotropic–baroclinic splitting for isopycnic models was developed by Bleck and Smith (1990). With that splitting, the barotropic velocity $\bar{\mathbf{u}}(x, y, t)$ is the mass-weighted vertical average of the horizontal velocity $\mathbf{u}(x, y, s, t) = \big(u(x, y, s, t), v(x, y, s, t)\big)$ appearing in the full three-dimensional system, in analogy to the approximate projection onto the external mode given in (4.16). A baroclinic velocity is then given by $\mathbf{u}'(x, y, s, t) = \mathbf{u}(x, y, s, t) - \bar{\mathbf{u}}(x, y, t)$, so that $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}'$. Here, the mass-weighted vertical averages are the same as those used in Section 7.16, except that the averages in that section involve a single coordinate layer, whereas in the present case the averages are taken over the entire depth of the fluid. The splitting of the mass field is based on the idea that an external wave causes all layers to thicken or thin by approximately the same proportion. In particular, the pressure is represented as $p(x, y, s, t) = \big(1 + \eta(x, y, t)\big)p'(x, y, s, t)$, where p' represents the baroclinic component of the pressure, and η is a barotropic mass variable that represents the relative thickening of coordinate layers.

With this formulation, the splittings of the velocity and mass fields are inexact, even in the linearized case where one can discuss decompositions of the solution into external and internal modes. In addition, as formulated by Bleck and Smith (1990), the pressure forcing in the barotropic momentum equation is the same as in the shallow water equations for a fluid of constant density, and this introduces a further approximation that contributes to the inexactness of the splitting. This inexactness implies that the baroclinic equations can represent a (small) portion of the fast external motions, even though these equations are intended to represent the slow motions. If the baroclinic equations are solved explicitly with a long time step that is appropriate for resolving the slow motions, then the Courant–Friedrichs–Lewy condition is violated, strictly speaking. This CFL violation raises the possibility of numerical instability.

Higdon and Bennett (1996) showed that this instability can in fact occur with the splitting of Bleck and Smith (1990), as applied to the linearized case. However, Higdon and de Szoeke (1997) subsequently showed that the instability can be removed by modifying the barotropic momentum equation so as to replace the pressure-gradient term from the shallow water equations with the mass-weighted vertical average of $\nabla M = (\partial M/\partial x, \partial M/\partial y)$. In this case the splitting is still inexact, due to the inexactness in splitting the velocity and mass fields. However, the amount of 'fast' energy in the 'slow' baroclinic equations is low enough to obtain stability, even in the face of the formal violation of the CFL condition. In other words, the splitting of the

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

external and internal modes does not need to be exact, but it does need to be sufficiently accurate.

The analysis of external and internal modes in Section 4.5 assumes that the bottom of the fluid domain is level. Without this assumption, the linearized system (4.7) is not separable, and the solutions do not decompose exactly into external and internal modes. In addition, realistic ocean circulation models include nonlinearity, and this provides another mechanism for interactions between fast and slow motions. Nonetheless, in the ocean modelling community it is common practice to split the dynamics of the governing equations based on the preceding ideas, even for nonlinear problems, and this approach has proved to be an effective method for gaining computational efficiency while maintaining numerical stability.

5.2. Leapfrog time-stepping

One time-stepping method that has been widely used in geophysical fluid dynamics is the leapfrog method. For example, this method is used to solve the baroclinic equations in many versions of the Bryan–Cox class of models (Griffies *et al.* 2000*a*) and in the Miami Isopycnic Coordinate Ocean Model (Bleck *et al.* 1992). The leapfrog method is a three-level scheme based on centred differencing about the middle time level, and for an equation of the form $u_t = F(t, u)$ it can be written in the form $u^{n+1} = u^{n-1} + 2\Delta t F(t_n, u^n)$. Here, u^n is an approximation to $u(t_n)$.

The leapfrog method is straightforward to implement, but it suffers the disadvantage of allowing a nonphysical computational mode consisting of sawtooth oscillations in t. For example, in the special case $F = 0$ the leapfrog method for $u_t = F(t, u)$ is simply $u^{n+1} = u^{n-1}$, and this scheme allows both constant solutions and sawtooth solutions of the form $u^n = c(-1)^n$. In a nonlinear model such a mode, once stimulated, can grow without bound unless measures are taken to suppress the oscillations. One widely used method for doing this is the Asselin filter (Asselin 1972), which uses a weighted average defined by $\bar{\phi}^n := \gamma\phi^{n+1} + (1-2\gamma)\phi^n + \gamma\bar{\phi}^{n-1}$. Here, γ is a positive constant, and $\bar{\phi}$ is a smoothed version of the quantity ϕ. After the quantity ϕ^{n+1} is computed, the filter is then used to smooth the solution at time t_n. The leapfrog method has order two, but using the Asselin filter reduces the method to first-order accuracy (Durran 1999). An alternative to this filter is given by Dukowicz and Smith (1994), who periodically interrupt the computation in order to average the solution between consecutive time levels. This is done for two consecutive pairs of time levels, and the leapfrog method is then restarted from the values obtained for the half-integer points.

If a filter succeeds in maintaining numerical stability but does not suppress the computational mode completely, then numerical difficulties can still persist. For example, Griffies (2004) describes some experience with

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

the Modular Ocean Model using leapfrog time-stepping and an Asselin filter. In those experiments, grid noise remained and was especially prevalent near the equator, and it caused a complete reversal of the directions of strong zonal (east–west) currents in that region from one time step to the next.

Additional difficulties arise if a computational mode is present and if the state of a model is altered when certain variables reach threshold values (Rainer Bleck, personal communcation). For example, in a hybrid-coordinate model, a coordinate layer is in an isopycnic state if and only if its target density lies within the range of densities that actually exist in the fluid. If this state is updated at each time step, and if the target density lies near the boundary of that range but the computed density field displays sawtooth oscillations in time, then the layer may be in one state at the even time steps and in another state at the odd steps.

5.3. Two-level time-stepping

The preceding difficulties can be avoided if a model employs a time-stepping scheme that uses only two time levels, as such a scheme cannot support the sawtooth mode described above. This matter has been a subject of recent attention.

For example, two-level methods are now available as options, along with the leapfrog method, in the Modular Ocean Model (Griffies 2004). The two-level method for the baroclinic equations is staggered in time, with velocity defined at integer time steps, and tracers, pressure, and density defined at half-integer steps. The staggered time grid enables second-order centred differencing without allowing the computational mode associated with the leapfrog method. In the case of the barotropic equations, the two-level method is a predictor–corrector method.

Hallberg (1997) developed a two-level predictor–corrector scheme for layered models. This method involves a prediction and correction of all of the dependent variables in the baroclinic and barotropic subsystems, and various weighted averages are used during the correction steps. In addition, Shchepetkin and McWilliams (2005) give an extensive survey of a variety of two-level and multi-level time-stepping methods.

A two-level non-staggered method for layered models was developed by Higdon (2002, 2005). After an initial forward-Euler prediction of the baroclinic velocity from time t_n to time t_{n+1}, all time discretizations for the baroclinic equations involve centred differencing and unweighted averaging about the middle of the time interval $[t_n, t_{n+1}]$. The Coriolis terms are implemented implicitly; if the horizontal velocity components u and v are defined at different points, as on the C-grid (Section 6.1), then the Coriolis terms are implemented with a simple iteration. In a linearized analysis

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

where the barotropic equations are assumed to be solved exactly in t, the algorithm is stable and essentially nondissipative. In some experiments with a nonlinear model and a standard dissipative advection scheme for mass and momentum, the algorithm behaves stably for very long times with zero explicit viscosity.

6. Spatial discretization and related issues

This section discusses some topics involving spatial discretization, the numerical simulation of advection, and the numerical solution of the momentum equations.

6.1. Horizontal grids: quadrilaterals

In existing numerical models of ocean circulation, it is commonplace to use finite difference and/or finite volume methods on quadrilateral spatial grids. This subsection focuses on such discretizations, and some alternatives are discussed in the following subsection.

A natural choice for a horizontal coordinate system for the Earth is the one based on latitude and longitude. However, this system admits a singularity at the North and South Poles, as the metric coefficient relating increments in longitude to increments in space tends to zero at those locations (see Section 7.18). This singularity is not an issue in the case of the South Pole, which lies on land, but it in the case of the North Pole the singularity lies within the fluid domain for a global model. If a spatial discretization is based on this coordinate system, then the convergence of grid lines causes substantial difficulties with explicit time-stepping methods, as the Courant–Friedrichs–Lewy condition implies that the time increment Δt must tend to zero as the spatial increment tends to zero.

This problem can be avoided with a coordinate system that does not place any singularities within the fluid domain. For example, Smith, Kortas and Meltz (1995) developed a class of orthogonal curvilinear grids with two coordinate poles that can be placed at arbitrary locations. In practice, the south coordinate pole can be located at the south geographical pole, and the latitude–longitude grid is deformed smoothly so that the north coordinate pole is located on a land mass.

With a coordinate system developed by Murray (1996), the coordinates are latitude and longitude south of a specified latitude. The remaining northern portion of the globe is covered with an orthogonal coordinate system that involves two poles that can be located on land. Models that have employed this grid include the Parallel Ocean Program, the Modular Ocean Model, and the Hybrid Coordinate Ocean Model. Figure 6.1 shows an example of such a grid.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

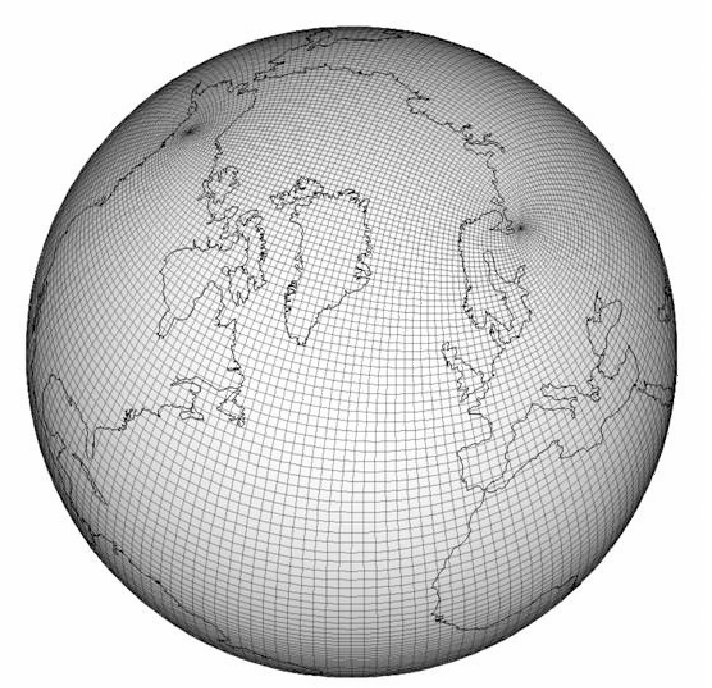


Figure 6.1. Tripolar grid. To the south of a specified latitude, the coordinate system uses latitude and longitude. To the north of that latitude, each line of constant longitude connects smoothly to a longitude line on the other side of the pair of coordinate poles. (Figure provided by Philip Jones, Los Alamos National Laboratory.)

A horizontal coordinate system developed by Rancic, Purser and Mesinger (1996) employs a conformally expanded spherical cube. With this formulation, a cube is inscribed within a sphere, and a rectangular mesh on the surface of the cube is then projected conformally onto the sphere. The resulting mesh on the sphere is orthogonal, except at the singular points corresponding to the corners of the cube. Adcroft, Campin, Hill and Marshall (2004) employ finite volume methods for the equations for mass and momentum on this grid, and they demonstrate that the singularities cause no difficulty in their setting. They also employ a re-scaling of the coordinates on the sphere so that the resulting grid is more uniform than the grid of Rancic *et al.* (1996). A non-orthogonal expanded spherical cube was used by Rossmanith (2006) to develop a finite volume method, based on the framework of LeVeque (2002), for hyperbolic systems on a sphere.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

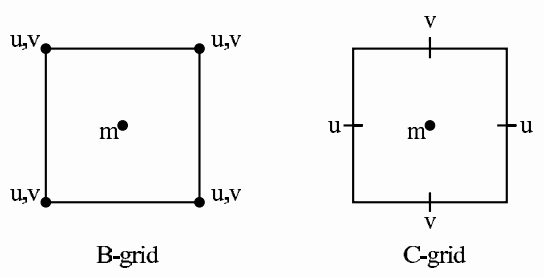


Figure 6.2. Arrangements of dependent variables on the B-grid and C-grid. The symbol m refers to mass variables such as density, layer thickness, and pressure; and u and v are the x- and y-components of velocity, respectively.

Relative to a given grid, there are several options for specifying the points where the various dependent variables are defined. For a rectangular grid, Winninghoff (1968) and Arakawa and Lamb (1977) analysed five such arrangements, which they labelled A–E. These grids are also illustrated and discussed in other references, such as Dukowicz (1995) and Haidvogel and Beckmann (1999). With the A-grid, all dependent variables are defined at the centres of grid cells. In the case of the B-grid, mass variables are defined at the centres of grid cells, and the components of velocity are defined at the corners. With the C-grid, the normal components of velocity are defined at the centres of the edges of the mass cells. That is, if u and v denote the x- and y-components of velocity, respectively, then the values of u are defined at the centres of edges corresponding to minimal and maximal x, and the values of v are defined at the centres of the edges corresponding to minimal and maximal y: see Figure 6.2.

Among the Arakawa grids, the B-grid and the C-grid are the most commonly used in ocean models. For example, the B-grid is used throughout the Bryan–Cox class of level models. Models that use the C-grid include the Miami Isopycnic Coordinate Ocean Model, the Hybrid Coordinate Ocean Model, and the Hallberg Isopycnal Model. A variation on the C-grid used by the MIT model (Adcroft, Hill and Marshall 1999) is described below.

One approach to comparing grid arrangements is to compare the accuracy of linear wave propagation as represented on those grids. For the case of gravity waves modelled by the linearized shallow water equations for a single-layer fluid, analyses of dispersion relations of numerical methods indicate that the C-grid is more accurate than the B-grid for relatively well-resolved waves, whereas the B-grid is more accurate in the case of relatively coarser resolution (Arakawa and Lamb 1977, Dukowicz 1995). Here, the resolution is defined in terms of the size of the grid spacing relative to the Rossby radius c/f, where c is the speed of gravity waves and f is the Coriolis

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

parameter, and it is assumed that the spatial discretizations are based on second-order centred finite differences and averages that are natural for the grid in question. Comparisons based on gravity waves thus suggest that the C-grid is preferable for higher-resolution models and the B-grid is preferable for coarser-resolution runs.

However, in a similar analysis of Rossby waves, Dukowicz (1995) concluded that for Rossby waves the B-grid is more accurate at higher resolution. This is the opposite case to gravity waves. For coarser resolution the B-grid is generally more accurate for Rossby waves, but this can depend on the location in wavenumber space. Rossby waves play a fundamental role in the development of large-scale circulation systems, whereas gravity waves are primarily involved in the details of adjustments between states of geostrophic balance (Gill 1982).

One practical problem with the B-grid is the possibility of a chessboard pattern in the mass fields (see, *e.g.*, Killworth *et al.* (1991).) On the B-grid, a given velocity point has four neighbouring mass cells, and a natural approximation to $\partial M/\partial x$ (equivalently, $\partial p/\partial x$) can be obtained by a difference in x of an average with respect to y. However, if the mass field displays a $+1/-1$ chessboard pattern then the computed pressure gradient is zero, and, more generally, a $+1/-1$ pattern can be added to an arbitrary mass field without affecting the computed pressure gradient. Grid noise of this nature can be introduced into a solution by grid-scale forcing, such as the forcing associated with variable bottom topography or variable coastlines. Spatial smoothing can be used to suppress this B-grid computational mode.

With the C-grid a given velocity point has only two neighbouring mass cells, and the above problem with the pressure gradient does not occur. However, there is an analogous problem with the Coriolis terms. The velocity components u and v are defined at different points, so some spatial averaging is required to implement the terms $-fv$ and fu in the u- and v-equations, respectively. Chessboard patterns in the velocity fields do not affect the computed values of those terms, and these patterns can persist once they are stimulated by grid-scale forcing.

Adcroft *et al.* (1999) state that the C-grid noise primarily occurs in low-resolution configurations, where again the resolution is defined in terms of the size of the grid spacing relative to the Rossby radius. For low-resolution simulations they obtain better results by implementing the Coriolis terms with a hybrid of the C-grid and the D-grid. With the D-grid, the *tangential* components of velocity are defined at the centres of the edges of mass cells; that is, the values of u and v are defined at the points where v and u, respectively, would be defined on the C-grid. In the method of Adcroft *et al.* (1999), the u-equation on the C-grid uses the value of $-fv$ from the D-grid. Equations are also solved in the D-grid, and for the u-equation on the D-grid, the value of $-fv$ is taken from the C-grid. The introduction of additional

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

equations and unknowns leads to the existence of nonphysical computational modes, which in this case take the form of spatially independent inertial oscillations. These modes are damped by using a weighted time average when implementing the Coriolis terms.

An alternate approach is given by Nechaev and Yaremchuk (2004). At a given u-point on the C-grid, consider implementing the Coriolis term $-fv$ by using the value of v at *one* of the neighbouring v-points; similarly, at that v-point, implement the Coriolis term fu by referring to the u-point just mentioned. This scheme, by itself, would give an uncentred implementation of the Coriolis terms. Instead, Nechaev and Yaremchuk use all of the four possible pairings of u-points and v-points and average the results. This procedure gives a centred approximation that avoids the C-grid noise described above. In their analysis, these authors use the leapfrog time discretization, except that the Coriolis terms are represented with a weighted time average involving three consecutive time levels.

6.2. Alternative spatial discretizations: spherical geodesic grids, spectral elements, and unstructured grids

One alternative to a quadrilateral grid is a spherical geodesic grid. The construction of such a grid is described and illustrated, for example, by Lipscomb and Ringler (2005). First, inscribe within the sphere a regular icosahedron, which has 20 faces that are equilateral triangles. Subdivide each triangle into four triangles by connecting the midpoints of the edges, and then project each of the resulting vertices onto the sphere. Continue this process, yielding finer and finer triangulations of the sphere at each step. For any such triangulation, regard each vertex as the centre of a grid cell; such a cell is defined by including all points on the sphere that are closer to that vertex than to any other vertex. The resulting cells are all hexagons, except that the cells corresponding to the 12 vertices of the original icosahedron are pentagons. The resulting mesh of hexagons and pentagons is nearly uniform, and it avoids the problems with coordinate poles described in the preceding subsection.

Ringler and Randall (2002) developed numerical methods for solving the shallow water equations on such a geodesic grid. In their formulation, scalar quantities such as mass variables, divergence of velocity, and the vertical component of the curl of velocity are defined at the centres of grid cells. Velocity vectors are defined at the vertices of the grid cells, and the cell averages of the divergence and curl are obtained by computing line integrals around the boundaries of the grid cells. The pressure gradient at a given vertex of a grid cell is obtained by linearly interpolating the pressure at the centres of the three neighbouring cells and then computing the gradient of the linear interpolant.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Another type of spatial discretization is used in the Spectral Element Ocean Model (Haidvogel and Beckmann 1999, Iskandarani, Haidvogel and Levin 2003). This model employs a finite element discretization using high-degree polynomials on hexahedral elements, which are cubes with curved surfaces. Numerical convergence is obtained by refining the grid and/or increasing the order of the approximating polynomials. The number of elements in the vertical direction is independent of the horizontal position, and the three-dimensional grid amounts to a stack of two-dimensional unstructured grids. The elements are constructed so as to conform to bottom topography and coastlines; given the structured nature of the grid in the vertical direction, the discrete algorithm is essentially a terrain-following σ-coordinate model.

The consideration of unstructured grids is a relatively recent development in ocean modelling. The review of Pain *et al.* (2005) describes some recent work involving three-dimensional unstructured grids, with an emphasis on finite element methods. Issues discussed include adapting the grid resolution to evolving flow conditions, parametrizing subgrid-scale processes in the presence of variable resolution, maintaining hydrostatic and geostrophic balance, advection schemes, iterative methods, and parallelization. This paper is the lead article in a special issue of *Ocean Modelling* devoted entirely to unstructured grids.

6.3. Advection

Among the processes represented in a numerical model of ocean circulation is the advection of quantities such as density, layer thickness, and tracers. The purpose of the present subsection is to describe some aspects of advection that are of particular interest in ocean modelling, namely, positive definiteness and compatibility, and to describe briefly some of the advection schemes that are used in this field. The literature on advection schemes is vast, and there is no point in trying to summarize the field here; a comprehensive development of the subject is given, for example, in the recent text by LeVeque (2002).

Section 7 of the present paper includes derivations of partial differential equations that describe the conservation of mass of the fluid and the conservation of tracers. In those derivations it is assumed that the horizontal coordinates are arbitrary curvilinear coordinates on a spheroid and that the generalized vertical coordinate s can satisfy $\dot{s} \neq 0$, where $\dot{s} = Ds/Dt$, so that s can change with time following fluid parcels. For the sake of simplicity in the present discussion, assume that $\dot{s} = 0$ and that the diffusion of tracers is zero. Also assume that the horizontal coordinates are Cartesian coordinates on a tangent plane, and consider the portion of the fluid bounded below and above by two coordinate surfaces $s = s_0$ and $s = s_1$,

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

respectively. Under these assumptions, equation (7.38) for the conservation of mass in such a coordinate layer is

$$\frac{\partial}{\partial t}(\Delta p) + \frac{\partial}{\partial x}(u\Delta p) + \frac{\partial}{\partial y}(v\Delta p) = \frac{\partial}{\partial t}(\Delta p) + \nabla \cdot \big(\mathbf{u}\Delta p\big) = 0, \tag{6.1}$$

and equation (7.107) for the conservation of a tracer in that layer is

$$\frac{\partial}{\partial t}(q\Delta p) + \nabla \cdot \big(\mathbf{u}(q\Delta p)\big) = 0. \tag{6.2}$$

Here, $\mathbf{u}(x, y, t) = \big(u(x, y, t), v(x, y, t)\big)$ is the mass-weighted vertical average of the horizontal velocity in the layer; $q(x, y, t)$ is the mass-weighted vertical average of the tracer concentration, which is the quantity of tracer per unit mass of fluid; and $\Delta p(x, y, t)$ is the vertical pressure difference across the layer. In Section 7, overbars are used to indicate mass-weighted vertical averages, but these are deleted in the present notation.

Because of the hydrostatic assumption, Δp is the weight per unit horizontal area in the layer, and it is thus g times the mass per unit horizontal area when viewed from above. The quantity $q\Delta p$ is then equal to g times the quantity of tracer per unit horizontal area. In the case where s is the geometrical height z, the hydrostatic condition $\partial p/\partial z = -\rho g$ (see (7.8)) implies that Δp is equal to g times the vertical integral of the density ρ in the coordinate layer. In the case of an isopycnic model, the quantity Δp can be regarded as a measure of layer thickness.

The tracers used in a model of ocean circulation typically include potential temperature and salinity, and depending on the usage of the model they may also include the concentrations of various chemical and/or biological species. In a model of the dynamics of sea ice, Lipscomb and Hunke (2004) transport 46 different scalar fields, including area fractions and thermodynamic variables associated with each of five different ice thickness categories. They also anticipate that the number of transported fields will increase as ice models become more realistic.

Numerical advection schemes can be evaluated according to standard requirements such as stability, accuracy, efficiency, and avoiding spurious oscillations. However, some additional requirements are of special interest in the context of ocean modelling.

In the case of an isopycnic model, the layer thickness Δp can tend to zero, as indicated by Figure 2.1. When equation (6.1) is solved numerically, it is then necessary to maintain nonnegative values of Δp in the computed solution; that is, solutions that are initially nonnegative must remain nonnegative in the absence of forcing. A numerical method with this property can be labelled positive definite.

An example of a positive definite method is MPDATA, the Multidimensional Positive Definite Advection Transport Algorithm (Smolarkiewicz and

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Margolin 1998). At each time step, this method employs multiple iterations. The first iteration uses the standard upwind (donor-cell) method, for which the mass flux at a cell edge is equal to the normal velocity at that edge times the density (such as Δp) in the upwind direction, *i.e.*, in the cell that is being drained. The upwind method is positive definite, assuming that a Courant–Friedrichs–Lewy condition is satisfied. Subsequent iterations are used to cancel the leading term in the truncation error, and these iterations are written in the form of upwind steps with non-physical pseudovelocities. The pseudovelocities are bounded by the physical velocities, so these subsequent iterations are also positive definite. The upwind method, by itself, is highly diffusive, and the subsequent iterations can be regarded as antidiffusive steps that cancel some of the numerical diffusion produced by the initial upwind step.

Another desirable property of an advection scheme is the synchronous transport of fluid and tracers. At a given time step, the mass equation (6.1) is used to update Δp, and the tracer equation (6.2) is used to update $q\Delta p$. The tracer concentration q is then obtained as a ratio of the computed $q\Delta p$ and the computed Δp. However, this ratio can display nonphysical oscillations and extrema, even if the fields $q\Delta p$ and Δp are individually well-behaved. This is especially a problem in cases where Δp tends to zero.

This problem can be characterized as follows. The mass equation (6.1) and the tracer equation (6.2) together imply

$$\frac{Dq}{Dt} = \frac{\partial q}{\partial t} + \mathbf{u} \cdot \nabla q = 0. \tag{6.3}$$

That is, the material derivative of q is zero, so q is constant along particle paths. The value of $q(x, y, t + \Delta t)$ is therefore bounded by values of q in a neighbourhood of (x, y) at time t. It is desirable that a numerical method produce solutions having an analogous property, namely, that the computed value of q at position (x_i, y_j) at time t_{n+1} be bounded by the values of q at time t_n at position (x_i, y_j) and the immediately adjacent grid points. Schär and Smolarkiewicz (1996) use the term 'compatible' to refer to a numerical method with this property, and they demonstrate that compabibility can be obtained through a suitable limiting of antidiffusive correction fluxes.

An alternate approach to compatibility, and to advection in general, is given by the method of incremental remapping developed by Dukowicz and Baumgardner (2000). This method was subsequently extended to the modelling of sea ice transport by Lipscomb and Hunke (2004). Lipscomb and Ringler (2005) developed a version of incremental remapping for use with geodesic grids, in contrast to the rectangular grids discussed by Dukowicz and Baumgardner.

With the method of incremental remapping, each fixed grid cell is regarded as an 'arrival cell' for the location of a portion of mass at time t_{n+1}.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

The values of the velocity at the corners of that cell are used to trace the cell backward in time to yield an approximate 'departure cell' for the location of that mass at time t_n. The solution at time t_n is regarded as linear in each cell. In general, a departure cell contains portions of several of the fixed cells, and the contributions from these various portions are summed to yield the total mass in the departure cell at time t_n. This mass is then the mass in the arrival cell at time t_{n+1}. The cell masses at time t_{n+1} yield cell averages that are used to construct linear approximations in preparation for the next step.

This process resembles the 'reconstruct-evolve-average' (REA) approach to deriving advection schemes, in which piecewise fields are reconstructed from cell averages, propagated forward in time, and then averaged to yield new cell averages (LeVeque 2002). One difference between incremental remapping and the REA approach is that the former traces cells backward in time, not forward.

In the linear approximations used by Dukowicz and Baumgardner (2000), the slopes are determined by centred differences of the values of the solution in the adjacent cells. However, when necessary, these slopes are limited to ensure that the method is monotone, *i.e.*, that the solution at all spatial positions at a fixed time is a monotone increasing function of the solution at all positions at the preceding time level. This condition implies that the method is positive definite. The method of incremental remapping is second-order accurate in space, except at locations where the limiting is applied.

If this method is applied to a pair of equations of the form (6.1) and (6.2), then the computed tracer concentration q in a given cell at time t_{n+1} is a weighted average of the values of q at time t_n in that cell and in neighbouring cells, and compatibility is therefore ensured. In addition, the method is also conservative; by construction, the total mass at time t_{n+1} is the sum of the masses in the departure cells at time t_n, and this is the total mass at time t_n.

With the method of incremental remapping, substantial computational cost is incurred when constructing the departure regions. However, if multiple quantities are transported with the same velocity field, then the geometrical constructions need to be performed only once per time step, as the same results are used for all of the transported quantities. This method is therefore well-suited for models that transport large numbers of tracers. In contrast, the pseudovelocities used in MPDATA are different for each transported quantity, so the marginal cost of introducing an additional tracer is greater with MPDATA than in the case of incremental remapping (Lipscomb and Hunke 2004).

The mass and tracer equations (6.1) and (6.2) are equivalent to stating that the masses of the fluid and tracers are conserved in Lagrangian volumes, *i.e.*, in volumes that move with the fluid; this principle is used in

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Section 7 to derive equations (7.38) and (7.107), of which (6.1) and (6.2) are special cases. The method of incremental remapping is a discrete, finite-volume, representation of this Lagrangian principle. This method is related to the class of semi-Lagrangian methods, for which grid *points* are traced backward from time t_{n+1} to departure points at time t_n. Values of the solution at departure points are obtained via interpolation from grid point values, and time differences represent approximations to material derivatives along particle paths. These methods are reviewed, for example, in the text by Durran (1999). In general, semi-Lagrangian methods do not conserve mass exactly. However, a conservative method by Lin and Rood (1996) has a semi-Lagrangian flavour, and it involves directional splitting and a composition of one-dimensional advection schemes that are essentially one-dimensional versions of the method of incremental remapping.

The preceding discussion of advection is not all-inclusive. For example, the review by Griffies *et al.* (2000*a*) includes a list of some advection schemes that were used in operational ocean models as of the date of that review. These include flux-corrected transport (Zalesak 1979), three different variations on the Quick scheme of Leonard (1979), an advection scheme of Easter (1993), various centred schemes, and MPDATA.

In addition, Griffies *et al.* (2005) describe a third-order upwind biased method, with flux limiting, that has been implemented in the Modular Ocean Model and in the MIT general circulation model. Iskandarani, Levin, Choi and Haidvogel (2005) compare four advection schemes for high-order finite element and finite volume methods, including the discontinuous Galerkin method and a spectral finite volume method with flux limiting. Hecht (2006) reviews several advection methods for ocean modelling, especially in the context of forward-in-time discretizations which involve only two time levels.

6.4. Solution of the momentum equations

We next give an overview of some different approaches that are taken to solving the momentum equations. In this subsection the horizontal coordinates are assumed to be Cartesian coordinates on a tangent plane, for the sake of notational simplicity.

Section 7 includes a derivation of partial differential equations that describe the conservation of momentum for a fluid that is in motion relative to a rotating spheroid. In the case of Cartesian coordinates, the momentum equations (7.92) can be written in the form

$$\begin{aligned} &\frac{\partial}{\partial t}(up_s) + \frac{\partial}{\partial x}\big(u(up_s)\big) + \frac{\partial}{\partial y}\big(v(up_s)\big) + \frac{\partial}{\partial s}\big(\dot{s}(up_s)\big) - fvp_s = F_u, \\ &\frac{\partial}{\partial t}(vp_s) + \frac{\partial}{\partial x}\big(u(vp_s)\big) + \frac{\partial}{\partial y}\big(v(vp_s)\big) + \frac{\partial}{\partial s}\big(\dot{s}(vp_s)\big) + fup_s = F_v. \end{aligned} \tag{6.4}$$

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Here, s is a generalized vertical coordinate, f is the Coriolis parameter, and F_u and F_v represent forcing due to pressure and viscosity.

A vertically discrete model can be obtained by partitioning the fluid domain into layers, each of which is bounded above and below by surfaces of constant s. The conservation of momentum in a coordinate layer bounded by the surfaces $s = s_0$ and $s = s_1$ is described by equations (7.101) and (7.104). In the case of Cartesian coordinates, the u-equation (7.101) can be written in the form

$$\begin{aligned}\frac{\partial}{\partial t}(\bar{u}\Delta p) + \frac{\partial}{\partial x}\big(\bar{u}(\bar{u}\Delta p)\big) + \frac{\partial}{\partial y}\big(\bar{v}(\bar{u}\Delta p)\big) \\ + \Big[\big(\dot{s}p_s u\big)_{s=s_0} - \big(\dot{s}p_s u\big)_{s=s_1}\Big] - f\bar{v}\Delta p = \bar{F}_u,\end{aligned} \tag{6.5}$$

and the v-equation (7.104) can be expressed similarly. The horizontal velocity components $\bar{u}(x, y, t)$ and $\bar{v}(x, y, t)$ are mass-weighted vertical averages within the layer. The quantities $\bar{u}\Delta p$ and $\bar{v}\Delta p$ are the components of momentum per unit horizontal area, times g, and so they will be regarded as components of momentum density. The second and third terms in (6.5) represent the lateral advection of momentum, and the terms in square brackets represent the transport of momentum between layers due to the movement of coordinate surfaces relative to the fluid. As noted in Section 7.8, the quantity $-\dot{s}p_s$ is the rate of flow of mass per unit horizontal area (times g) across a coordinate surface due to material changes in s. In the case of a hydrostatic z-coordinate model, $-\dot{s}p_s = -\dot{z}p_z = \rho g w$, where w is the vertical component of velocity. In the case of a hybrid coordinate model, the rate of mass flow depends on the actions of the grid generator that establishes and moves the coordinate surfaces (Bleck 2002). With an isopycnic model, the flow across coordinate surfaces is due to the diapycnal diffusion of heat and salt, which causes the surfaces to move relative to the fluid; the modelling of this transport is discussed, for example, by McDougall and Dewar (1998), Hallberg (2000), and de Szoeke and Springer (2003).

One approach to solving the momentum equations is to use an advection scheme to implement the advective terms in the layer u-equation (6.5) and in the analogous v-equation. The Coriolis terms and the effects of pressure and viscosity can be regarded as forcing terms that are implemented with standard differencing and/or averaging. This approach was used by Smolarkiewicz and Margolin (1998) in the context of the shallow water equations for a single layer. In their formulation, the forcing is implemented with a Strang splitting; starting with the computed solution at time t_n, add half the forcing at time t_n (times Δt), apply an advection scheme to the result, and then add half the forcing at time t_{n+1} (times Δt).

A different strategy for the momentum equations is used in the Bryan–Cox class of z-coordinate models. The initial model in this class is described

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

by Bryan (1969). In that reference it is assumed that the Boussinesq approximation applies, *i.e.*, that the density is constant in the momentum equations except in the buoyancy term in the vertical component. In that case $p_s = p_z = -\rho g \approx -\rho_0 g$, and the u-equation in (6.4) has the form

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(uu) + \frac{\partial}{\partial y}(vu) + \frac{\partial}{\partial z}(wu) - fv = -F_u/(\rho_0 g). \tag{6.6}$$

In this equation the nonlinear terms on the left side represent an advection of the quantity u with advective velocity (u, v, w). To discretize these terms, consider grid cells centred at velocity points on the B-grid, and use spatial averages to obtain the necessary fluxes along the faces of those cells. The choice of discretization is guided by a desire to maintain energetic consistency in the model, namely, that the discrete nonlinear terms have no effect on the total kinetic energy in the solution and that the exchanges between kinetic and potential energy are represented correctly. These choices constrain the fluxes at cell faces to be calculated in terms of centred second-order averages.

In their description of the current version of the Modular Ocean Model, a direct descendant of the original Bryan–Cox model, Griffies *et al.* (2004) cite additional aspects of energetic consistency. One example is that the work done by the fluid against the pressure gradient can be converted into compressibility effects and/or work against gravity. The Modular Ocean Model no longer uses the Boussinesq approximation; one reason is that this approximation causes a model to preserve volume, which precludes the accurate modelling of sea-level rise due to thermal expansion. In order to employ the approach outlined above, Griffies *et al.* (2004) use a substitution $\rho_0 \tilde{\mathbf{u}} = \rho \mathbf{u}$ to convert the momentum equation into a form similar to (6.6). They give a detailed discussion of the averaging that they use to compute the necessary fluxes.

A third approach to solving the momentum equations is to convert the dependent variables to velocity and write the horizontal transport terms in terms of kinetic energy and vorticity. In the present case of Cartesian coordinates, the mass equation (7.32) is

$$\frac{\partial}{\partial t}(p_s) + \frac{\partial}{\partial x}(up_s) + \frac{\partial}{\partial y}(vp_s) + \frac{\partial}{\partial s}(\dot{s}p_s) = 0;$$

when this equation is combined with the u-momentum equation in (6.4), the result is

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + \dot{s}\frac{\partial u}{\partial s} - fv = F_u/p_s,$$

or

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left[\frac{1}{2}(u^2 + v^2)\right] - (\zeta + f)v + \dot{s}p_s\frac{\partial u}{\partial p} = F_u/p_s. \tag{6.7}$$

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

The last equation is obtained by writing $uu_x + vu_y$ in terms of the relative vorticity $\zeta = v_x - u_y$ and the kinetic energy per unit mass, $\frac{1}{2}(u^2 + v^2)$. The same procedure was used in Section 4.4 in preparation for proving the conservation of potential vorticity for the shallow water equations. As noted in that section, $\zeta + f$ is the local vertical component of the absolute (relative plus planetary) vorticity.

The kinetic energy and vorticity terms can be approximated with centred finite differences; if the C-grid is used, some spatial averaging is also required. In the case of a vertically discrete layered model, the vorticity term can be written as

$$(\zeta + f)v = \left(\frac{\zeta + f}{\Delta p}\right)(v\Delta p). \tag{6.8}$$

The first factor on the right side of (6.8) is proportional to the potential vorticity in a coordinate layer, and the second factor is proportional to mass flux. Sadourny (1975) describes two methods for implementing (6.8) on a C-grid, each of which uses various kinds of spatial averages to compute the potential vorticity and mass flux. With one such discretization, the total energy is conserved when applied to the shallow water equations with exact integration in time. The other discretization yields the conservation of potential enstrophy in the same context; in the case of the shallow water equations, the potential enstrophy is the spatial integral of $\frac{1}{2}(\zeta + f)^2/h$. The conservation of this quantity is of interest because it inhibits the accumulation of energy at grid scales and thus promotes numerical stability. The potential-enstrophy-conserving method is used, for example, in the Miami Isopycnic Coordinate Ocean Model (Bleck 2002), whereas the energy-conserving method is used in the Hallberg Isopycnal Model (Hallberg 1997). Because of the particular nature of the spatial averaging used in the potential-enstrophy-conserving method, special procedures must be used in situations where Δp tends to zero. However, no special procedures are needed for the energy-conserving method.

Adcroft *et al.* (2004) use the formulation (6.7) when solving the momentum equations on the conformally expanded spherical cube (see Section 6.1). In their setting the vorticity is computed by a finite volume approach; a vorticity point is regarded as the centre of a grid cell, and the circulation around the boundary of that cell, divided by area, gives the vorticity. This formulation avoids problems associated with the singular points corresponding to the corners of the cube.

7. Governing equations in general coordinates

We now address the derivation of the governing equations that are to be solved numerically. In particular, the goal of the present section is to derive partial differential equations that describe the conservation of mass,

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

momentum, and tracers for a hydrostatic and stratified fluid in a rotating reference frame. In this derivation the vertical coordinate is a generalized coordinate that includes all of the cases described in Section 3, and the horizontal coordinates are arbitrary orthogonal coordinates on a spheroid. Some aspects of the following discussion are also contained in various references, such as Bleck (2002), Gill (1982), Griffies (2004), Holton (1992), Miller (2006), and Pedlosky (1987).

7.1. Definition of level surfaces

The Earth's surface is not exactly a sphere, but instead can be described more accurately as an oblate spheroid (Gill 1982). In particular, the surface of constant gravitational potential at sea level is approximated by a spheroid with a polar radius approximately equal to 6357 km and an equatorial radius of approximately 6378 km. The maximum deviation from the spheroid of best fit is on the order of 100 metres.

In the following, we consider fluid motion relative to a spheroid that rotates with uniform angular velocity Ω. Assume that the axis of rotation is stationary in a rectangular coordinate system that serves as an inertial reference frame. As a first approximation, one can define directions tangent to the spheroid as horizontal, and directions normal to the spheroid as vertical. However, for later use it will be useful to incorporate the effects of centripetal acceleration into such a definition.

Suppose that a layer of fluid is stationary relative to the rotating spheroid, and assume that the only forces acting on the fluid are due to gravity and pressure. For a given fluid parcel and a given time, let $\mathbf{a}_p$ denote the net acceleration (force per unit mass) due to pressure forcing, and denote the acceleration due to gravity by $\mathbf{a}_g = -\nabla\phi$. Here, ϕ is the gravitational potential, ∇ is the three-dimensional spatial gradient, and all quantities are viewed in the inertial frame. Since the parcel rotates with uniform angular velocity Ω, the resultant force per unit mass must be the centripetal acceleration that is required to maintain the rotation. The centripetal acceleration is directed toward the axis of rotation and has magnitude $\Omega^2 r_\perp$, where $r_\perp$ denotes the distance from the parcel to the axis. This acceleration can also be written in terms of a potential function as $-\nabla\phi_c$, where $\phi_c = \frac{1}{2}\Omega^2 r_\perp^2$ (Pedlosky 1987). It then follows that the centripetal acceleration is $-\nabla\phi_c = \mathbf{a}_p + \mathbf{a}_g = \mathbf{a}_p - \nabla\phi$: see Figure 7.1.

With the Earth's rotation rate of approximately $2\pi + 2\pi/360.24$ radians in 24 hours (Griffies 2004), the centripetal acceleration at radius 6300 km has a magnitude of approximately 0.03 m/sec^2. In contrast, the acceleration due to gravity is approximately 9.8 m/sec^2. It is thus a slight imbalance between gravitational and pressure forces that provides the necessary centripetal force.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

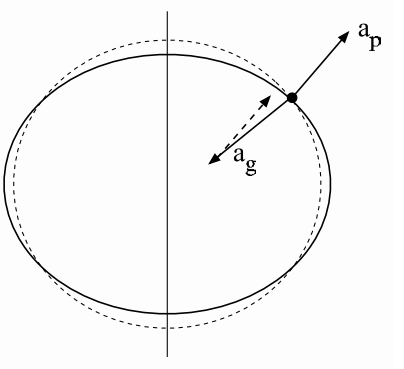


Figure 7.1. Combined effect of gravitational and pressure accelerations. The vertical line segment is the axis of rotation. The dashed ellipse is a cross-section of a spheroid of constant gravitational potential ϕ; the gravitational acceleration $\mathbf{a}_g = -\nabla\phi$ is orthogonal to this spheroid. The solid ellipse is a cross-section of the spheroid of constant gravitational–centripetal potential $\Phi = \phi - \phi_c$ that intersects the first spheroid at the indicated point. The pressure acceleration $\mathbf{a}_p$ for a relatively stationary fluid is orthogonal to the surface of constant Φ. The dashed arrow is a translation of $\mathbf{a}_p$, and the sum $\mathbf{a}_g + \mathbf{a}_p = -\nabla\phi + \mathbf{a}_p$ is the centripetal acceleration that is required to maintain the rotation. Locally horizontal and vertical directions are defined relative to surfaces of constant Φ.

The effects of the gravitational and centripetal potentials can be combined into a potential function $\Phi = \phi - \phi_c = \phi - \frac{1}{2}\Omega^2 r_\perp^2$. Surfaces of constant Φ are also spheroids, and they bulge out slightly near the equator, compared to surfaces of constant ϕ. For a layer of fluid that is stationary relative to the rotating spheroid, $\mathbf{a}_p - \nabla\Phi = 0$, and $\mathbf{a}_p$ is then orthogonal to the surfaces of constant Φ.

In the following discussions, directions that are tangent to surfaces of constant Φ will be regarded as 'level', or locally 'horizontal', and directions normal to such surfaces will be regarded as locally 'vertical'.

In the case of a fluid that is relatively stationary, the pressure acceleration $\mathbf{a}_p$ points vertically, and it exactly cancels the effect of the gravitational–centripetal potential Φ. More generally, for a moving fluid we will impose the 'hydrostatic assumption' that the vertical component of the pressure acceleration equals $\nabla\Phi$. The justification for this assumption follows from the scaling of large-scale flows that is discussed in Section 2.1.

Assume that $\mathbf{a}_p$ is proportional to the pressure gradient ∇p. For a relatively stationary fluid, $\mathbf{a}_p$ is parallel to $\nabla\Phi$, and the surfaces of constant

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

pressure coincide with the surfaces of constant Φ. As noted above, the latter surfaces bulge out slightly near the equator, relative to the rotating spheroid. This bulging of the pressure surfaces provides the centripetal acceleration that enables the fluid to rotate with the spheroid without assistance from any forcing other than gravity and hydrostatic pressure.

The magnitude of the gravitational acceleration is nearly constant over the entire fluid domain. At latitude 45°, this magnitude is approximately 9.806 m/sec^2, whereas the area average over the entire Earth is approximately 9.7976 m/sec^2 (Griffies 2004). In the following, it will be assumed that $|\nabla\Phi| = g$, where g is a constant.

7.2. Horizontal and vertical coordinates

Let Σ denote a surface of constant gravitational–centripetal potential Φ, such as at mean sea level. As before, assume that Σ rotates with uniform angular velocity Ω about an axis that is fixed in an inertial reference frame. Parametrize all or part of Σ with coordinates $\mathbf{x} = (x_1, x_2)$, and let z denote a distance perpendicular to Σ, with z increasing outward and $z = 0$ corresponding to points on Σ. The coordinate z has units of length, but x_1 and x_2 do not necessarily have those units.

The coordinates $(\mathbf{x}, z) = (x_1, x_2, z)$ define the position of a point relative to the rotating surface Σ. For any such point and any time t, let $\mathbf{r}(\mathbf{x}, z, t)$ denote the position of that point relative to the inertial reference frame at time t. That is, $\mathbf{r}(\mathbf{x}, z, t)$ is a vector of three numbers, with units of length, that gives the position of the point relative to the rectangular coordinate system in which the axis of rotation is stationary. The effects of rotation can be expressed as

$$\mathbf{r}(\mathbf{x}, z, t) = Q(\Omega t)\ \mathbf{r}(\mathbf{x}, z, 0), \tag{7.1}$$

where $Q(\Omega t)$ is an orthogonal matrix that represents a rotation through angle Ωt about the given axis. In (7.1), $\mathbf{r}(\mathbf{x}, z, t)$ and $\mathbf{r}(\mathbf{x}, z, 0)$ are regarded as column vectors.

Define the metric coefficients m_1 and m_2 by

$$m_i(\mathbf{x}, z) = \left| \frac{\partial \mathbf{r}}{\partial x_i}(\mathbf{x}, z, t) \right|, \qquad i = 1, 2. \tag{7.2}$$

Here, the Euclidean vector norm is used. A comparison with (7.1) shows that the right side of (7.2) is independent of t, so the notation on the left side indicates no dependence on t. Equation (7.2) implies that an increment Δx_i in the parameter x_i corresponds to a horizontal spatial displacement approximately equal to $m_i \Delta x_i$. In the case of an ocean model the metric coefficient m_i is nearly independent of z, as the ocean is a few kilometres deep, whereas the Earth has a radius of over 6300 km.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Next define the unit vectors **i** and **j** by

$$\begin{aligned}\mathbf{i}(\mathbf{x},t) &= \frac{1}{m_1(\mathbf{x},0)}\frac{\partial \mathbf{r}}{\partial x_1}(\mathbf{x},0,t),\\ \mathbf{j}(\mathbf{x},t) &= \frac{1}{m_2(\mathbf{x},0)}\frac{\partial \mathbf{r}}{\partial x_2}(\mathbf{x},0,t).\end{aligned} \tag{7.3}$$

The partial derivatives in (7.3) are taken for fixed $z = 0$, so the vectors **i** and **j** are tangent to the surface Σ and thus locally horizontal. These vectors can be regarded as unit vectors pointing in the x_1- and x_2-directions, respectively, that are attached to Σ and rotate with Σ. Throughout the following discussion, the horizontal coordinate system is assumed to be orthogonal, in the sense that **i** and **j** are everywhere orthogonal. The unit vector in the upward vertical direction is

$$\mathbf{k}(\mathbf{x},t) = \frac{\partial \mathbf{r}}{\partial z}(\mathbf{x},z,t). \tag{7.4}$$

Since the distance z is taken orthogonal to the surface Σ, the vectors **i**, **j**, and **k** are mutually orthogonal. In Sections 7.12 and 7.15 it is assumed that Σ is parametrized so that the coordinate system is right-handed, in the sense that $\mathbf{i}\times\mathbf{j} = \mathbf{k}$, $\mathbf{k}\times\mathbf{j} = -\mathbf{i}$, and $\mathbf{i}\times\mathbf{k} = -\mathbf{j}$; however, this assumption is not used elsewhere.

In later discussions we use the approximations

$$\begin{aligned}\frac{\partial \mathbf{r}}{\partial x_1}(\mathbf{x},z,t) &= m_1(\mathbf{x},z)\ \mathbf{i}(\mathbf{x},t),\\ \frac{\partial \mathbf{r}}{\partial x_2}(\mathbf{x},z,t) &= m_2(\mathbf{x},z)\ \mathbf{j}(\mathbf{x},t),\end{aligned} \tag{7.5}$$

which are exact when $z = 0$. In effect, these approximations assume that the tangent vectors to surfaces of constant z are parallel.

For an example of a coordinate system, assume that Σ is a sphere of radius a centred at the origin in the inertial frame. Also assume that the axis of rotation aligns with the third coordinate axis in that reference frame and that the rotation is anticlockwise when the sphere is viewed from the positive portion of that axis. Use spherical coordinates, with longitude $x_1 = \lambda$ and latitude $x_2 = \theta$, and assume that at time $t = 0$ the points with $\lambda = 0$ align with the first coordinate in the inertial frame. Then

$$\mathbf{r}(\mathbf{x},z,t) = \mathbf{r}(\lambda,\theta,z,t) = \big(r\cos\theta\cos(\lambda+\Omega t),\ r\cos\theta\sin(\lambda+\Omega t),\ r\sin\theta\big), \tag{7.6}$$

with $r = a + z \approx a$. In this case the unit vectors **i** and **j** point eastward and

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

northward, respectively; the metric coefficients are

$$m_\lambda = \left|\frac{\partial \mathbf{r}}{\partial \lambda}\right| = (a+z)\cos\theta \approx a\cos\theta,$$
$$m_\theta = \left|\frac{\partial \mathbf{r}}{\partial \theta}\right| = a + z \approx a; \tag{7.7}$$

and the relations (7.5) are exact.

Gill (1982) points out that the horizontal spheroidal coordinates for the Earth can be approximated accurately in terms of spherical coordinates. However, as noted in Section 6, for a numerical model of the global ocean circulation it is inadvisable to use spherical (or spheroidal) coordinates, because of the convergence of grid lines at the North Pole.

7.3. Generalized vertical coordinate

Here we develop a generalized vertical coordinate s for a hydrostatic, stratified fluid lying in a basin that rotates along with the equipotential surface Σ described in the preceding section. Possible choices for s include the level coordinate z, an isopycnic coordinate using the reciprocal of potential density (or a related quantity), the terrain-following coordinate σ, or some hybrid of the preceding.

Assume that, for fixed time t and horizontal coordinates $\mathbf{x} = (x_1, x_2)$, a quantity s is an increasing function of z. Equivalently, z is an increasing function of s. A surface of constant s can move upward and downward with time, and for fixed time the elevation of that surface can vary with horizontal position. For each $\mathbf{x}$ and t, let $z(\mathbf{x}, s, t)$ denote the elevation of such a surface.

The derivative $\partial z/\partial s$ relates increments in the parameter s to increments in space, and it can therefore serve as a metric coefficient for the vertical direction. For later use, it will be useful to relate this quantity to pressure and density.

Let $P(\mathbf{x}, z, t)$ denote the pressure in the fluid corresponding to horizontal position $\mathbf{x}$, elevation z, and time t. The hydrostatic assumption of Section 7.1 implies

$$\frac{\partial P}{\partial z} = -\rho g, \tag{7.8}$$

where ρ is the density of the fluid. In particular, the net vertical component of the pressure force acting on an element of fluid having horizontal area ΔA and height Δz is approximately $-\frac{\partial P}{\partial z}\Delta z \Delta A = -\frac{1}{\rho}\frac{\partial P}{\partial z}(\rho \Delta z \Delta A)$, so the pressure force per unit mass is approximately $-\frac{1}{\rho}\frac{\partial P}{\partial z}$. According to the hydrostatic assumption, this quantity equals $|\nabla \Phi|$, which was assumed to equal a constant value g at the end of Section 7.1.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Now let $p(\mathbf{x}, s, t)$ denote the pressure in terms of s, so that $p(\mathbf{x}, s, t) = P(\mathbf{x}, z(\mathbf{x}, s, t), t)$. Then

$$\begin{aligned} \frac{\partial p}{\partial s}(\mathbf{x}, s, t) &= \frac{\partial P}{\partial z}(\mathbf{x}, z(\mathbf{x}, s, t), t)\ \frac{\partial z}{\partial s}(\mathbf{x}, s, t) \\ &= \frac{-g}{\alpha(\mathbf{x}, s, t)}\ \frac{\partial z}{\partial s}(\mathbf{x}, s, t). \end{aligned} \tag{7.9}$$

Here, $\alpha(\mathbf{x}, s, t) = 1/\rho(\mathbf{x}, s, t)$ denotes the specific volume (volume per unit mass). A metric coefficient for the coordinate s is then defined by

$$m_s(\mathbf{x}, s, t) = z_s(\mathbf{x}, s, t) = -\frac{\alpha(\mathbf{x}, s, t) p_s(\mathbf{x}, s, t)}{g}. \tag{7.10}$$

The subscripts on z and p denote partial derivatives. The coefficient m_s is positive, since $\partial p/\partial s < 0$.

An increment Δs in s corresponds to a vertical distance approximately equal to $m_s \Delta s$, in analogy to the horizontal displacement $m_i \Delta x_i$ corresponding to an increment Δx_i in the horizontal parameter x_i. Note that the lateral displacements are not measured along surfaces of constant s, but instead are measured in terms of projections onto the horizontal. This formulation is also used, for example, by Bleck (1978, 2002).

7.4. Integration in parameter space

The partial differential equations for conservation of mass and momentum will be derived so that the independent spatial variables are the parameters $(\mathbf{x}, s) = (x_1, x_2, s)$. However, quantities such as mass per unit volume and momentum per unit volume are naturally expressed in terms of rectangular coordinates having units of length. In the following, the term 'parameter space' refers to the coordinates $(\mathbf{x}, s)$ attached to the rotating spheroid Σ, and the term 'rectangular coordinates' refers to the rectangular coordinate system in which the axis of rotation is stationary. The inertial nature of this frame is immaterial to the formulation of conservation of mass, but it is essential to the formulation of conservation of momentum. The purpose of the present subsection is to develop the necessary change of variables between integrals in these two coordinate systems.

Let $A(t)$ be a region in parameter space, and let $B(t)$ denote the corresponding region in the rectangular coordinate system in the inertial reference frame at time t. ($A(t)$ and $B(t)$ provide different mathematical descriptions of the same physical entity, and in the following they are regarded as different mathematical objects. The dependence on t is immaterial to the present subsection, but it is included in the notation for the sake of further developments.) The coordinate mapping from $A(t)$ to $B(t)$ can be represented as

$$\tilde{\mathbf{r}}(\mathbf{x}, s, t) = \mathbf{r}\big(\mathbf{x}, z(\mathbf{x}, s, t), t\big) \tag{7.11}$$

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

for all $(\mathbf{x}, s) \in A(t)$. Let $\tilde{\psi}(\cdot, t)$ denote an integrable function defined on $B(t)$, and let $\psi(\cdot, t)$ be the corresponding function on $A(t)$ defined by $\psi(\mathbf{x}, s, t) = \tilde{\psi}\big(\tilde{\mathbf{r}}(\mathbf{x}, s, t), t\big)$. Then

$$\begin{aligned}\int_{B(t)} \tilde{\psi}(\cdot, t)\, \mathrm{d}V_{B(t)} &= \int_{A(t)} \tilde{\psi}\big(\tilde{\mathbf{r}}(\mathbf{x}, s, t), t\big)\ J(\mathbf{x}, s, t)\, \mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}s \\ &= \int_{A(t)} \psi(\mathbf{x}, s, t)\ J(\mathbf{x}, s, t)\, \mathrm{d}\mathbf{x}\, \mathrm{d}s,\end{aligned} \tag{7.12}$$

where $\mathrm{d}V_{B(t)}$ denotes integration on $B(t)$, and $J(\mathbf{x}, s, t)$ is the absolute value of the Jacobian,

$$J(\mathbf{x}, s, t) = \left|\det\left[\frac{\partial\tilde{\mathbf{r}}}{\partial x_1}(\mathbf{x}, s, t),\ \frac{\partial\tilde{\mathbf{r}}}{\partial x_2}(\mathbf{x}, s, t),\ \frac{\partial\tilde{\mathbf{r}}}{\partial s}(\mathbf{x}, s, t)\right]\right|. \tag{7.13}$$

Here, $\tilde{\mathbf{r}}$ is interpreted as a column vector, and the quantity in square brackets is a 3×3 matrix. Insert the definition (7.11) of $\tilde{\mathbf{r}}$ into (7.13) to obtain

$$J(\mathbf{x}, s, t) = \left|\det\left[\frac{\partial\mathbf{r}}{\partial x_1} + \frac{\partial\mathbf{r}}{\partial z}\frac{\partial z}{\partial x_1},\ \frac{\partial\mathbf{r}}{\partial x_2} + \frac{\partial\mathbf{r}}{\partial z}\frac{\partial z}{\partial x_2},\ \frac{\partial\mathbf{r}}{\partial z}\frac{\partial z}{\partial s}\right]\right| \tag{7.14}$$

$$= \left|\det\left[\frac{\partial\mathbf{r}}{\partial x_1},\ \frac{\partial\mathbf{r}}{\partial x_2},\ \frac{\partial\mathbf{r}}{\partial z}\frac{\partial z}{\partial s}\right]\right| \tag{7.15}$$

$$= m_1 m_2 m_s\ \big|\det\big[\mathbf{i}(\mathbf{x}, t), \mathbf{j}(\mathbf{x}, t), \mathbf{k}(\mathbf{x}, t)\big]\big|. \tag{7.16}$$

In (7.14)–(7.15), the partial derivatives of $\mathbf{r}$ are evaluated at $\big(\mathbf{x}, z(\mathbf{x}, s, t), t\big)$, and the partial derivatives of z are evaluated at $(\mathbf{x}, s, t)$. The representation (7.15) arises from expanding the sums in (7.14) and using the fact that a determinant is zero if one column is a multiple of another. The representation (7.16) relies on the relations (7.5); the metric coefficients m_1 and m_2 are evaluated at $\big(\mathbf{x}, z(\mathbf{x}, s, t)\big)$; and m_s is evaluated at $(\mathbf{x}, s, t)$. The unit vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are mutually orthogonal, so the determinant in (7.16) has absolute value equal to 1. It then follows that

$$J(\mathbf{x}, s, t) = m_1\big(\mathbf{x}, z(\mathbf{x}, s, t)\big)\ m_2\big(\mathbf{x}, z(\mathbf{x}, s, t)\big)\ m_s(\mathbf{x}, s, t) \tag{7.17}$$

and thus

$$\int_{B(t)} \tilde{\psi}(\cdot, t)\, \mathrm{d}V_{B(t)} = \int_{A(t)} \psi(\mathbf{x}, s, t)\ m_1 m_2 m_s\, \mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}s. \tag{7.18}$$

For an interpretation of the preceding, consider a rectangular solid having sides Δx_1, Δx_2, Δs in the region $A(t)$ in parameter space. The corresponding subset of $B(t)$ is approximately a parallelepiped generated by the vectors

$$\frac{\partial\tilde{\mathbf{r}}}{\partial x_1}\Delta x_1, \quad \frac{\partial\tilde{\mathbf{r}}}{\partial x_2}\Delta x_2, \quad \frac{\partial\tilde{\mathbf{r}}}{\partial s}\Delta s.$$

The first two of these vectors are tangent to a surface of constant s, and

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

the expressions in (7.14) amount to resolving these vectors into horizontal and vertical components. The subsequent expressions (7.15)–(7.16) give the volume of a rectangular solid having the same volume as the parallelepiped. The appropriate volume element is then $m_1 m_2 m_s \, \mathrm{d}x_1 \, \mathrm{d}x_2 \, \mathrm{d}s$.

7.5. Integration on a material volume

Now consider integration on a time-dependent region that moves with a fluid flow. In particular, suppose that a volume of fluid occupies a region $A(t)$ in parameter space at time t, and let $B(t)$ denote the corresponding region in rectangular coordinates in the inertial reference frame. Assume that a function $\tilde{\psi}(\cdot, t)$ is defined and integrable on $B(t)$ for each t, and let $\psi(\cdot, t)$ be the corresponding function on $A(t)$ defined by $\psi(\mathbf{x}, s, t) = \tilde{\psi}\big(\tilde{\mathbf{r}}(\mathbf{x}, s, t), t\big)$ for all $(\mathbf{x}, s) \in A(t)$.

In the following usage, $\tilde{\psi}$ represents a fluid property such as mass density or a component of momentum density, and an integral on $B(t)$ is used when representing a conservation principle for that mass of fluid. In particular, it will be necessary to compute a time derivative of such an integral, and for that purpose it is useful to relate an integral on $B(t)$ to an integral on a fixed region in parameter space at time $t = 0$.

Let $(\mathbf{X}, S) \in A(0)$ denote the position of a fluid parcel in parameter space at time 0, and denote the position of that parcel in parameter space at any time t by $\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t)\big)$. That is, $\mathbf{X}$ and S are Lagrangian variables, and $\mathbf{x}$ and s are Eulerian variables. Then

$$\begin{aligned} \int_{B(t)} \tilde{\psi}(\cdot, t) \, \mathrm{d}V_{B(t)} &= \int_{A(t)} \psi(\mathbf{x}, s, t) \; J(\mathbf{x}, s, t) \, \mathrm{d}\mathbf{x} \, \mathrm{d}s \\ &= \int_{A(0)} \psi(\mathbf{x}, s, t) \; J(\mathbf{x}, s, t) \; H(\mathbf{X}, S, t) \, \mathrm{d}\mathbf{X} \, \mathrm{d}s. \end{aligned} \tag{7.19}$$

Here, $\mathbf{x}$ and s are functions of $(\mathbf{X}, S, t)$, and $H(\mathbf{X}, S, t)$ is the Jacobian of the transformation from $(\mathbf{X}, S)$ to $(\mathbf{x}, s)$, *i.e.*,

$$H(\mathbf{X}, S, t) = \begin{pmatrix} \frac{\partial x_1}{\partial X_1} & \frac{\partial x_1}{\partial X_2} & \frac{\partial x_1}{\partial S} \\ \frac{\partial x_2}{\partial X_1} & \frac{\partial x_2}{\partial X_2} & \frac{\partial x_2}{\partial S} \\ \frac{\partial s}{\partial X_1} & \frac{\partial s}{\partial X_2} & \frac{\partial s}{\partial S} \end{pmatrix}. \tag{7.20}$$

The last integral in (7.19) uses H instead of $|H|$, as $H(\mathbf{X}, S, t) > 0$ for all $(\mathbf{X}, S, t)$. This can be demonstrated as follows. Impose the physically realistic assumption that this Jacobian is everywhere nonzero for all t, so that positive volumes are mapped to positive volumes, and also assume that the flow is smooth enough that H is a continuous function. At time $t = 0$ the mapping $(\mathbf{X}, S) \mapsto (\mathbf{x}, s)$ is the identity mapping. In that case H is the determinant of the identity matrix, so $H(\mathbf{X}, S, 0) = 1$ for all $(\mathbf{X}, S)$.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

If $H(\mathbf{X}, S, t) < 0$ for some $(\mathbf{X}, S, t)$, then continuity in time implies that $H(\mathbf{X}, S, t_0) = 0$ for some t_0, which is a contradiction.

7.6. Time derivative of a material integral

Next consider the time derivative of the integral (7.19) over a material volume. For that purpose, it will be useful to define the following quantities.

For any point $(\mathbf{x}, s) \in A(t)$, let $\dot{\mathbf{x}}(\mathbf{x}, s, t)$ and $\dot{s}(\mathbf{x}, s, t)$ denote the time derivatives of $\mathbf{x}$ and s, respectively, as seen by the fluid parcel that is located at position $(\mathbf{x}, s)$ in parameter space at time t. More precisely,

$$\begin{aligned} \frac{\partial \mathbf{x}}{\partial t}(\mathbf{X}, S, t) &= \dot{\mathbf{x}}\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big), \\ \frac{\partial s}{\partial t}(\mathbf{X}, S, t) &= \dot{s}\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big) \end{aligned} \tag{7.21}$$

for all $(\mathbf{X}, S) \in A(0)$ and all t.

In general, the quantities $\dot{x}_1$, $\dot{x}_2$, and $\dot{s}$ are not components of linear velocity. For example, if spherical coordinates are used for the horizontal parameters, then $\dot{x}_1 = \dot{\lambda}$ and $\dot{x}_2 = \dot{\theta}$ represent the time derivatives of longitude and latitude, respectively, as seen by an observer moving with the fluid. Corresponding components of linear velocity are obtained by using the appropriate metric coefficients, as described in Section 7.9. In addition, the values of $\dot{s}$ need not be related to any geometrical motion at all; for example, if s is the reciprocal of potential density, then nonzero values of $\dot{s}$ could result entirely from a warming or cooling of the fluid.

Now define a material derivative of a quantity in terms of the coordinates used here. Assume that $F(\mathbf{x}, s, t)$ is defined for all $(\mathbf{x}, s) \in A(t)$, and let $\hat{F}(\mathbf{X}, S, t) = F\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big)$ for all $(\mathbf{X}, S) \in A(0)$. Then

$$\begin{aligned} \frac{\partial \hat{F}}{\partial t}(\mathbf{X}, S, t) &= \frac{\partial F}{\partial t} + \frac{\partial \mathbf{x}}{\partial t}(\mathbf{X}, S, t) \cdot \nabla F + \frac{\partial s}{\partial t}(\mathbf{X}, S, t) \frac{\partial F}{\partial s} \\ &= \frac{\partial F}{\partial t} + \dot{\mathbf{x}} \cdot \nabla F + \dot{s} \frac{\partial F}{\partial s}. \end{aligned} \tag{7.22}$$

Here, $\nabla F(\mathbf{x}, s, t) = \big(\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}\big)$ for all $(\mathbf{x}, s) \in A(t)$ and all t. The notation ∇ denotes the gradient with respect to (x_1, x_2) for fixed s; the notation ∇_s is also used in this context, but a subscript s is unnecessary here, given the definition of F. In (7.22), $\dot{\mathbf{x}}$, $\dot{s}$, and all derivatives of F are evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big)$. Now define the material derivative

$$\frac{DF}{Dt}(\mathbf{x}, s, t) = \frac{\partial F}{\partial t} + \dot{\mathbf{x}} \cdot \nabla F + \dot{s} F_s \tag{7.23}$$

for all $(\mathbf{x}, s) \in A(t)$ and all t, where all quantities on the right side are evaluated at $(\mathbf{x}, s, t)$, and $F_s = \partial F / \partial s$. The quantity DF/Dt is the time

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

derivative of F following fluid parcels, as opposed to the partial derivative $\partial F/\partial t$ for fixed position, and $DF/Dt(\mathbf{x}, s, t)$ refers to the parcel that is located at position $(\mathbf{x}, s)$ at time t.

The time derivative of the integral (7.19) includes the time derivative of the Jacobian H in (7.20). The calculation of this integral is an analogue of the calculation done for standard Cartesian coordinates (*e.g.*, Chorin and Marsden (1990)), and it will not be given here. Instead, we simply state the result, which is

$$\frac{\partial H}{\partial t}(\mathbf{X}, S, t) = \left(\frac{\partial \dot{x}_1}{\partial x_1} + \frac{\partial \dot{x}_2}{\partial x_2} + \frac{\partial \dot{s}}{\partial s}\right) H(\mathbf{X}, S, t). \tag{7.24}$$

Each of the terms in the parentheses is evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big)$.

The result (7.24) can be interpreted as follows. The Jacobian H represents the 'volume' of a fluid parcel in parameter space at time t, relative to its volume at time 0. If $\partial \dot{x}_1/\partial x_1 > 0$ then the rate of fluid flow in the x_1-direction increases with x_1, so the fluid expands in that direction. The quantity in parentheses in (7.24) represents the net effect of such quantities over the three coordinate directions, and it is a generalization of the usual divergence of velocity. If the generalized divergence is positive, then $\partial H/\partial t > 0$, and the fluid parcel expands as t increases, as expected.

Now let $E(t)$ denote the time-dependent integral (7.19), and compute $E'(t)$. The quantities ψ and J in (7.19) are evaluated at the point $\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big)$, so the time derivative of $J\psi$ in (7.19) is the material derivative. Then

$$\begin{aligned} E'(t) &= \int_{A(0)} \left[\frac{D(J\psi)}{Dt} + J\psi\left(\frac{\partial \dot{x}_1}{\partial x_1} + \frac{\partial \dot{x}_2}{\partial x_2} + \frac{\partial \dot{s}}{\partial s}\right)\right] H(\mathbf{X}, S, t)\, \mathrm{d}\mathbf{X}\, \mathrm{d}S \\ &= \int_{A(t)} \left[\frac{D(J\psi)}{Dt} + J\psi\left(\frac{\partial \dot{x}_1}{\partial x_1} + \frac{\partial \dot{x}_2}{\partial x_2} + \frac{\partial \dot{s}}{\partial s}\right)\right] \mathrm{d}\mathbf{x}\, \mathrm{d}s. \end{aligned} \tag{7.25}$$

In the first integral, the terms in the square brackets are evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big)$, and in the second integral they are evaluated at $(\mathbf{x}, s, t)$. A comparison with the expression (7.23) for the material derivative yields

$$E'(t) = \int_{A(t)} D_f(J\psi)\, \mathrm{d}\mathbf{x}\, \mathrm{d}s, \tag{7.26}$$

where

$$D_f(J\psi) = \frac{\partial}{\partial t}\big(J\psi\big) + \frac{\partial}{\partial x_1}\big(\dot{x}_1 J\psi\big) + \frac{\partial}{\partial x_2}\big(\dot{x}_2 J\psi\big) + \frac{\partial}{\partial s}\big(\dot{s} J\psi\big). \tag{7.27}$$

The quantity $D_f(J\psi)$ will be termed a 'flux derivative' of $J\psi$, as the spatial terms are expressed in terms of derivatives of fluxes.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

7.7. Conservation of mass

Now suppose that $\tilde{\psi}(\cdot,t) = \tilde{\rho}(\cdot,t)$ is the density function (mass per unit volume) at time t in some fluid domain, as seen in rectangular coordinates in the inertial frame. The corresponding function on parameter space is then defined by $\psi(\mathbf{x},s,t) = \rho(\mathbf{x},s,t) = \tilde{\psi}\big(\tilde{\mathbf{r}}(\mathbf{x},s,t),t\big)$. Consider a volume of fluid within this domain, and let $A(t)$ and $B(t)$ denote the regions in parameter space and rectangular coordinates, respectively, that are occupied by this volume at time t. The total mass of this volume of fluid at time t is

$$\int_{B(t)} \tilde{\psi}(\cdot,t)\,\mathrm{d}V_{B(t)} = \int_{B(t)} \tilde{\rho}(\cdot,t)\,\mathrm{d}V_{B(t)}. \tag{7.28}$$

Since the region $B(t)$ follows the flow, the mass (7.28) must remain constant in time. Equation (7.26) then implies

$$\int_{A(t)} D_f(J\rho)\,\mathrm{d}\mathbf{x}\,\mathrm{d}s = 0$$

for all t. This relation holds for all regions $A(t)$ within the fluid domain. If the function $D_f(J\rho)$ is continuous everywhere, then $D_f(J\rho) = 0$ at all positions in space at all times, *i.e.*,

$$\frac{\partial}{\partial t}\big(J\rho\big) + \frac{\partial}{\partial x_1}\big(\dot{x}_1 J\rho\big) + \frac{\partial}{\partial x_2}\big(\dot{x}_2 J\rho\big) + \frac{\partial}{\partial s}\big(\dot{s} J\rho\big) = 0. \tag{7.29}$$

All quantities in (7.29) are functions of $(\mathbf{x},s,t)$.

Equation (7.29) is the basic statement of conservation of mass, but it is useful to re-write it as follows. The definition (7.10) of the metric coefficient m_s implies

$$\rho(\mathbf{x},s,t) = \frac{1}{\alpha(\mathbf{x},s,t)} = -\frac{p_s(\mathbf{x},s,t)}{g\ m_s(\mathbf{x},s,t)}. \tag{7.30}$$

The definition $J = m_1 m_2 m_s$ from (7.17) implies

$$J\rho = -m_1 m_2 p_s g^{-1}, \tag{7.31}$$

and equation (7.29) becomes

$$\frac{\partial}{\partial t}\big(Gp_s\big) + \frac{\partial}{\partial x_1}\big(\dot{x}_1 Gp_s\big) + \frac{\partial}{\partial x_2}\big(\dot{x}_2 Gp_s\big) + \frac{\partial}{\partial s}\big(\dot{s} Gp_s\big) = 0, \tag{7.32}$$

or equivalently,

$$\frac{\partial}{\partial t}\big(Gp_s\big) + \nabla\cdot\big(\dot{\mathbf{x}} Gp_s\big) + \frac{\partial}{\partial s}\big(\dot{s} Gp_s\big) = 0. \tag{7.33}$$

Here,

$$\begin{aligned} G(\mathbf{x},s,t) &= m_1\big(\mathbf{x},z(\mathbf{x},s,t)\big)\ m_2\big(\mathbf{x},z(\mathbf{x},s,t)\big) \\ &\approx m_1(\mathbf{x}) m_2(\mathbf{x}), \end{aligned} \tag{7.34}$$

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

and $p_s = \partial p/\partial s$. As noted in Section 7.2, the metric coefficients m_1 and m_2 are nearly independent of z, so to good approximation one can use $G(\mathbf{x}) = m_1(\mathbf{x})m_2(\mathbf{x})$. In equations (7.32)–(7.33), the operations $\partial/\partial x_i$ and ∇ are taken for fixed s, as the functions involved depend on $(\mathbf{x}, s, t)$.

In the mass conservation equation (7.32)–(7.33), the dependent variable for mass is p_s. This quantity indicates the change in hydrostatic pressure, and thus the amount of mass, over the vertical distance between nearby surfaces of constant s. In the special case where s is the elevation z, $p_s = p_z = -\rho g$, and in that case the density ρ can be used in place of p_s in (7.32)–(7.33).

7.8. Conservation of mass in layers

The role of $p_s = \partial p/\partial s$ as the mass variable can be illuminated by examining the conservation of mass in a layer bounded above and below by surfaces of constant s. The following analysis also anticipates the vertical discretization that is needed for solving the governing equations numerically.

Consider the fluid lying between the coordinate surfaces $s = s_0$ and $s = s_1$, where $s_0 < s_1$. For any $\mathbf{x}$ and t, let

$$\Delta p(\mathbf{x}, t) = p(\mathbf{x}, s_0, t) - p(\mathbf{x}, s_1, t) = \int_{s_0}^{s_1} (-p_s)\, \mathrm{d}s > 0. \tag{7.35}$$

Due to the hydrostatic assumption, $\Delta p(\mathbf{x}, t)$ represents the weight per unit horizontal area in this layer, at horizontal position $\mathbf{x}$ at time t. Therefore Δp equals the mass per unit horizontal area in the layer, times g, so Δp can be regarded as a two-dimensional density when the layer is viewed from above. In an isopycnic setting this quantity indicates the thickness of the layer, and Δp is often labelled informally as the 'layer thickness'.

Also define the mass-weighted vertical average of $\dot{\mathbf{x}}$ over the layer by

$$\overline{\dot{\mathbf{x}}}(\mathbf{x}, t) = \frac{1}{\Delta p} \int_{s_0}^{s_1} \dot{\mathbf{x}}(\mathbf{x}, s, t)(-p_s)\, \mathrm{d}s, \tag{7.36}$$

and use the approximation $G(\mathbf{x}) = m_1(\mathbf{x})m_2(\mathbf{x})$. Now multiply the mass equation (7.33) by -1 and integrate over s from s_0 to s_1 to obtain

$$\frac{\partial}{\partial t}\big(G\Delta p\big) + \nabla \cdot \big(G\bar{\dot{\mathbf{x}}}\Delta p\big) + G\Big[\big(\dot{s}p_s\big)_{s=s_0} - \big(\dot{s}p_s\big)_{s=s_1}\Big] = 0. \tag{7.37}$$

The independent variables in (7.37) are the horizontal coordinates x_1 and x_2 and the time t.

Equation (7.37) can be expressed in terms of linear velocity as follows. In Section 7.9 it is shown that the horizontal components of linear velocity in the x_1- and x_2-directions, relative to the rotating spheroid Σ, are $u(\mathbf{x}, s, t) = m_1\dot{x}_1(\mathbf{x}, s, t)$ and $v(\mathbf{x}, s, t) = m_2\dot{x}_2(\mathbf{x}, s, t)$, respectively. Let $\bar{u}(\mathbf{x}, t)$ and

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

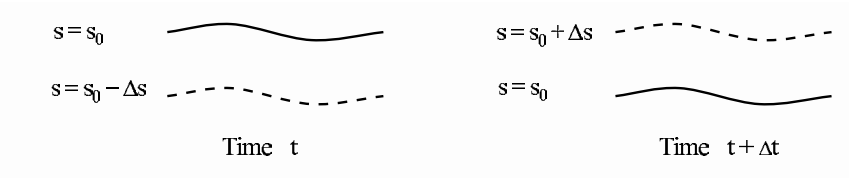


Figure 7.2. Illustration of mass transport across coordinate surfaces due to material changes in the vertical coordinate s. If s is a thermodynamic variable, then a warming or cooling of the fluid could cause changes in s for given fluid parcels. Fluid can then cross coordinate surfaces, even if the fluid is motionless in space. Similarly, if s is a hybrid coordinate, then such transport could occur during a conversion between isopycnic and geometric coordinates.

$\bar{v}(\mathbf{x}, t)$ denote the mass-weighted vertical averages of u and v, respectively, over the layer. Then $\bar{u}(\mathbf{x}, t) = m_1 \bar{\dot{x}}_1$ and $\bar{v}(\mathbf{x}, t) = m_2 \bar{\dot{x}}_2$, and equation (7.37) can be written as

$$\begin{aligned} \frac{\partial}{\partial t}\big(G\Delta p\big) + \frac{\partial}{\partial x_1}\big(m_2 \bar{u} \Delta p\big) + \frac{\partial}{\partial x_2}\big(m_1 \bar{v} \Delta p\big) \\ + G\Big[\big(\dot{s} p_s\big)_{s=s_0} - \big(\dot{s} p_s\big)_{s=s_1}\Big] = 0. \end{aligned} \tag{7.38}$$

Equation (7.38) can be interpreted as follows. Integrate (7.38) with respect to x_1 and x_2 on a rectangle R (in the $\mathbf{x}$ domain) having sides Δx_1 and Δx_2. The quantities $m_1 \Delta x_1$ and $m_2 \Delta x_2$ give the (approximate) linear dimensions of the corresponding rectangular region $\tilde{R}$ on the surface of the rotating spheroid Σ. The area of $\tilde{R}$ is approximately $m_1 m_2 \Delta x_1 \Delta x_2 = G \Delta x_1 \Delta x_2$, so G is the factor that relates horizontal 'area' in parameter space to physical area on the surface of Σ. The quantity $G \Delta x_1 \Delta x_2 \Delta p$ is the mass in the given layer in $\tilde{R}$, times g, and the integrals on R of the terms in (7.38) involving $\partial/\partial x_1$ and $\partial/\partial x_2$ can be interpreted in terms of lateral mass transport across the edges of $\tilde{R}$.

The terms in (7.38) involving $\dot{s} p_s$ are less standard and represent transport between layers due to material changes in the coordinate s. Consider the coordinate surface $s = s_0$, and for definiteness assume that $\dot{s} > 0$ on that surface. The quantity $\dot{s}$ is the time derivative of s as seen by an observer that is fixed relative to the fluid, and nonzero values of $\dot{s}$ may or may not be due to any motion of the fluid. For example, if s is the reciprocal of potential density or a related quantity, then a warming of the fluid could cause $\dot{s} > 0$, even if the fluid is motionless relative to Σ. Over a time

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

increment Δt, the value of s seen by a fluid parcel changes by approximately the amount $\Delta s = \dot{s}\Delta t$. Now consider the two coordinate surfaces $s = s_0$ and $s = s_0 - \Delta s$ at time t. From time t to time $t + \Delta t$, the fluid parcels on these surfaces experience a change from s_0 and $s_0 - \Delta s$ to $s_0 + \Delta s$ and s_0, respectively (see Figure 7.2). The coordinate surface $s = s_0$ thus moves downward relative to the fluid, but an observer on that surface sees fluid crossing the surface in the upward direction. For the fluid that crosses the surface $s = s_0$ during the time increment Δt, the weight per unit area is approximately $\Delta s(-\partial p/\partial s) = \dot{s}\Delta t(-\partial p/\partial s)$, so the rate of mass flow per unit time per unit horizontal area (times g) is $-\dot{s}p_s$. Here, a positive rate of flow indicates an upward movement of fluid relative to the coordinate surface. The last two terms in equation (7.38) represent the net effect of transport into or out of the layer by this mechanism.

7.9. Velocity

In preparation for deriving the equation for conservation of momentum, the present subsection develops a representation for fluid velocity. This velocity is expressed as a motion relative to the rotating spheroid Σ, plus a rigid-body rotation of points attached to Σ.

As in Section 7.5, let $(\mathbf{X}, S)$ denote the position of a fluid parcel in parameter space at time 0, and denote the position of that parcel in parameter space at any time t by $\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t)\big)$. The position of that parcel, relative to the rectangular coordinate system in the inertial reference frame, can then be written in the form

$$\mathbf{R}(\mathbf{X}, S, t) = \mathbf{r}\big(\mathbf{x}(\mathbf{X}, S, t), \hat{z}(\mathbf{X}, S, t), t\big).$$

Here, $\mathbf{r}(\mathbf{x}, z, t)$ denotes the position in rectangular coordinates of the point with coordinates $(\mathbf{x}, z)$ relative to Σ, as defined in Section 7.2, and $\hat{z}(\mathbf{X}, S, t)$ denotes the elevation of the fluid parcel in question. More precisely,

$$\hat{z}(\mathbf{X}, S, t) = z\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big),$$

where $z(\mathbf{x}, s, t)$ is the elevation of the coordinate surface as defined in Section 7.3.

The velocity of a fluid parcel, as seen in rectangular coordinates in the inertial frame, is then

$$\frac{\partial \mathbf{R}}{\partial t}(\mathbf{X}, S, t) = \frac{\partial \mathbf{r}}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial \mathbf{r}}{\partial x_2}\frac{\partial x_2}{\partial t} + \frac{\partial \mathbf{r}}{\partial z}\frac{\partial \hat{z}}{\partial t} + \frac{\partial \mathbf{r}}{\partial t}. \tag{7.39}$$

Here, the partial derivatives of $\mathbf{r}$ are evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), \hat{z}(\mathbf{X}, S, t), t\big)$, and the partial derivatives of x_1, x_2, and $\hat{z}$ are evaluated at $(\mathbf{X}, S, t)$. The relations (7.4) and (7.5) involving $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, and the definitions (7.21)

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

of $\dot{\mathbf{x}}$ and $\dot{s}$, imply

$$\frac{\partial \mathbf{R}}{\partial t}(\mathbf{X}, S, t) = m_1 \dot{x}_1 \mathbf{i} + m_2 \dot{x}_2 \mathbf{j} + \frac{\partial \hat{z}}{\partial t}\mathbf{k} + \frac{\partial \mathbf{r}}{\partial t}. \tag{7.40}$$

In (7.40), m_1 and m_2 are evaluated at the point $\big(\mathbf{x}(\mathbf{X}, S, t), \hat{z}(\mathbf{X}, S, t)\big)$; $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), t\big)$; and $\dot{x}_1$ and $\dot{x}_2$ are evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big)$.

In equations (7.39) and (7.40), the partial derivative $\partial \mathbf{r}/\partial t$ is taken for fixed position $(\mathbf{x}, z)$ relative to the rotating surface Σ, and it therefore represents the velocity associated with a rigid-body rotation about the given axis. The remaining terms in (7.39) and (7.40) involve derivatives of $\mathbf{r}$ with respect to position on Σ. They thus represent motion relative to Σ, and their sum is termed the 'relative velocity'.

For the horizontal components of the relative velocity, define

$$\begin{aligned} u(\mathbf{x}, s, t) &= m_1\big(\mathbf{x}, z(\mathbf{x}, s, t)\big)\ \dot{x}_1(\mathbf{x}, s, t), \\ v(\mathbf{x}, s, t) &= m_2\big(\mathbf{x}, z(\mathbf{x}, s, t)\big)\ \dot{x}_2(\mathbf{x}, s, t) \end{aligned} \tag{7.41}$$

and define a vertical component $w(\mathbf{x}, s, t)$ by

$$\frac{\partial \tilde{z}}{\partial t}(\mathbf{X}, S, t) = w\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big). \tag{7.42}$$

The particle velocity (7.40) can then be written as

$$\frac{\partial \mathbf{R}}{\partial t}(\mathbf{X}, S, t) = u\mathbf{i} + v\mathbf{j} + w\mathbf{k} + \frac{\partial \mathbf{r}}{\partial t}, \tag{7.43}$$

where the components u, v, and w are evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big)$; the unit vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), t\big)$; and the rigid-body velocity $\partial \mathbf{r}/\partial t$ is evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), \hat{z}(\mathbf{X}, S, t), t\big)$.

The terms $u\mathbf{i}$ and $v\mathbf{j}$ are defined in terms of $\partial \mathbf{r}/\partial x_1$ and $\partial \mathbf{r}/\partial x_2$, via (7.39). Given the definition of the position vector $\mathbf{r}$, these partial derivatives are taken for fixed z, so the vectors $\mathbf{i}$ and $\mathbf{j}$ are horizontal, *i.e.*, tangent to Σ. The quantities u and v thus represent components of velocity that are truly horizontal, not components along surfaces of constant s. This is consistent with the observation, made at the end of Section 7.3, that lateral displacements are measured in terms of projections onto the horizontal instead of along s-coordinate surfaces.

7.10. Momentum

In the present subsection we formulate the momentum of a volume of fluid and calculate its time derivative.

Assume that a volume of fluid occupies a region $A(t)$ in parameter space at time t, and let $B(t)$ denote the corresponding region in rectangular

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

coordinates in the inertial reference frame. Also let $\mathbf{Q}(t)$ denote the total momentum of this volume of fluid, as seen in the inertial frame. Then $\mathbf{Q}(t)$ is the integral on $B(t)$ of the velocity times mass density; the change of variables given in (7.19) implies

$$\begin{aligned}\mathbf{Q}(t) &= \int_{A(t)} \rho(\mathbf{x},s,t)\left[u\mathbf{i}+v\mathbf{j}+w\mathbf{k}+\frac{\partial \mathbf{r}}{\partial t}\right] J(\mathbf{x},s,t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}s \qquad (7.44)\\ &= \int_{A(0)} \rho(\mathbf{x},s,t)\left[u\mathbf{i}+v\mathbf{j}+w\mathbf{k}+\frac{\partial \mathbf{r}}{\partial t}\right] J(\mathbf{x},s,t)\ H(\mathbf{X},S,t)\,\mathrm{d}\mathbf{X}\,\mathrm{d}S.\end{aligned}$$

Equation (7.30) and subsequent discussions imply $\rho J = -m_1 m_2 p_s g^{-1} = -G p_s g^{-1}$, so

$$\mathbf{Q}(t) = -\frac{1}{g}\int_{A(0)}\left[\big(Gup_s\big)\mathbf{i}+\big(Gvp_s\big)\mathbf{j}+\big(Gwp_s\big)\mathbf{k}+Gp_s\frac{\partial \mathbf{r}}{\partial t}\right] H(\mathbf{X},S,t)\,\mathrm{d}\mathbf{X}\,\mathrm{d}S. \qquad (7.45)$$

Here, G, u, v, w, p_s are evaluated at $\big(\mathbf{x}(\mathbf{X},S,t),s(\mathbf{X},S,t),t\big)$; $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are evaluated at $\big(\mathbf{x}(\mathbf{X},S,t),t\big)$; $\partial\mathbf{r}/\partial t$ is evaluated at $\big(\mathbf{x}(\mathbf{X},S,t),\hat{z}(\mathbf{X},S,t),t\big)$; and $H(\mathbf{X},S,t)$ is the Jacobian from $(\mathbf{X},S)$ to $(\mathbf{x},s)$ defined in (7.20).

The next main task is to calculate the time derivative of (7.45). Several terms in the integrand have the form

$$\hat{F}(\mathbf{X},S,t) = F\big(\mathbf{x}(\mathbf{X},S,t),s(\mathbf{X},S,t),t\big)\ H(\mathbf{X},S,t),$$

and some calculations similar to those in (7.25)–(7.27) imply

$$\frac{\partial \hat{F}}{\partial t}(\mathbf{X},S,t) = D_f(F)\ H(\mathbf{X},S,t),$$

where the flux derivative

$$\begin{aligned}D_f(F) &= \frac{\partial F}{\partial t}+\frac{\partial}{\partial x_1}\big(\dot{x}_1 F\big)+\frac{\partial}{\partial x_2}\big(\dot{x}_2 F\big)+\frac{\partial}{\partial s}\big(\dot{s}F\big) \qquad (7.46)\\ &= \frac{\partial F}{\partial t}+\nabla\cdot\big(\dot{\mathbf{x}}F\big)+\frac{\partial}{\partial s}\big(\dot{s}F\big)\end{aligned}$$

is evaluated at $\big(\mathbf{x}(\mathbf{X},S,t),s(\mathbf{X},S,t),t\big)$. It then follows that

$$\begin{aligned}\mathbf{Q}'(t) = -\frac{1}{g}\int_{A(0)}\Big\{&D_f\big(Gup_s\big)\mathbf{i}+D_f\big(Gvp_s\big)\mathbf{j}+D_f\big(Gwp_s\big)\mathbf{k}\\ &+\big(Gup_s\big)\frac{D\mathbf{i}}{Dt}+\big(Gvp_s\big)\frac{D\mathbf{j}}{Dt}+\big(Gwp_s\big)\frac{D\mathbf{k}}{Dt} \qquad (7.47)\\ &+Gp_s\frac{\partial}{\partial t}\left[\frac{\partial \mathbf{r}}{\partial t}\big(\mathbf{x}(\mathbf{X},S,t),\hat{z}(\mathbf{X},S,t),t\big)\right]\Big\}\\ &H(\mathbf{X},S,t)\,\mathrm{d}\mathbf{X}\,\mathrm{d}S.\end{aligned}$$

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

The derivation of (7.47) uses the relation $D_f\big(Gp_s\big) = 0$, which is a way of stating the equation (7.32)–(7.33) of conservation of mass. In (7.47), the notation $D\mathbf{i}/Dt$ refers to the time derivative of $\mathbf{i}\big(\mathbf{x}(\mathbf{X}, S, t), t\big)$, so

$$\frac{D\mathbf{i}}{Dt} = \frac{\partial \mathbf{i}}{\partial t} + \dot{x}_1 \frac{\partial \mathbf{i}}{\partial x_1} + \dot{x}_2 \frac{\partial \mathbf{i}}{\partial x_2}, \tag{7.48}$$

where $\dot{x}_1$ and $\dot{x}_2$ are evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big)$. The terms $D\mathbf{j}/Dt$ and $D\mathbf{k}/Dt$ are analogous.

The time derivative in the third line of (7.47) is

$$\frac{\partial^2 \mathbf{r}}{\partial t \partial x_1} \dot{x}_1 + \frac{\partial^2 \mathbf{r}}{\partial t \partial x_2} \dot{x}_2 + \frac{\partial^2 \mathbf{r}}{\partial t \partial z} \frac{\partial \hat{z}}{\partial t} + \frac{\partial^2 \mathbf{r}}{\partial t^2}, \tag{7.49}$$

where the partial derivatives of $\mathbf{r}$ are evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), \hat{z}(\mathbf{X}, S, t), t\big)$, $\dot{x}_1$ and $\dot{x}_2$ are evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big)$, and $\partial \hat{z}/\partial t$ is evaluated at $(\mathbf{X}, S, t)$. According to the approximation introduced in (7.5),

$$\begin{aligned} \frac{\partial^2 \mathbf{r}}{\partial x_1 \partial t}(\mathbf{x}, z, t) &= m_1(\mathbf{x}, z)\ \frac{\partial \mathbf{i}}{\partial t}(\mathbf{x}, t), \\ \frac{\partial^2 \mathbf{r}}{\partial x_2 \partial t}(\mathbf{x}, z, t) &= m_2(\mathbf{x}, z)\ \frac{\partial \mathbf{j}}{\partial t}(\mathbf{x}, t), \end{aligned} \tag{7.50}$$

and the definition (7.4) of $\mathbf{k}$ implies

$$\frac{\partial^2 \mathbf{r}}{\partial z \partial t}(\mathbf{x}, z, t) = \frac{\partial \mathbf{k}}{\partial t}(\mathbf{x}, t). \tag{7.51}$$

The definitions (7.41) and (7.42) of the velocity components u, v, and w imply that the expression (7.49) is equal to

$$u \frac{\partial \mathbf{i}}{\partial t} + v \frac{\partial \mathbf{j}}{\partial t} + w \frac{\partial \mathbf{k}}{\partial t} + \mathbf{a}_c, \tag{7.52}$$

where $\mathbf{a}_c = \partial^2 \mathbf{r}/\partial t^2$; u, v, and w are evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), s(\mathbf{X}, S, t), t\big)$; and the derivatives of $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are evaluated at $\big(\mathbf{x}(\mathbf{X}, S, t), t\big)$.

The quantity $\partial^2 \mathbf{r}/\partial t^2(\mathbf{x}, z, t)$ is the second-order time derivative of the position vector $\mathbf{r}$, for a fixed position $(\mathbf{x}, z)$ relative to the rotating spheroid Σ. The vector $\mathbf{a}_c$ is thus the centripetal acceleration associated with a rigid-body rotation. For present purposes, $\mathbf{a}_c$ will be regarded as a function of $(\mathbf{x}, s, t)$, where $\mathbf{x}$ and s are evaluated at $(\mathbf{X}, S, t)$ in (7.52).

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Equation (7.47) can now be written as

$$\mathbf{Q}'(t) = -\frac{1}{g}\int_{A(t)} \Big[D_f\big(Gup_s\big)\mathbf{i} + D_f\big(Gvp_s\big)\mathbf{j} + D_f\big(Gwp_s\big)\mathbf{k} + Gp_s\mathbf{a}_c + \mathbf{\Psi}\Big]\,\mathrm{d}\mathbf{x}\,\mathrm{d}s \tag{7.53}$$

where

$$\begin{aligned}\mathbf{\Psi} = {} & 2\big(Gup_s\big)\frac{\partial \mathbf{i}}{\partial t} + 2\big(Gvp_s\big)\frac{\partial \mathbf{j}}{\partial t} + 2\big(Gwp_s\big)\frac{\partial \mathbf{k}}{\partial t} \\ & + \big(Gup_s\big)\left(\dot{x}_1\frac{\partial \mathbf{i}}{\partial x_1} + \dot{x}_2\frac{\partial \mathbf{i}}{\partial x_2}\right) \\ & + \big(Gvp_s\big)\left(\dot{x}_1\frac{\partial \mathbf{j}}{\partial x_1} + \dot{x}_2\frac{\partial \mathbf{j}}{\partial x_2}\right) \\ & + \big(Gwp_s\big)\left(\dot{x}_1\frac{\partial \mathbf{k}}{\partial x_1} + \dot{x}_2\frac{\partial \mathbf{k}}{\partial x_2}\right).\end{aligned} \tag{7.54}$$

In equations (7.53) and (7.54), $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ and their derivatives depend on $(\mathbf{x}, t)$, and all other quantities depend on $(\mathbf{x}, s, t)$.

The time derivatives $\partial\mathbf{i}/\partial t$, $\partial\mathbf{j}/\partial t$, and $\partial\mathbf{k}/\partial t$ in (7.54) are taken with the position fixed relative to Σ. The terms containing these derivatives are therefore due to the rotation of Σ. The remaining terms in (7.54) involve the variation of the unit vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ with respect to position on Σ, so these terms are due to the curvature of Σ and/or properties of the parametrization of Σ in terms of $\mathbf{x} = (x_1, x_2)$.

According to the principle of conservation of momentum, $\mathbf{Q}'(t)$ is equal to the sum of the various forces acting on the fluid volume. These forces are due to pressure, gravity, stresses, and viscosity, and they will be discussed in the next subsections.

7.11. Pressure and the Montgomery potential

As in the preceding subsection, assume that a volume of fluid occupies a region $A(t)$ in parameter space at time t, and let $B(t)$ denote the corresponding region in rectangular coordinates in the inertial reference frame. Also, as in Section 7.3, let $P(\mathbf{x}, z, t)$ denote the pressure in the fluid corresponding to horizontal position $\mathbf{x}$, elevation z, and time t. Now let $\tilde{P}(\mathbf{r}, t)$ denote the pressure as seen in rectangular coordinates in the inertial frame, *i.e.*, $P(\mathbf{x}, z, t) = \tilde{P}\big(\mathbf{r}(\mathbf{x}, z, t), t\big)$.

Let $\partial B(t)$ denote the boundary of the region $B(t)$ in rectangular coordinates, and let $\mathbf{n}(\mathbf{r}, t)$ denote the unit outward normal vector to $\partial B(t)$. The

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

net pressure force acting on the volume of fluid is then

$$\begin{aligned}\int_{\partial B(t)} -\tilde{P}\mathbf{n}\,\mathrm{d}A &= -\int_{\partial B(t)} \big(\tilde{P}n_1, \tilde{P}n_2, \tilde{P}n_3\big)\,\mathrm{d}A \\ &= -\int_{\partial B(t)} \big((\tilde{P},0,0)\cdot\mathbf{n},\ (0,\tilde{P},0)\cdot\mathbf{n},\ (0,0,\tilde{P})\cdot\mathbf{n}\big)\,\mathrm{d}A \\ &= -\int_{B(t)} \big(\mathrm{div}(\tilde{P},0,0),\ \mathrm{div}(0,\tilde{P},0),\ \mathrm{div}(0,0,\tilde{P})\big)\,\mathrm{d}V_{B(t)} \\ &= -\int_{B(t)} \left(\frac{\partial\tilde{P}}{\partial r_1},\ \frac{\partial\tilde{P}}{\partial r_2},\ \frac{\partial\tilde{P}}{\partial r_3}\right)\mathrm{d}V_{B(t)} \\ &= -\int_{A(t)} \nabla\tilde{P}\big(\mathbf{r}(\mathbf{x}, z(\mathbf{x},s,t),t),t\big)J(\mathbf{x},s,t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}s. \end{aligned} \tag{7.55}$$

Here, the integration is performed component-by-component, the divergence theorem is used on each component to obtain the third line, and the change of variables described in Section 7.4 is used to obtain the last line. In the last line, the notation $\nabla\tilde{P}\big(\mathbf{r}\big(\mathbf{x}, z(\mathbf{x},s,t),t\big),t\big)$ refers to the value of $\nabla\tilde{P}$ at the indicated point; it does not indicate the derivative of a composite function.

The integral (7.55) can also be be expressed as

$$\int_{\partial B(t)} -\tilde{P}\mathbf{n}\,\mathrm{d}A = -\int_{A(t)} \big((\nabla\tilde{P}\cdot\mathbf{i})\mathbf{i} + (\nabla\tilde{P}\cdot\mathbf{j})\mathbf{j} + (\nabla\tilde{P}\cdot\mathbf{k})\mathbf{k}\big)J(\mathbf{x},s,t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}s, \tag{7.56}$$

where $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are evaluated at $(\mathbf{x},t)$. The relation (7.5) implies

$$\begin{aligned}\nabla\tilde{P}\cdot\mathbf{i} &= \frac{1}{m_1}\nabla\tilde{P}\big(\mathbf{r}\big(\mathbf{x}, z(\mathbf{x},s,t),t\big),t\big)\cdot\frac{\partial\mathbf{r}}{\partial x_1}(\mathbf{x}, z(\mathbf{x},s,t),t) \\ &= \frac{1}{m_1}\frac{\partial P}{\partial x_1}\big(\mathbf{x}, z(\mathbf{x},s,t),t\big),\end{aligned} \tag{7.57}$$

where m_1 is evaluated at $\big(\mathbf{x}, z(\mathbf{x},s,t)\big)$. Given the definition of the function P, the partial derivative $\partial P/\partial x_1$ is taken for fixed z and then evaluated at $\big(\mathbf{x}, z(\mathbf{x},s,t),t\big)$. Similarly,

$$\nabla\tilde{P}\cdot\mathbf{j} = \frac{1}{m_2}\frac{\partial P}{\partial x_2}\big(\mathbf{x}, z(\mathbf{x},s,t),t\big), \tag{7.58}$$

and the definition (7.4) of $\mathbf{k}$ implies

$$\nabla\tilde{P}\cdot\mathbf{k} = \frac{\partial P}{\partial z}\big(\mathbf{x}, z(\mathbf{x},s,t),t\big). \tag{7.59}$$

Now express the preceding in terms of functions of $(\mathbf{x},s,t)$, as these will be the independent variables in the final form of the governing equations.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

As in Section 7.3, let $p(\mathbf{x}, s, t)$ denote the pressure in terms of s, so that $p(\mathbf{x}, s, t) = P\big(\mathbf{x}, z(\mathbf{x}, s, t), t\big)$. Then, for $i = 1$ and $i = 2$,

$$\begin{aligned}\frac{\partial p}{\partial x_i}(\mathbf{x}, s, t) &= \frac{\partial P}{\partial x_i} + \frac{\partial P}{\partial z}\frac{\partial z}{\partial x_i}(\mathbf{x}, s, t) \\ &= \frac{\partial P}{\partial x_i} - \frac{g}{\alpha(\mathbf{x}, s, t)}\frac{\partial z}{\partial x_i}(\mathbf{x}, s, t),\end{aligned} \tag{7.60}$$

where $\alpha(\mathbf{x}, s, t)$ is the specific volume (reciprocal of density). The second line is obtained by using the hydrostatic condition (7.8). Equation (7.60) then implies

$$\begin{aligned}\alpha(\mathbf{x}, s, t)\frac{\partial P}{\partial x_i}\big(\mathbf{x}, z(\mathbf{x}, s, t), t\big) &= \alpha\frac{\partial p}{\partial x_i} + g\frac{\partial z}{\partial x_i} \\ &= \frac{\partial M}{\partial x_i} - p\frac{\partial \alpha}{\partial x_i},\end{aligned} \tag{7.61}$$

where

$$M(\mathbf{x}, s, t) = \alpha(\mathbf{x}, s, t)p(\mathbf{x}, s, t) + gz(\mathbf{x}, s, t) \tag{7.62}$$

is the Montgomery potential (Montgomery 1937). The hydrostatic condition (7.8) also implies

$$\alpha(\mathbf{x}, s, t)\frac{\partial P}{\partial z}\big(\mathbf{x}, z(\mathbf{x}, s, t), t\big) = -g.$$

The pressure force (7.56) can then be expressed in the form

$$\begin{aligned}\int_{\partial B(t)} -\tilde{P}\mathbf{n}\,\mathrm{d}A = -\int_{A(t)} \Bigg[&\frac{1}{m_1}\left(\frac{\partial M}{\partial x_1} - p\frac{\partial \alpha}{\partial x_1}\right)\mathbf{i} \\ &+ \frac{1}{m_2}\left(\frac{\partial M}{\partial x_2} - p\frac{\partial \alpha}{\partial x_2}\right)\mathbf{j} - g\mathbf{k}\Bigg]\frac{J(\mathbf{x}, s, t)}{\alpha(\mathbf{x}, s, t)}\,\mathrm{d}\mathbf{x}\,\mathrm{d}s,\end{aligned} \tag{7.63}$$

where m_1 and m_2 are evaluated at $\big(\mathbf{x}, z(\mathbf{x}, s, t)\big)$. According to equations (7.31) and (7.34), $J/\alpha = J\rho = -m_1 m_2 p_s g^{-1} = -Gp_s g^{-1}$, and (7.63) becomes

$$\begin{aligned}\int_{\partial B(t)} -\tilde{P}\mathbf{n}\,\mathrm{d}A = \frac{1}{g}\int_{A(t)} \Bigg[&\frac{1}{m_1}\left(\frac{\partial M}{\partial x_1} - p\frac{\partial \alpha}{\partial x_1}\right)\mathbf{i} \\ &+ \frac{1}{m_2}\left(\frac{\partial M}{\partial x_2} - p\frac{\partial \alpha}{\partial x_2}\right)\mathbf{j} - g\mathbf{k}\Bigg]Gp_s\,\mathrm{d}\mathbf{x}\,\mathrm{d}s.\end{aligned} \tag{7.64}$$

In the earlier expression (7.56) for the pressure force, the coefficients of $\mathbf{i}$ and $\mathbf{j}$ represent the components that are horizontal, *i.e.*, in directions tangent to the rotating spheroid Σ, and the derivatives in (7.57) and (7.58) are taken for fixed z. However, when the independent variables are $(\mathbf{x}, s, t)$, derivatives with respect to x_1 and x_2 are taken for fixed s, and these directions need not be horizontal. The usage of the Montgomery potential (7.62)

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

enables derivatives for fixed s to yield components of the pressure force that are in the directions of the horizontal vectors $\mathbf{i}$ and $\mathbf{j}$.

The vertical variation of the Montgomery potential can be determined by observing

$$\begin{aligned}\frac{\partial M}{\partial s}(\mathbf{x},s,t) &= p\frac{\partial \alpha}{\partial s} + \alpha\frac{\partial p}{\partial s} + g\frac{\partial z}{\partial s} \\ &= p\frac{\partial \alpha}{\partial s}.\end{aligned} \tag{7.65}$$

The condition $\alpha p_s + g z_s = 0$ follows from equation (7.9), which in turn follows from the hydrostatic condition (7.8). Equation (7.65) can be regarded as a statement of the hydrostatic condition in a generalized vertical coordinate.

7.12. Diffusion of momentum

One mechanism for the transport of momentum within a fluid is the mixing caused by small-scale turbulence and molecular diffusion. In the ocean the effects of molecular diffusion are negligible compared to the effects of turbulence, so molecular diffusion will not be considered here. In numerical models of ocean circulation it is generally not possible to represent the details of turbulent motions, owing to insufficient grid resolution, so instead it is necessary to parametrize the large-scale effects of these subgrid-scale motions in terms of the dependent variables that are used in the model. This problem has long been an active area of research; see, *e.g.*, Griffies (2004) or Pedlosky (1987). In present-day practice, operational ocean models often use a parametrization that involves some sort of Laplacian or biharmonic diffusion.

As noted in Section 2.4, mixing in the ocean's interior is highly anisotropic, with the principle directions of mixing being approximately horizontal. In the present subsection, we assume that the vertical coordinate s is an isopycnic coordinate and that the mixing occurs primarily along surfaces of constant s. In addition, we represent the subgrid-scale processes in terms of a diffusion that is proportional to the gradient of velocity, in analogy to molecular diffusion, and the diffusion is separated into a component tangent to coordinate surfaces and a component normal to such surfaces. The following analysis produces a parametrization similar to the one used, for example, by Bleck and Smith (1990) and Bleck (2002) for isopycnic models.

As in the preceding two subsections, assume that a volume of fluid occupies a region $A(t)$ in parameter space at time t, and denote by $B(t)$ the corresponding region in rectangular coordinates in the inertial reference frame. For each $(\mathbf{x},s) \in A(t)$, let $\tilde{\mathbf{r}}(\mathbf{x},s,t) = \mathbf{r}\big(\mathbf{x}, z(\mathbf{x},s,t), t\big)$ denote the corresponding point in $B(t)$, as in equation (7.11).

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

In order to formulate the highly anisotropic diffusion considered here, it is useful to define a local coordinate system, at each point in the fluid, that involves a plane tangent to a surface of constant s and a vector normal to that surface. A starting point is provided by the tangent vectors $\partial\tilde{\mathbf{r}}/\partial x_1$ and $\partial\tilde{\mathbf{r}}/\partial x_2$. Equations (7.11), (7.4), and (7.5) imply

$$\begin{aligned}\frac{\partial\tilde{\mathbf{r}}}{\partial x_1}(\mathbf{x},s,t) &= \frac{\partial\mathbf{r}}{\partial x_1}+\frac{\partial\mathbf{r}}{\partial z}\frac{\partial z}{\partial x_1}\\ &= m_1\mathbf{i}+\frac{\partial z}{\partial x_1}\mathbf{k}\\ &= m_1\big(\mathbf{i}+\delta_1\mathbf{k}\big),\end{aligned}$$

where $\delta_1(\mathbf{x},s,t)$ is the slope of a surface of constant s in the x_1-direction, m_1 is evaluated at $\big(\mathbf{x},z(\mathbf{x},s,t)\big)$, and $\mathbf{i}$ and $\mathbf{k}$ are evaluated at $(\mathbf{x},t)$. Similarly,

$$\frac{\partial\tilde{\mathbf{r}}}{\partial x_2}(\mathbf{x},s,t)=m_2\big(\mathbf{j}+\delta_2\mathbf{k}\big),$$

where $\delta_2(\mathbf{x},s,t)$ is the slope of a surface of constant s in the x_2-direction. Unit tangent vectors $\tilde{\mathbf{i}}$ and $\tilde{\mathbf{j}}$ to a coordinate surface are then defined by

$$\begin{aligned}\frac{\partial\tilde{\mathbf{r}}}{\partial x_1}(\mathbf{x},s,t)&=\tilde{m}_1(\mathbf{x},s,t)\frac{\mathbf{i}+\delta_1\mathbf{k}}{\sqrt{1+\delta_1^2}}\equiv\tilde{m}_1(\mathbf{x},s,t)\;\tilde{\mathbf{i}}(\mathbf{x},s,t),\\ \frac{\partial\tilde{\mathbf{r}}}{\partial x_2}(\mathbf{x},s,t)&=\tilde{m}_2(\mathbf{x},s,t)\frac{\mathbf{j}+\delta_2\mathbf{k}}{\sqrt{1+\delta_2^2}}\equiv\tilde{m}_2(\mathbf{x},s,t)\;\tilde{\mathbf{j}}(\mathbf{x},s,t),\end{aligned}$$

where

$$\tilde{m}_i(\mathbf{x},s,t)=\left|\frac{\partial\tilde{\mathbf{r}}}{\partial x_i}(\mathbf{x},s,t)\right|=m_i\sqrt{1+\delta_i^2}=m_i\big(1+O(\delta^2)\big)\qquad(7.66)$$

for $i=1,2$. Here, δ is an upper bound for $|\delta_1|$ and $|\delta_2|$. In the interior of the ocean, the slopes of isopycnals generally do not reach much above 10^{-2} (Griffies *et al.* 2000*a*), so typically $\delta\approx 10^{-2}$. The quantity $\tilde{m}_i$ can be regarded as a metric coefficient that relates increments in the parameter x_i to spatial increments along s-coordinate surfaces, whereas the metric coefficient m_i relates increments in x_i to spatial increments that are horizontal.

For any $(\mathbf{x},s,t)$, the vectors $\tilde{\mathbf{i}}(\mathbf{x},s,t)$ and $\tilde{\mathbf{j}}(\mathbf{x},s,t)$ provide a basis for the tangent plane to the coordinate surface at that position and time. However, these vectors are not exactly orthogonal, as $\tilde{\mathbf{i}}\cdot\tilde{\mathbf{j}}=O(\delta^2)$. Orthogonal unit vectors can be obtained by rotating $\tilde{\mathbf{i}}$ and/or $\tilde{\mathbf{j}}$ through an angle of magnitude $O(\delta^2)$ in that plane. This process is not uniquely determined, but assume

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

that orthogonal unit vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ in that plane have been obtained so that

$$\begin{aligned}\hat{\mathbf{i}}(\mathbf{x},s,t) &= \tilde{\mathbf{i}} + \mathbf{O}(\delta^2) = \frac{1}{\tilde{m}_1(\mathbf{x},s,t)}\frac{\partial\tilde{\mathbf{r}}}{\partial x_1}(\mathbf{x},s,t) + \mathbf{O}(\delta^2),\\ \hat{\mathbf{j}}(\mathbf{x},s,t) &= \tilde{\mathbf{j}} + \mathbf{O}(\delta^2) = \frac{1}{\tilde{m}_2(\mathbf{x},s,t)}\frac{\partial\tilde{\mathbf{r}}}{\partial x_2}(\mathbf{x},s,t) + \mathbf{O}(\delta^2).\end{aligned} \tag{7.67}$$

The boldface font in the order terms is used to indicate vector quantities.

Now define a unit normal vector $\hat{\mathbf{k}}$ by

$$\begin{aligned}\hat{\mathbf{k}}(\mathbf{x},s,t) &= \hat{\mathbf{i}}\times\hat{\mathbf{j}}\\ &= \left[\frac{\mathbf{i}+\delta_1\mathbf{k}}{\sqrt{1+\delta_1^2}} + \mathbf{O}(\delta^2)\right]\times\left[\frac{\mathbf{j}+\delta_2\mathbf{k}}{\sqrt{1+\delta_2^2}} + \mathbf{O}(\delta^2)\right]\\ &= \mathbf{k} - \delta_1\mathbf{i} - \delta_2\mathbf{j} + \mathbf{O}(\delta^2).\end{aligned} \tag{7.68}$$

Here, it is assumed that the coordinate system on the rotating spheroid Σ is right-handed, in the sense that $\mathbf{i}\times\mathbf{j} = \mathbf{k}$, $\mathbf{k}\times\mathbf{j} = -\mathbf{i}$, and $\mathbf{i}\times\mathbf{k} = -\mathbf{j}$. Equation (7.11), $\tilde{\mathbf{r}}(\mathbf{x},s,t) = \mathbf{r}\big(\mathbf{x},z(\mathbf{x},s,t),t\big)$, together with equations (7.4) and (7.10), imply

$$\frac{\partial\tilde{\mathbf{r}}}{\partial s}(\mathbf{x},s,t) = \frac{\partial\mathbf{r}}{\partial z}\frac{\partial z}{\partial s} = m_s(\mathbf{x},s,t)\ \mathbf{k}(\mathbf{x},t).$$

It then follows from (7.68) that

$$\hat{\mathbf{k}}(\mathbf{x},s,t) = \frac{1}{m_s(\mathbf{x},s,t)}\frac{\partial\tilde{\mathbf{r}}}{\partial s}(\mathbf{x},s,t) + \mathbf{O}(\delta). \tag{7.69}$$

The representation of diffusion will initially be formulated in the region $B(t)$ in rectangular coordinates in the inertial reference frame. For that purpose, define the unit vectors $\hat{\mathbf{I}}$, $\hat{\mathbf{J}}$, and $\hat{\mathbf{K}}$ on $B(t)$ by

$$\begin{aligned}\hat{\mathbf{i}}(\mathbf{x},s,t) &= \hat{\mathbf{I}}\big(\tilde{\mathbf{r}}(\mathbf{x},s,t),t\big),\\ \hat{\mathbf{j}}(\mathbf{x},s,t) &= \hat{\mathbf{J}}\big(\tilde{\mathbf{r}}(\mathbf{x},s,t),t\big),\\ \hat{\mathbf{k}}(\mathbf{x},s,t) &= \hat{\mathbf{K}}\big(\tilde{\mathbf{r}}(\mathbf{x},s,t),t\big),\end{aligned}$$

for all $(\mathbf{x},s)\in A(t)$. Also define components U and V of horizontal velocity, with the independent spatial variables in $B(t)$, by

$$\begin{aligned}u(\mathbf{x},s,t) &= U\big(\tilde{\mathbf{r}}(\mathbf{x},s,t),t\big),\\ v(\mathbf{x},s,t) &= V\big(\tilde{\mathbf{r}}(\mathbf{x},s,t),t\big),\end{aligned} \tag{7.70}$$

for all $(\mathbf{x},s)\in A(t)$.

Now consider a parametrization of the diffusion of the component u. This will provide a term in the coefficient of $\mathbf{i}$ in equation (7.53), which gives the derivative $\mathbf{Q}'(t)$ of the momentum of the volume of fluid considered here. The diffusion of the component v is represented similarly. A component in

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

the direction of $\mathbf{k}$ will not be considered here, as the vertical component of momentum conservation will be represented with the hydrostatic balance.

Assume that the momentum flux associated with u, *i.e.*, the rate of transport of u-momentum per unit cross-sectional area per unit time, is given by

$$-\rho A_H\big(\nabla U\cdot\hat{\mathbf{I}}\big)\hat{\mathbf{I}}-\rho A_H\big(\nabla U\cdot\hat{\mathbf{J}}\big)\hat{\mathbf{J}}-\rho A_D\big(\nabla U\cdot\hat{\mathbf{K}}\big)\hat{\mathbf{K}}, \tag{7.71}$$

where ρ is the density of the fluid, A_H is a kinematic viscosity coefficient for diffusion in directions that are tangent to s-coordinate surfaces, and A_D is a kinematic viscosity for diapycnal diffusion normal to such surfaces. All quantities in (7.71) are evaluated at points $(\mathbf{r},t)$ for $\mathbf{r}\in B(t)$, except that A_H and A_D could perhaps be taken as constant, and $\nabla U = (\partial U/\partial r_1, \partial U/\partial r_2, \partial U/\partial r_3)$.

The net rate of diffusion of momentum into the region $B(t)$ is given by the integral of the negative of the outward normal component of (7.71) over $\partial B(t)$. This integral needs to be represented in terms of an integral over the region $A(t)$, as the quantity $\mathbf{Q}'(t)$ in (7.53) also involves an integral over that region. Toward that end, represent the boundary $\partial A(t)$ of $A(t)$ in terms of parameters $\sigma=(\sigma_1,\sigma_2)$ by

$$\mathbf{b}(\sigma,t)=\mathbf{b}(\sigma_1,\sigma_2,t)=\big(\mathbf{x}(\sigma,t),s(\sigma,t)\big) \tag{7.72}$$

for all σ in a parameter region Γ. The boundary $\partial B(t)$ of $B(t)$ is then represented by

$$\mathbf{B}(\sigma,t)=\tilde{\mathbf{r}}\big(\mathbf{b}(\sigma_1,\sigma_2,t),t\big)=\tilde{\mathbf{r}}\big(\mathbf{x}(\sigma,t),s(\sigma,t),t\big) \tag{7.73}$$

for $\sigma\in\Gamma$. The net rate of diffusion of momentum into the region $B(t)$ in rectangular coordinates in the inertial frame is then

$$\int_\Gamma\Big[\rho A_H\big(\nabla U\cdot\hat{\mathbf{I}}\big)\hat{\mathbf{I}}+\rho A_H\big(\nabla U\cdot\hat{\mathbf{J}}\big)\hat{\mathbf{J}}+\rho A_D\big(\nabla U\cdot\hat{\mathbf{K}}\big)\hat{\mathbf{K}}\Big]\cdot\left(\frac{\partial\mathbf{B}}{\partial\sigma_1}\times\frac{\partial\mathbf{B}}{\partial\sigma_2}\right)\mathrm{d}\sigma_1\,\mathrm{d}\sigma_2. \tag{7.74}$$

Here, it is assumed that the parameters $\sigma=(\sigma_1,\sigma_2)$ are chosen so that the cross product in (7.74) points out of the region $B(t)$. The quantities in the square brackets are evaluated at $\mathbf{B}(\sigma,t)$.

Equations (7.73), (7.67), and (7.69) imply

$$\begin{aligned}\frac{\partial\mathbf{B}}{\partial\sigma_i}&=\frac{\partial\tilde{\mathbf{r}}}{\partial x_1}\frac{\partial x_1}{\partial\sigma_i}+\frac{\partial\tilde{\mathbf{r}}}{\partial x_2}\frac{\partial x_2}{\partial\sigma_i}+\frac{\partial\tilde{\mathbf{r}}}{\partial s}\frac{\partial s}{\partial\sigma_i}\\&=\tilde{m}_1\big(\hat{\mathbf{i}}+\mathbf{O}(\delta^2)\big)\frac{\partial x_1}{\partial\sigma_i}\\&\quad+\tilde{m}_2\big(\hat{\mathbf{j}}+\mathbf{O}(\delta^2)\big)\frac{\partial x_2}{\partial\sigma_i}\\&\quad+m_s\big(\hat{\mathbf{k}}+\mathbf{O}(\delta)\big)\frac{\partial s}{\partial\sigma_i}\\&=D(\sigma,t)\mathbf{q}_i(\sigma,t)+\mathbf{e}_i,\end{aligned} \tag{7.75}$$

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

where

$$D(\sigma,t) = \big[\hat{\mathbf{i}},\ \hat{\mathbf{j}},\ \hat{\mathbf{k}}\big]$$
$$\mathbf{q}_i(\sigma,t) = \left(\tilde{m}_1\frac{\partial x_1}{\partial\sigma_i},\ \tilde{m}_2\frac{\partial x_2}{\partial\sigma_i},\ m_s\frac{\partial s}{\partial\sigma_i}\right)^T, \tag{7.76}$$

where $\mathbf{e}_i$ is an error vector involving $\delta^2(\tilde{m}_1\partial x_1/\partial\sigma_i)$, $\delta^2(\tilde{m}_2\partial x_2/\partial\sigma_i)$, and $\delta(m_s\partial s/\partial\sigma_i)$. The notation for $D(\sigma,t)$ specifies a 3×3 matrix having the indicated columns, and $\mathbf{q}_i$ is a column vector. In (7.75) and (7.76), the quantities $\tilde{m}_1$, $\tilde{m}_2$, m_s, $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$ are evaluated at $\big(\mathbf{x}(\sigma,t), s(\sigma,t), t\big)$; and the partial derivatives of x_1, x_2, and s are evaluated at (σ,t). Since D is a unitary matrix, it follows that

$$\frac{\partial\mathbf{B}}{\partial\sigma_1}\times\frac{\partial\mathbf{B}}{\partial\sigma_2} = D(\sigma,t)\big(\mathbf{q}_1\times\mathbf{q}_2\big) + \mathbf{O}(\mathbf{e}\times\mathbf{q}),$$

where $\mathbf{q}_1\times\mathbf{q}_2$ is regarded as a column vector, and the error term $\mathbf{O}\big(\mathbf{e}\times\mathbf{q}\big)$ involves δ, δ^2, and higher powers.

The net rate of diffusion (7.74), after neglecting the error terms involving δ, can then be written in the form

$$\int_\Gamma \big[F_1\hat{\mathbf{i}} + F_2\hat{\mathbf{j}} + F_3\hat{\mathbf{k}}\big]\cdot\big[D(\sigma,t)\big(\mathbf{q}_1\times\mathbf{q}_2\big)\big]\,\mathrm{d}\sigma_1\,\mathrm{d}\sigma_2, \tag{7.77}$$

where $F_1\hat{\mathbf{i}} + F_2\hat{\mathbf{j}} + F_3\hat{\mathbf{k}}$ is evaluated at $\big(\mathbf{x}(\sigma,t), s(\sigma,t), t\big)$, and the terms F_1, F_2, and F_3 are defined by comparing with (7.74). The dot product in (7.77) can be represented as a row vector times a column vector, so (7.77) is equal to

$$\begin{aligned}
&\int_\Gamma \big(F_1,F_2,F_3\big)D^TD\big(\mathbf{q}_1\times\mathbf{q}_2\big)\,\mathrm{d}\sigma_1\,\mathrm{d}\sigma_2 \qquad (7.78)\\
&\quad= \int_\Gamma \big(F_1,F_2,F_3\big)\mathbf{q}_1\times\mathbf{q}_2\,\mathrm{d}\sigma_1\,\mathrm{d}\sigma_2\\
&\quad= \int_\Gamma \big(\tilde{m}_2m_sF_1,\ \tilde{m}_1m_sF_2,\ \tilde{m}_1\tilde{m}_2F_3\big)\frac{\partial\mathbf{b}}{\partial\sigma_1}\times\frac{\partial\mathbf{b}}{\partial\sigma_2}\,\mathrm{d}\sigma_1\,\mathrm{d}\sigma_2\\
&\quad= \int_{\partial A(t)} \big(\tilde{m}_2m_sF_1,\ \tilde{m}_1m_sF_2,\ \tilde{m}_1\tilde{m}_2F_3\big)\cdot\mathbf{n}_{\partial A(t)}\,\mathrm{d}A_{\partial A(t)}\\
&\quad= \int_{A(t)}\left[\frac{\partial}{\partial x_1}\big(\tilde{m}_2m_sF_1\big) + \frac{\partial}{\partial x_2}\big(\tilde{m}_1m_sF_2\big) + \frac{\partial}{\partial s}\big(\tilde{m}_1\tilde{m}_2F_3\big)\right]\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}s.
\end{aligned}$$

The third line in (7.78) arises from examining the roles of the factors $\tilde{m}_1$, $\tilde{m}_2$, and m_s in $\mathbf{q}_i$ on the cross product and observing that the components of $\mathbf{b}(\sigma,t)$ are $x_1(\sigma,t)$, $x_2(\sigma,t)$, and $s(\sigma,t)$. The integrands in the second and third lines have the form of a row vector times a column vector. The fourth

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

line is an integral on the boundary $\partial A(t)$ of the region $A(t)$ in parameter space, and $\mathbf{n}_{\partial A(t)}(\mathbf{x}, s, t)$ is the unit outward normal to that boundary. The last line is obtained by using the divergence theorem in parameter space, with independent variables (x_1, x_2, s).

Equations (7.74), (7.77), (7.67), and (7.70) imply

$$\begin{aligned} F_1(\mathbf{x}, s, t) &= \rho A_H \nabla U\big(\tilde{\mathbf{r}}(\mathbf{x}, s, t), t\big) \cdot \hat{\mathbf{i}}(\mathbf{x}, s, t) \\ &= \rho A_H \nabla U \cdot \left[\frac{1}{\tilde{m}_1(\mathbf{x}, s, t)} \frac{\partial \tilde{\mathbf{r}}}{\partial x_1}(\mathbf{x}, s, t) + \mathbf{O}(\delta^2)\right] \\ &= \rho A_H \left[\frac{1}{\tilde{m}_1(\mathbf{x}, s, t)} \frac{\partial u}{\partial x_1}(\mathbf{x}, s, t) + O(\delta^2)\right]. \end{aligned}$$

Similarly,

$$F_2(\mathbf{x}, s, t) = \rho A_H \left[\frac{1}{\tilde{m}_2(\mathbf{x}, s, t)} \frac{\partial u}{\partial x_2}(\mathbf{x}, s, t) + O(\delta^2)\right],$$

$$F_3(\mathbf{x}, s, t) = \rho A_D \left[\frac{1}{m_s(\mathbf{x}, s, t)} \frac{\partial u}{\partial s}(\mathbf{x}, s, t) + O(\delta)\right].$$

According to equation (7.66), $\tilde{m}_i$ can be replaced by m_i for $i = 1$ and $i = 2$, to order δ^2. If the error terms involving δ and higher powers are neglected, then (7.78) can be written as

$$\begin{aligned} \int_{A(t)} \Bigg[&\frac{\partial}{\partial x_1}\left(\rho A_H \left(\frac{1}{m_1}\frac{\partial u}{\partial x_1}\right) m_2 m_s\right) + \frac{\partial}{\partial x_2}\left(\rho A_H \left(\frac{1}{m_2}\frac{\partial u}{\partial x_2}\right) m_1 m_s\right) \\ &+ \frac{\partial}{\partial s}\left(\rho A_D \left(\frac{1}{m_s}\frac{\partial u}{\partial s}\right) m_1 m_2\right)\Bigg] \mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}s. \end{aligned} \tag{7.79}$$

The integral (7.79) represents the net rate of diffusion of the u-component of momentum into the volume of fluid that occupies the region $A(t)$ in parameter space and the region $B(t)$ in rectangular coordinates in the inertial reference frame. The rate of diffusion of the v-component is obtained by replacing u with v in (7.79).

7.13. Applied stresses plus diffusion

The ocean is subjected to forces due to wind stress at the top of the fluid and frictional stress along the bottom. In the present subsection we develop a representation of these forces and then combine that result with the representation (7.79) of diffusion.

Here, it is assumed that each of these stresses acts as a body force near the top or bottom boundary, with the amplitude of the stress decaying with distance from the boundary. In the case of an isopycnic-coordinate ocean model, layer thicknesses can tend to zero due to interfaces intersecting the

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

top or bottom of the fluid domain. If the wind or bottom stress were applied strictly as a surface force, then nonzero forcing could be applied to layers having essentially zero thickness. This would have adverse algorithmic consequences, whereas the assumption of a decaying body force enables the forcing to be distributed proportionately when thin layers are found near the boundary.

Assume that the wind and bottom stresses act horizontally, and denote the sum of the stresses by $\boldsymbol{\tau}^{wb} = \tau_u^{wb}\mathbf{i} + \tau_v^{wb}\mathbf{j}$. Near the upper boundary $\boldsymbol{\tau}^{wb}$ is due to the wind, near the bottom boundary $\boldsymbol{\tau}^{wb}$ is due to bottom friction, and in the ocean's interior $\boldsymbol{\tau}^{wb}$ is zero. Along any horizontal plane, $\boldsymbol{\tau}^{wb}$ represents a horizontal force, per unit horizontal area, that is exerted by the upper region on the lower region. The net force, due to $\boldsymbol{\tau}^{wb}$, acting on the volume of fluid contained in the region $B(t)$ in rectangular coordinates is then given by

$$\int_{\partial B(t)} \boldsymbol{\tau}^{wb}(\mathbf{k}\cdot\mathbf{n})\,\mathrm{d}A, \tag{7.80}$$

where $\mathbf{n}(\mathbf{r},t)$ denotes the unit outward normal to the boundary $\partial B(t)$.

The statement (7.80) can be justified as follows. The quantity $\mathbf{k}\cdot\mathbf{n}$ is the vertical component of the unit outward normal. If an element ΔA of $\partial B(t)$ is horizontal and lies on the upper part of $\partial B(t)$, then $\mathbf{k}\cdot\mathbf{n} = 1$, and the corresponding contribution to (7.80) is $\boldsymbol{\tau}^{wb}\Delta A$. This is the product of the area and the force per unit area, and it is consistent with the sign convention that a stress represents the force exerted by an upper region on a lower region. On the other hand, if ΔA is horizontal and lies on the lower part of $\partial B(t)$, then $\mathbf{k}\cdot\mathbf{n} = -1$, and the corresponding contribution to (7.80) is $-\boldsymbol{\tau}^{wb}\Delta A$. The minus sign indicates the effect of a lower region on an upper region.

For a general element of area ΔA on $\partial B(t)$, the quantity $(\mathbf{k}\cdot\mathbf{n})\Delta A$ is a signed projection of that area onto a horizontal plane. Consider the element of fluid bounded by ΔA, the horizontal projection, and a vertical surface. The quantity $\mathbf{F}_H = -\boldsymbol{\tau}^{wb}(\mathbf{k}\cdot\mathbf{n})\Delta A$ represents the force exerted on the element of fluid along that horizontal projection. If $\mathbf{F}_{\Delta A}$ represents the force exerted along the area element ΔA, then the net force acting on the fluid element, due to the horizontal stress $\boldsymbol{\tau}^{wb}$, is $\mathbf{F}_{\Delta A} + \mathbf{F}_H$. This net force is proportional to the volume of the fluid element and thus varies as the cube of distance. However, $\mathbf{F}_{\Delta A}$ and $\mathbf{F}_H$ vary as the square of distance, so for sufficiently small elements $\mathbf{F}_{\Delta A} \approx -\mathbf{F}_H = \boldsymbol{\tau}^{wb}(\mathbf{k}\cdot\mathbf{n})\Delta A$.

The force (7.80) can be represented as an integral on the region $A(t)$ in parameter space by using techniques similar to those used in the preceding subsection. In the present case, parametrize $\partial B(t)$ by

$$\mathbf{B}^z(\sigma,t) = \mathbf{r}\big(\mathbf{x}(\sigma,t), z(\sigma,t), t\big), \tag{7.81}$$

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

so that

$$\begin{aligned}\frac{\partial \mathbf{B}^z}{\partial \sigma_i} &= \frac{\partial \mathbf{r}}{\partial x_1}\frac{\partial x_1}{\partial \sigma_i} + \frac{\partial \mathbf{r}}{\partial x_2}\frac{\partial x_2}{\partial \sigma_i} + \frac{\partial \mathbf{r}}{\partial z}\frac{\partial z}{\partial \sigma_i} \\ &= m_1\frac{\partial x_1}{\partial \sigma_i}\mathbf{i} + m_2\frac{\partial x_2}{\partial \sigma_i}\mathbf{j} + m_s\frac{\partial s}{\partial \sigma_i}\mathbf{k}.\end{aligned} \tag{7.82}$$

Calculations similar to those in (7.74)–(7.78) imply that the force (7.80) is equal to

$$\int_{A(t)}\left[\frac{\partial}{\partial s}\big(m_1 m_2 \tau_u^{wb}\big)\mathbf{i} + \frac{\partial}{\partial s}\big(m_1 m_2 \tau_v^{wb}\big)\mathbf{j}\right] \mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}s. \tag{7.83}$$

During these calculations, $\tilde{\mathbf{i}}$, $\tilde{\mathbf{j}}$, $\tilde{\mathbf{k}}$, $\tilde{m}_1$, and $\tilde{m}_2$ are replaced by $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, m_1, and m_2, respectively, and there are no error terms involving δ.

Now combine the representation (7.83) of the applied stresses with the representation (7.79) for the diffusion of the u-component of momentum and the analogous formula for the v-component. According to (7.10), the hydrostatic condition implies $\rho m_s = -p_s/g$, and this relation can be inserted into the first two terms in (7.79). For the u-component, the combined effect of diffusion and applied stress is

$$\begin{aligned}-\frac{1}{g}\int_{A(t)}\Bigg[&\frac{\partial}{\partial x_1}\left(A_H\left(\frac{1}{m_1}\frac{\partial u}{\partial x_1}\right)m_2 p_s\right) + \frac{\partial}{\partial x_2}\left(A_H\left(\frac{1}{m_2}\frac{\partial u}{\partial x_2}\right)m_1 p_s\right) \\ &- \frac{\partial}{\partial s}\big(g\tau_u G\big)\Bigg]\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}s,\end{aligned} \tag{7.84}$$

where $G = m_1 m_2$ and

$$\tau_u = \tau_u^{wb} + \rho A_D\left(\frac{1}{m_s}\frac{\partial u}{\partial s}\right). \tag{7.85}$$

The corresponding formulas for the v-component are obtained by replacing u with v in (7.84) and (7.85).

The second term in (7.85) can also be regarded as $\rho A_D \partial u/\partial z$, and it represents the internal friction due to vertical variations in horizontal velocity. This term represents a form of shear stress, and in (7.85) it is combined with the shear stress τ_u^{wb} due to wind forcing and bottom friction to yield the total shear stress τ_u in the u-direction. Analogous remarks apply to the total shear stress τ_v in the v-direction.

7.14. Conservation of momentum

Equation (7.53) gives the derivative $\mathbf{Q}'(t)$ of the momentum of the volume of fluid that occupies the region $A(t)$ in parameter space and the region $B(t)$ in rectangular coordinates in the inertial reference frame, at time t. According to the principle of conservation of momentum, $\mathbf{Q}'(t)$ is equal to

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

the sum of the forces acting on that volume of fluid due to pressure, gravity, applied stresses, and diffusion of momentum. In the present subsection the results of the preceding subsections are combined to express this principle in terms of partial differential equations.

The formula for $\mathbf{Q}'(t)$ in (7.53) includes a term involving the centripetal acceleration $\mathbf{a}_c$. According to the hydrostatic assumption stated in Section 7.1, $\mathbf{a}_c$ is equal to the gravitational acceleration plus the vertical component of the pressure force per unit mass. It then follows that the term involving centripetal acceleration in the formula for $\mathbf{Q}'(t)$ is balanced by terms involving gravity and the vertical component of the pressure force. These terms will then be deleted from the following discussion. The vertical velocity w will also be neglected, due to the hydrostatic assumption, and from now on the discussion focuses on the horizontal component of $\mathbf{Q}'(t)$.

This component can be expressed as

$$\mathbf{Q}'_H(t) = -\frac{1}{g} \int_{A(t)} \Big[\big(D_f(Gup_s) + \mathbf{\Psi} \cdot \mathbf{i} \big) \mathbf{i} + \big(D_f(Gvp_s) + \mathbf{\Psi} \cdot \mathbf{j} \big) \mathbf{j} \Big] \, d\mathbf{x} \, ds, \quad (7.86)$$

where $\mathbf{\Psi}$ is defined in (7.54). When $\mathbf{\Psi} \cdot \mathbf{i}$ and $\mathbf{\Psi} \cdot \mathbf{j}$ are calculated, the terms in $\mathbf{\Psi}$ involving $\mathbf{k}$ can be neglected, as these involve the vertical velocity w. In addition, $\mathbf{i}(\mathbf{x}, t) \cdot \mathbf{i}(\mathbf{x}, t) = 1$, $\mathbf{j}(\mathbf{x}, t) \cdot \mathbf{j}(\mathbf{x}, t) = 1$, and $\mathbf{i}(\mathbf{x}, t) \cdot \mathbf{j}(\mathbf{x}, t) = 0$ for all $(\mathbf{x}, t)$, so

$$\begin{aligned} \mathbf{i} \cdot \frac{\partial \mathbf{i}}{\partial t} = \mathbf{i} \cdot \frac{\partial \mathbf{i}}{\partial x_i} = \mathbf{j} \cdot \frac{\partial \mathbf{j}}{\partial t} = \mathbf{j} \cdot \frac{\partial \mathbf{j}}{\partial x_i} = 0, \\ \mathbf{i} \cdot \frac{\partial \mathbf{j}}{\partial t} = -\mathbf{j} \cdot \frac{\partial \mathbf{i}}{\partial t}, \\ \mathbf{i} \cdot \frac{\partial \mathbf{j}}{\partial x_i} = -\mathbf{j} \cdot \frac{\partial \mathbf{i}}{\partial x_i}. \end{aligned} \quad (7.87)$$

It then follows from (7.54) that

$$\begin{aligned} \mathbf{\Psi} \cdot \mathbf{i} &= \big(Gvp_s \big) \left[2 \, \mathbf{i} \cdot \frac{\partial \mathbf{j}}{\partial t} + \mathbf{i} \cdot \left(\dot{x}_1 \frac{\partial \mathbf{j}}{\partial x_1} + \dot{x}_2 \frac{\partial \mathbf{j}}{\partial x_2} \right) \right], \\ \mathbf{\Psi} \cdot \mathbf{j} &= \big(Gup_s \big) \left[2 \, \mathbf{j} \cdot \frac{\partial \mathbf{i}}{\partial t} + \mathbf{j} \cdot \left(\dot{x}_1 \frac{\partial \mathbf{i}}{\partial x_1} + \dot{x}_2 \frac{\partial \mathbf{i}}{\partial x_2} \right) \right]. \end{aligned} \quad (7.88)$$

The last two lines in (7.87) imply that the bracketed quantities in (7.88) are negatives of each other.

Define the Coriolis parameter f by

$$f(\mathbf{x}) = 2 \, \mathbf{j} \cdot \frac{\partial \mathbf{i}}{\partial t} \quad (7.89)$$

for all $\mathbf{x}$. The partial derivative $\partial \mathbf{i} / \partial t$ is taken for a fixed position on the rotating spheroid Σ, so the presence of f is due to the rotation of Σ. The right side of (7.89) appears to depend on t; however, the alternate

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

representation developed in Section 7.15 shows that f is independent of t and is also independent of the choice of coordinate system on Σ.

The remaining terms in the square brackets in (7.88) involve the variation of $\mathbf{i}$ and $\mathbf{j}$ with respect to position on Σ, for fixed t. These terms are therefore due to the curvature of Σ and/or the properties of the parametrization of Σ in terms of $\mathbf{x}$. The quantities $\dot{x}_1$ and $\dot{x}_2$ can depend on $(\mathbf{x}, s, t)$, but the dot products of the quantities involving $\mathbf{i}$ and $\mathbf{j}$ depend only on $\mathbf{x}$.

Let

$$\tilde{f}(\mathbf{x}, s, t) = f(\mathbf{x}) + \mathbf{j} \cdot \left(\dot{x}_1 \frac{\partial \mathbf{i}}{\partial x_1} + \dot{x}_2 \frac{\partial \mathbf{i}}{\partial x_2} \right) \tag{7.90}$$

denote the sum of the Coriolis and geometric parameters. The derivative (7.86) of the horizontal component of momentum then becomes

$$\begin{aligned} \mathbf{Q}'_H(t) = -\frac{1}{g} \int_{A(t)} \Big[& \big(D_f(Gup_s) - \tilde{f} Gvp_s \big) \mathbf{i} \\ & + \big(D_f(Gvp_s) + \tilde{f} Gup_s \big) \mathbf{j} \Big] \, \mathrm{d}\mathbf{x} \, \mathrm{d}s. \end{aligned} \tag{7.91}$$

Now compare to the pressure force (7.64) and the combined effect (7.84) of the diffusion and applied stresses to obtain the momentum equations

$$\begin{aligned} D_f\big(Gup_s\big) - \tilde{f} Gvp_s &= -\frac{1}{m_1} \left(\frac{\partial M}{\partial x_1} - p \frac{\partial \alpha}{\partial x_1} \right) Gp_s - \frac{\partial}{\partial s} \big(g \tau_u G \big) + \psi_u, \\ D_f\big(Gvp_s\big) + \tilde{f} Gup_s &= -\frac{1}{m_2} \left(\frac{\partial M}{\partial x_2} - p \frac{\partial \alpha}{\partial x_2} \right) Gp_s - \frac{\partial}{\partial s} \big(g \tau_v G \big) + \psi_v. \end{aligned} \tag{7.92}$$

The flux derivative D_f is defined in (7.46), $G = m_1 m_2$, the Montgomery potential $M = \alpha p + gz$ is defined in (7.62), the total shear stress τ_u is defined in (7.85), and

$$\psi_u = \frac{\partial}{\partial x_1} \left(A_H \left(\frac{1}{m_1} \frac{\partial u}{\partial x_1} \right) m_2 p_s \right) + \frac{\partial}{\partial x_2} \left(A_H \left(\frac{1}{m_2} \frac{\partial u}{\partial x_2} \right) m_1 p_s \right) \tag{7.93}$$

represents the diffusion of u along the s-coordinate surfaces. The quantities τ_v and ψ_v are analogous. The terms ψ_u and ψ_v, and the terms involving A_D in τ_u and τ_v, depend on the particular parametrization of subgrid-scale mixing developed in Section 7.12.

The pressure forcing on the right sides in (7.92) can be expressed in the form $-(\nabla M - p \nabla \alpha) G p_s$. Depending on the exact choice of the vertical coordinate s, it is possible for the terms ∇M and $p \nabla \alpha$ to be far larger than their difference, under certain conditions. This situation can lead to numerical inaccuracy and even numerical instability. However, these problems can be remedied by proper choices of the details of the vertical coordinate and a proper implementation of the pressure term: see, *e.g.*, Sun *et al.* (1999), Bleck (2002), and Hallberg (2005).

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

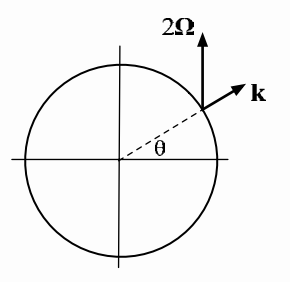


Figure 7.3. The Coriolis parameter $f = 2\mathbf{\Omega} \cdot \mathbf{k} = 2\Omega \sin\theta$ is the local vertical component of the planetary vorticity $2\mathbf{\Omega}$. This parameter is zero at the equator and has maximum magnitude at the poles.

7.15. The Coriolis parameter

Here we develop a more transparent representation of the Coriolis parameter $f(\mathbf{x}) = 2\mathbf{j} \cdot \frac{\partial \mathbf{i}}{\partial t}$ given in (7.89). Equations (7.1) and (7.3) imply

$$\mathbf{i}(\mathbf{x}, t) = Q(\Omega t)\mathbf{i}(\mathbf{x}, 0) \quad \text{and} \quad \mathbf{j}(\mathbf{x}, t) = Q(\Omega t)\mathbf{j}(\mathbf{x}, 0),$$

where $Q(\Omega t)$ is a rotation matrix. Then

$$\frac{\partial \mathbf{i}}{\partial t}(\mathbf{x}, t) = \Omega Q'(\Omega t)\mathbf{i}(\mathbf{x}, 0) = \Omega Q'(\Omega t)Q^T(\Omega t)\mathbf{i}(\mathbf{x}, t),$$

where the superscript T denotes the transpose. A calculation shows that $\frac{\partial \mathbf{i}}{\partial t}(\mathbf{x}, t) = \mathbf{\Omega} \times \mathbf{i}(\mathbf{x}, t)$, which is the velocity associated with a rigid-body rotation. Here, $\mathbf{\Omega}$ is a vector with length Ω that is aligned with the axis of rotation, and the direction of $\mathbf{\Omega}$ and the direction of rotation are related by the right-hand rule. It then follows that $f = 2\mathbf{j} \cdot (\mathbf{\Omega} \times \mathbf{i}) = 2\mathbf{\Omega} \cdot (\mathbf{i} \times \mathbf{j})$. Now assume that the local coordinate system on Σ is right-handed, in the sense that $\mathbf{i} \times \mathbf{j} = \mathbf{k}$. The Coriolis parameter can then be written as

$$f(\mathbf{x}) = 2\mathbf{\Omega} \cdot \mathbf{k} = 2\Omega \sin\theta, \tag{7.94}$$

where θ is the latitude. The second equality neglects the slight departure of Σ from being a perfect sphere.

A calculation shows that if the fluid is motionless relative to the rotating spheroid Σ, then the vorticity (curl of velocity) of that fluid relative to the inertial reference frame is $2\mathbf{\Omega}$. The term 'planetary vorticity' is commonly applied to this vorticity. At any point on the spheroid, the value of the Coriolis parameter is then the local vertical component of the planetary vorticity, as illustrated in Figure 7.3.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

The Coriolis parameter is zero at the equator, positive in the northern hemisphere, and negative in the southern hemisphere. The Earth rotates at a rate of approximately $2\pi + 2\pi/360.24$ radians in 24 hours, and for the Earth the maximum value of $|f|$ is $2\Omega \approx 1.46 \times 10^{-4}$ sec^{-1}.

7.16. Conservation of momentum in layers

The partial differential equations (7.92) describe the conservation of momentum in the x_1- and x_2-directions, respectively. In anticipation of solving these equations numerically, we now develop the vertically discrete equations obtained by integrating the equations (7.92) with respect to s between the coordinate surfaces $s = s_0$ and $s = s_1$, with $s_0 < s_1$. The present discussion is an analogue of the discussion of conservation of mass in layers given in Section 7.8.

As in Section 7.8, let

$$\Delta p(\mathbf{x}, t) = p(\mathbf{x}, s_0, t) - p(\mathbf{x}, s_1, t) = \int_{s_0}^{s_1} (-p_s)\, \mathrm{d}s > 0, \tag{7.95}$$

and define mass-weighted vertical averages of the velocity components u and v by

$$\begin{aligned} \bar{u}(\mathbf{x}, t) &= \frac{1}{\Delta p} \int_{s_0}^{s_1} u(\mathbf{x}, s, t)(-p_s)\, \mathrm{d}s, \\ \bar{v}(\mathbf{x}, t) &= \frac{1}{\Delta p} \int_{s_0}^{s_1} v(\mathbf{x}, s, t)(-p_s)\, \mathrm{d}s. \end{aligned} \tag{7.96}$$

Denote the deviations from these averages by δ_u and δ_v, respectively, so that $u(\mathbf{x}, s, t) = \bar{u}(\mathbf{x}, t) + \delta_u(\mathbf{x}, s, t)$ and $v(\mathbf{x}, s, t) = \bar{v}(\mathbf{x}, t) + \delta_v(\mathbf{x}, s, t)$. It then follows that

$$\int_{s_0}^{s_1} \delta_u(\mathbf{x}, s, t)(-p_s)\, \mathrm{d}s = \int_{s_0}^{s_1} \delta_v(\mathbf{x}, s, t)(-p_s)\, \mathrm{d}s = 0 \tag{7.97}$$

and $\delta_u = O(\Delta s)$ and $\delta_v = O(\Delta s)$, where $\Delta s = s_1 - s_0$.

Now consider the vertical integral of the u-equation in (7.92). The flux derivative in the u-equation is

$$\begin{aligned} D_f\big(Gup_s\big) &= \frac{\partial}{\partial t}\big(Gup_s\big) + \frac{\partial}{\partial x_1}\big(\dot{x}_1 Gup_s\big) + \frac{\partial}{\partial x_2}\big(\dot{x}_2 Gup_s\big) + \frac{\partial}{\partial s}\big(\dot{s} Gup_s\big) \\ &= \frac{\partial}{\partial t}\big(Gup_s\big) + \frac{\partial}{\partial x_1}\big(uup_s m_2\big) + \frac{\partial}{\partial x_2}\big(vup_s m_1\big) + \frac{\partial}{\partial s}\big(\dot{s} Gup_s\big). \end{aligned} \tag{7.98}$$

The second equality uses the relations $u = m_1\dot{x}_1$, $v = m_2\dot{x}_2$, and $G = m_1 m_2$. From now on, use the approximation $G(\mathbf{x}) = m_1(\mathbf{x})m_2(\mathbf{x})$, as discussed in

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Section 7.7. Multiply (7.98) by -1 and integrate over s to obtain

$$\begin{aligned} -\int_{s_0}^{s_1} D_f\big(Gup_s\big)\,\mathrm{d}s = \frac{\partial}{\partial t}\big(G\bar{u}\Delta p\big) + \frac{\partial}{\partial x_1}\left(m_2\int_{s_0}^{s_1} uu(-p_s)\,\mathrm{d}s\right) \\ + \frac{\partial}{\partial x_2}\left(m_1\int_{s_0}^{s_1} vu(-p_s)\,\mathrm{d}s\right) \\ + G\Big[\big(\dot{s}p_s u\big)_{s=s_0} - \big(\dot{s}p_s u\big)_{s=s_1}\Big]. \end{aligned} \tag{7.99}$$

The relations (7.95) and (7.97) imply

$$\begin{aligned} \int_{s_0}^{s_1} uu(-p_s)\,\mathrm{d}s &= \int_{s_0}^{s_1} (\bar{u}+\delta_u)(\bar{u}+\delta_u)(-p_s)\,\mathrm{d}s \\ &= \bar{u}\bar{u}\Delta p + O(\Delta s)^3, \end{aligned}$$

so equation (7.99) becomes

$$\begin{aligned} -\int_{s_0}^{s_1} D_f(Gup_s)\,\mathrm{d}s = \frac{\partial}{\partial t}\big(G(\bar{u}\Delta p)\big) + \frac{\partial}{\partial x_1}\big(m_2\bar{u}(\bar{u}\Delta p)\big) \\ + \frac{\partial}{\partial x_2}\big(m_1\bar{v}(\bar{u}\Delta p)\big) + O(\Delta s)^3 \\ + G\Big[\big(\dot{s}p_s u\big)_{s=s_0} - \big(\dot{s}p_s u\big)_{s=s_1}\Big]. \end{aligned} \tag{7.100}$$

The remaining terms in the u-equation in (7.92) can be handled in a similar manner. The resulting vertically integrated u-equation, after the deletion of error terms, is

$$\begin{aligned} \frac{\partial}{\partial t}\big(G(\bar{u}\Delta p)\big) + \frac{\partial}{\partial x_1}\big(m_2\bar{u}(\bar{u}\Delta p)\big) + \frac{\partial}{\partial x_2}\big(m_1\bar{v}(\bar{u}\Delta p)\big) \\ + G\Big[\big(\dot{s}p_s u\big)_{s=s_0} - \big(\dot{s}p_s u\big)_{s=s_1}\Big] \\ - \bar{f}G\big(\bar{v}\Delta p\big) \\ = -G\Delta p\,\frac{1}{m_1}\left(\frac{\partial\overline{M}}{\partial x_1} - \bar{p}\frac{\partial\bar{\alpha}}{\partial x_1}\right) + Gg\Delta\tau_u + \bar{\psi}_u \end{aligned} \tag{7.101}$$

where $\Delta\tau_u = \big(\tau_u\big)_{s=s_1} - \big(\tau_u\big)_{s=s_0}$ denotes the vertical difference of the total shear stress defined in (7.85); $\bar{f}(\mathbf{x},t)$ is the mass-weighted vertical average of the Coriolis/geometric parameter (7.90), *i.e.*,

$$\bar{f}(\mathbf{x},t) = f(\mathbf{x}) + \mathbf{j}\cdot\left(\bar{u}\frac{1}{m_1}\frac{\partial\mathbf{i}}{\partial x_1} + \bar{v}\frac{1}{m_2}\frac{\partial\mathbf{i}}{\partial x_2}\right); \tag{7.102}$$

$\overline{M}$, $\bar{\alpha}$, and $\bar{p}$, respectively, are the mass-weighted vertical averages of M, α, and p; and $\bar{\psi}_u$ is the vertical integral of the lateral diffusion term (7.93), *i.e.*,

$$\bar{\psi}_u = \frac{\partial}{\partial x_1}\left(A_H\left(\frac{1}{m_1}\frac{\partial\bar{u}}{\partial x_1}\right)m_2\Delta p\right) + \frac{\partial}{\partial x_2}\left(A_H\left(\frac{1}{m_2}\frac{\partial\bar{u}}{\partial x_2}\right)m_1\Delta p\right). \tag{7.103}$$

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

The u-momentum equation (7.101) can be interpreted as follows. As noted in Section 7.8, $\Delta p(\mathbf{x}, t)$ is the weight per unit horizontal area in the fluid layer lying between the coordinate surfaces $s = s_0$ and $s = s_1$. The quantities $\bar{u}\Delta p$ and $\bar{v}\Delta p$ then represent components of momentum density (momentum per unit horizontal area), times g. In analogy to the discussion of the mass equation given in Section 7.8, integrate the u-momentum equation (7.101) on a rectangle R in the $\mathbf{x}$ domain having sides Δx_1 and Δx_2. The quantities $m_1\Delta x_1$ and $m_2\Delta x_2$ give the (approximate) linear dimensions of the corresponding rectangular region $\tilde{R}$ on the surface of the rotating spheroid Σ. The quantity $G\Delta x_1\Delta x_2\big(\bar{u}\Delta p\big) = m_1\Delta x_1 m_2\Delta x_2\big(\bar{u}\Delta p\big)$ is then the u-component of momentum in the given layer in $\tilde{R}$, times g. The integrals of the second and third terms in (7.101) (involving $\partial/\partial x_1$ and $\partial/\partial x_2$) can be interpreted in terms of the lateral advection of momentum across the edges of $\tilde{R}$.

As described in Section 7.8, the quantity $-\dot{s}p_s$ is the rate of flow of mass per unit area (times g) across a coordinate surface due to material changes in s. The terms involving $\dot{s}p_s u$ then represent the rate of transport of u-momentum across coordinate surfaces due to the movement of such surfaces relative to the fluid.

The vertical integral of the v-equation in (7.92) is derived in a manner similar to that of (7.101), and the result is

$$\begin{aligned}\frac{\partial}{\partial t}\big(G(\bar{v}\Delta p)\big) + \frac{\partial}{\partial x_1}\big(m_2\bar{u}(\bar{v}\Delta p)\big) + \frac{\partial}{\partial x_2}\big(m_1\bar{v}(\bar{v}\Delta p)\big) \\ + G\Big[\big(\dot{s}p_s v\big)_{s=s_0} - \big(\dot{s}p_s v\big)_{s=s_1}\Big] \\ + \bar{f}G\big(\bar{u}\Delta p\big) \\ = -G\Delta p\,\frac{1}{m_2}\left(\frac{\partial\overline{M}}{\partial x_2} - \bar{p}\frac{\partial\bar{\alpha}}{\partial x_2}\right) + Gg\Delta\tau_v + \bar{\psi}_v,\end{aligned} \tag{7.104}$$

where $\bar{\psi}_v$ is obtained by replacing $\bar{u}$ with $\bar{v}$ in (7.103).

Equations (7.101) and (7.104) describe the conservation of horizontal momentum in a coordinate layer. In these equations the dependent variables are the components of momentum density, and the second and third terms in each equation are written in terms of the horizontal flux of that density. This formulation facilitates the usage of (nearly) nonoscillatory advection schemes for solving these equations numerically. However, it has been commonplace in the ocean modelling community to use the components of velocity as the dependent variables, and the corresponding form of the momentum equations can be obtained by combining equations (7.101) and (7.104) with the layer mass equation (7.38). A derivation for the case of Cartesian coordinates on a tangent plane is given in Section 6.4.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Also needed for solving the momentum equations is a vertically discrete analogue of equation (7.65), $\partial M/\partial s = p\partial\alpha/\partial s$, which is essentially a statement of the hydrostatic condition in a generalized vertical coordinate. This relation can also be expressed in the form $\partial M/\partial\alpha = p$, and this latter form provides a guide for the following discretization.

Define $\widetilde{M}(\mathbf{x},\alpha,t)$ by $M(\mathbf{x},s,t) = \widetilde{M}\big(\mathbf{x},\alpha(\mathbf{x},s,t),t\big)$ for all $\mathbf{x}$ and t. Then $p\partial\alpha/\partial s = \partial M/\partial s = (\partial\widetilde{M}/\partial\alpha)(\partial\alpha/\partial s)$, so

$$\frac{\partial\widetilde{M}}{\partial\alpha}\big(\mathbf{x},\alpha(\mathbf{x},s,t),t\big) = p(\mathbf{x},s,t).$$

Consider two adjacent coordinate layers bounded by the surfaces $s = s_0$, $s = s_1$, and $s = s_2$, with $s_0 < s_1 < s_2$. The interface between the two layers is then defined by $s = s_1$. Denote the values of M, α, and p on that interface by $M_{\rm int}(\mathbf{x},t)$, $\alpha_{\rm int}(\mathbf{x},t)$, and $p_{\rm int}(\mathbf{x},t)$, respectively. Then

$$\begin{aligned} M(\mathbf{x},s,t) &= \widetilde{M}\big(\mathbf{x},\alpha(\mathbf{x},s,t),t\big) \\ &= M_{\rm int}(\mathbf{x},t) + p_{\rm int}(\mathbf{x},t)\big(\alpha(\mathbf{x},s,t) - \alpha_{\rm int}(\mathbf{x},t)\big) + O\big(\alpha-\alpha_{\rm int}\big)^2 \end{aligned}$$

for all $(\mathbf{x},s)$ in each of the two layers. Denote the mass-weighted vertical averages of M and α on the upper layer $s_1 < s < s_2$ by $\overline{M}_{\rm upper}(\mathbf{x},t)$ and $\bar\alpha_{\rm upper}(\mathbf{x},t)$, respectively; and denote their averages on the lower layer $s_0 < s < s_1$ by $\overline{M}_{\rm lower}(\mathbf{x},t)$ and $\bar\alpha_{\rm lower}(\mathbf{x},t)$, respectively. Then

$$\begin{aligned} \overline{M}_{\rm upper}(\mathbf{x},t) &= M_{\rm int}(\mathbf{x},t) + p_{\rm int}(\mathbf{x},t)\big(\bar\alpha_{\rm upper}(\mathbf{x},t) - \alpha_{\rm int}(\mathbf{x},t)\big) + O(\Delta s)^3, \\ \overline{M}_{\rm lower}(\mathbf{x},t) &= M_{\rm int}(\mathbf{x},t) + p_{\rm int}(\mathbf{x},t)\big(\bar\alpha_{\rm lower}(\mathbf{x},t) - \alpha_{\rm int}(\mathbf{x},t)\big) + O(\Delta s)^3, \end{aligned}$$

and thus $\overline{M}_{\rm upper} - \overline{M}_{\rm lower} = p_{\rm int}\big(\bar\alpha_{\rm upper} - \bar\alpha_{\rm lower}\big) + O(\Delta s)^3$. A second-order approximation to the hydrostatic condition (7.65) is then

$$\frac{\Delta\overline{M}}{\Delta\bar\alpha} = \frac{\overline{M}_{\rm upper} - \overline{M}_{\rm lower}}{\bar\alpha_{\rm upper} - \bar\alpha_{\rm lower}} = p_{\rm int}. \tag{7.105}$$

The relation (7.105) provides a means for communicating pressure effects between layers.

7.17. Transport of tracers

Next consider the transport by the fluid of a tracer such as heat, salt, or a chemical or biological component. Let $q(\mathbf{x},s,t)$ denote the quantity of tracer per unit mass of the fluid. Then the quantity of tracer per unit volume is $\rho(\mathbf{x},s,t)q(\mathbf{x},s,t)$, where ρ is the density (mass per unit volume) of the fluid itself.

In the discussion of conservation of mass of the fluid given in Section 7.7, $A(t)$ denotes the region of parameter space occupied by a volume of fluid,

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

and $B(t)$ denotes the corresponding region in rectangular coordinates in the inertial reference frame. The total mass of the fluid occupying those regions remains constant in time, since the regions follow the fluid. In the case of a tracer, it is assumed here that the amount of tracer in the volume of fluid can vary due to diffusion across the boundary. It is also assumed that the diffusion is due to the gradient of q and occurs predominantly along surfaces of constant s, as in the discussion of diffusion of momentum in Section 7.12.

In analogy to the discussions in Sections 7.5–7.7, the total quantity of tracer in the region $A(t)$ in parameter space at time t is

$$\int_{A(t)} \rho(\mathbf{x}, s, t) q(\mathbf{x}, s, t)\ J(\mathbf{x}, s, t)\, \mathrm{d}\mathbf{x}\, \mathrm{d}s,$$

where $J = m_1 m_2 m_s$ is the Jacobian defined in (7.17). The time derivative of this quantity is

$$\int_{A(t)} D_f(J\rho q)\, \mathrm{d}\mathbf{x}\, \mathrm{d}s,$$

where D_f is the flux derivative given in (7.27) and (7.46). The diffusion of the tracer can be modelled by the same form (7.71) as used for the diffusion of momentum, with q replacing U in that formula. Calculations analogous to those in Sections 7.12 and 7.14 imply

$$D_f\big(Gqp_s\big) = \psi_q + \frac{\partial}{\partial s}(GgF_D). \tag{7.106}$$

Here,

$$F_D = -\rho A_D \left(\frac{1}{m_s}\frac{\partial q}{\partial s}\right)$$

denotes the rate of diapycnal diffusion of q per unit cross-sectional area, and

$$\psi_q = \frac{\partial}{\partial x_1}\left(A_H\left(\frac{1}{m_1}\frac{\partial q}{\partial x_1}\right)m_2 p_s\right) + \frac{\partial}{\partial x_2}\left(A_H\left(\frac{1}{m_2}\frac{\partial q}{\partial x_2}\right)m_1 p_s\right).$$

Integration over a layer between two coordinate surfaces $s = s_0$ and $s = s_1$, analogous to the calculations in Section 7.16, yields

$$\begin{aligned} \frac{\partial}{\partial t}\big(G(\bar{q}\Delta p)\big) + \frac{\partial}{\partial x_1}\big(m_2\bar{u}(\bar{q}\Delta p)\big) + \frac{\partial}{\partial x_2}\big(m_1\bar{v}(\bar{q}\Delta p)\big) \\ + G\Big[\big(\dot{s}p_s q\big)_{s=s_0} - \big(\dot{s}p_s q\big)_{s=s_1}\Big] \\ = \bar{\psi}_q + Gg\Big[\big(F_D\big)_{s=s_0} - \big(F_D\big)_{s=s_1}\Big], \end{aligned} \tag{7.107}$$

where $\bar{q}(\mathbf{x}, t)$ is the mass-weighted vertical average of q over the layer, and

$$\bar{\psi}_q = \frac{\partial}{\partial x_1}\left(A_H\left(\frac{1}{m_1}\frac{\partial \bar{q}}{\partial x_1}\right)m_2\Delta p\right) + \frac{\partial}{\partial x_2}\left(A_H\left(\frac{1}{m_2}\frac{\partial \bar{q}}{\partial x_2}\right)m_1\Delta p\right). \tag{7.108}$$

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

The term $\bar{q}\Delta p$ is the quantity of tracer per unit horizontal area, times g. The terms involving $\dot{s}p_s q$ represent the rate of transport of tracer across coordinate surfaces due to material changes in s, *i.e.*, due to the movement of coordinate surfaces relative to the fluid. The velocity components $\bar{u}$ and $\bar{v}$ in the momentum equations (7.101) and (7.104) are components of momentum per unit mass, so these quantities can be regarded as examples of the quantity $\bar{q}$, except for the forcing due to the Coriolis, pressure, and applied stress terms. If those terms are deleted from equations (7.101) and (7.104), then the resulting equations have the same form as (7.107).

7.18. Planar and spherical coordinates

The layer mass equation (7.38), the layer momentum equations (7.101) and (7.104), and the layer tracer equation (7.107) are derived under the assumption that the horizontal coordinates are general orthogonal curvilinear coordinates $\mathbf{x} = (x_1, x_2)$. The present subsection discusses the forms of these equations in the special cases of planar and spherical coordinates.

For the case of planar coordinates, *i.e.*, Cartesian coordinates on a tangent plane, x_1 and x_2 can be taken as literal measures of distance. The metric coefficients (7.2) are then given by $m_1 = 1$ and $m_2 = 1$, and in addition $G = m_1 m_2 = 1$. The factors m_1, m_2, and G can then be deleted from the partial differential equations (7.38), (7.101), (7.104), and (7.107). In the Coriolis/geometric parameter (7.102), the geometric terms (involving $\partial \mathbf{i}/\partial x_1$ and $\partial \mathbf{i}/\partial x_2$) are equal to zero, and the values of the Coriolis parameter can be taken from the formula $f = 2\Omega \sin\theta$ given in (7.94).

Spherical coordinates are defined in equation (7.6). In that case the horizontal coordinates are the longitude $x_1 = \lambda$ and the latitude $x_2 = \theta$. The metric coefficients (7.7) are $m_1 = m_\lambda = r\cos\theta \approx a\cos\theta$ and $m_2 = m_\theta = r \approx a$, where a is the radius of the sphere, and thus $G = m_\lambda m_\theta = r^2 \cos\theta \approx a^2 \cos\theta$. It follows from (7.6) and (7.3) that

$$\mathbf{i}(\lambda,\theta,t) = \frac{1}{m_\lambda}\frac{\partial \mathbf{r}}{\partial \lambda} = \big(-\sin(\lambda+\Omega t),\ \cos(\lambda+\Omega t),\ 0\big), \tag{7.109}$$

$$\mathbf{j}(\lambda,\theta,t) = \frac{1}{m_\theta}\frac{\partial \mathbf{r}}{\partial \theta} = \big(-\sin\theta\cos(\lambda+\Omega t),\ -\sin\theta\sin(\lambda+\Omega t),\ \cos\theta\big).$$

A calculation reproduces the formula (7.94), $f = 2\,\mathbf{j}\cdot\frac{\partial \mathbf{i}}{\partial t} = 2\Omega\sin\theta$. The geometric parameter in (7.102) is

$$\mathbf{j}\cdot\left(\bar{u}\frac{1}{m_\lambda}\frac{\partial \mathbf{i}}{\partial \lambda} + \bar{v}\frac{1}{m_\theta}\frac{\partial \mathbf{i}}{\partial \theta}\right) = \mathbf{j}\cdot\left[\frac{\bar{u}}{r\cos\theta}\big(-\cos(\lambda+\Omega t),\ -\sin(\lambda+\Omega t),\ 0\big)\right]$$
$$= \frac{\bar{u}\tan\theta}{r},$$

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

so the combined Coriolis/geometric parameter (7.102) is

$$\bar{f} = 2\Omega \sin\theta + \frac{\bar{u}\tan\theta}{r}. \tag{7.110}$$

As noted in Section 7.15, $2\Omega \approx 1.46 \times 10^{-4}$ sec^{-1} in the case of the rotating Earth. For a fluid velocity $\bar{u} = 1$ m/sec, the factor $\bar{u}/r$ in the geometric parameter in (7.110) is approximately 1.6×10^{-7} sec^{-1}. In the low and middle latitudes, the Coriolis parameter thus greatly exceeds the geometric parameter for spherical coordinates. However, as the latitude θ approaches 90°, $\tan\theta$ increases without bound, owing to the singularity in the spherical coordinate system at that point. As pointed out in Section 6.1, this singularity corresponds to a convergence of grid lines at the poles, and for a global ocean model it is better to use a coordinate system that does not produce a convergence of grid lines within the fluid domain.

8. Summary

The goal of this paper is to provide an introduction to the mathematical and computational modelling of ocean circulation. The paper includes a detailed derivation of partial differential equations that describe the conservation of mass, momentum, and tracers for a hydrostatic and stratified fluid on a rotating spheroid. Also included are a description of some physical properties of oceanic flows, a discussion of some of the issues that are encountered when the governing equations are solved numerically, and a summary of some of the numerical methods that are used in this field.

In the derivation of the governing equations, it is assumed that the vertical coordinate is a generalized coordinate that includes the cases of level, isopycnic, sigma, and hybrid coordinates. Each of these coordinates presents advantages and disadvantages that are discussed here. Among the considerations in the choice and implementation of a vertical coordinate is the need to represent accurately the vertical exchanges between water masses that are often subtle but nevertheless important for long-term integrations. It is also assumed that the horizontal coordinates are general orthogonal curvilinear coordinates instead of spherical coordinates, in anticipation of using a grid that is suitable for a global model.

Several numerical issues are discussed here. These include time-stepping methods and the numerical treatment of multiple time scales; spatial grids and spatial discretization; the solution of the nonlinear equations for the conservation of momentum; and the numerical simulation of advection, especially in the context of maintaining nonnegative solutions and transporting the multiple tracers that are typically included in numerical models of ocean circulation.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

Acknowledgements

Over a period of several years I have benefited greatly from conversations with numerous ocean modellers, especially members of the physical oceanography group at Oregon State University, members of the ocean and climate modelling group at Los Alamos National Laboratory, and attendees at the annual Layered Ocean Model Users' Workshop at the University of Miami. Special thanks are extended to Rainer Bleck, Andrew Bennett, Mats Bentsen, Roland de Szoeke, John Dukowicz, Roger Samelson, and Scott Springer. This work was supported by National Science Foundation grant DMS-0511782.

REFERENCES

A. J. Adcroft, C. N. Hill and J. C. Marshall (1999), A new treatment of the Coriolis terms in C-grid models at both high and low resolutions, *Monthly Weather Review* **127**, 1928–1936.

A. Adcroft, J.-M. Campin, C. Hill and J. Marshall (2004), Implementation of an atmosphere-ocean general circulation model on the expanded spherical cube, *Monthly Weather Review* **132**, 2845–2863.

A. Arakawa and V. R. Lamb (1977), Computational design of the basic dynamical processes of the UCLA general circulation model, *Methods in Computational Physics* **17**, 173–265.

R. Asselin (1972), Frequency filter for time integrations, *Monthly Weather Review* **100**, 487–490.

A. F. Bennett (2002), *Inverse Modeling of the Ocean and Atmosphere*, Cambridge University Press, Cambridge.

R. Bleck (1978), Finite difference equations in generalized vertical coordinates, Part I: Total energy conservation, *Contributions to Atmospheric Physics* **51**, 360–372.

R. Bleck (2002), An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates, *Ocean Modelling* **4**, 55–88.

R. Bleck (2005), On the use of hybrid vertical coordinates in ocean circulation modeling, in *An Integrated View of Oceanography: Ocean Weather Forecasting in the 21st Century* (E. Chassignet and J. Verron, eds), Kluwer, Dordrecht, pp. 109–126.

R. Bleck and L. T. Smith (1990), A wind-driven isopycnic coordinate model of the north and equatorial Atlantic Ocean 1: Model development and supporting experiments, *J. Geophysical Research* **95C**, 3273–3285.

R. Bleck, C. Rooth, D. Hu and L. T. Smith (1992), Salinity-driven thermocline transients in a wind- and thermohaline-forced isopycnic coordinate model of the North Atlantic, *J. Physical Oceanography* **22**, 1486–1505.

A. F. Blumberg and G. L. Mellor (1987), A description of a three-dimensional coastal ocean circulation model, in *Three-dimensional Coastal Ocean Models* (N. Heaps, ed.), American Geophysical Union, Washington, DC.

K. Bryan (1969), A numerical method for the study of the circulation of the world ocean, *J. Comput. Phys.* **4**, 347–376.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

A. J. Chorin and J. E. Marsden (1990), *A Mathematical Introduction to Fluid Mechanics*, 2nd edn, Springer, New York.

R. E. Davis (1994), Diapycnal mixing in the ocean: equations for large-scale budgets, *J. Physical Oceanography* **24**, 777–800.

R. A. de Szoeke (2000), Equations of motion using thermodynamic coordinates, *J. Physical Oceanography* **30**, 2814–2829.

R. A. de Szoeke and S. R. Springer (2003), A diapycnal diffusion algorithm for isopycnal ocean circulation models with special application to mixed layers, *Ocean Modelling* **5**, 297–323.

R. A. de Szoeke, S. R. Springer and D. M. Oxilia (2000), Orthobaric density: A thermodynamic variable for ocean circulation studies, *J. Physical Oceanography* **30**, 2830–2852.

J. K. Dukowicz (1995), Mesh effects for Rossby waves, *J. Comput. Phys.* **119**, 188–194.

J. K. Dukowicz (2004), HYPOP Summary (Hybrid vertical coordinate version of POP) Current C-Grid Version. Technical Report LA-UR-04-8586, Los Alamos National Laboratory.

J. K. Dukowicz (2006), Structure of the barotropic mode in layered ocean models, *Ocean Modelling* **11**, 49–68.

J. K. Dukowicz and J. R. Baumgardner (2000), Incremental remapping as a transport/advection algorithm, *J. Comput. Phys.* **160**, 318–335.

J. K. Dukowicz and R. D. Smith (1994), Implicit free-surface method for the Bryan–Cox–Semtner ocean model, *J. Geophysical Research* **99**, 7991–8014.

D. R. Durran (1999), *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*, Springer, New York.

R. C. Easter (1993), Two modified versions of Bott's positive-definite numerical advection scheme, *Monthly Weather Review* **121**, 297–304.

C. Eden and J. Willebrand (1999), Neutral density revisited, *Deep-Sea Research Part II* **46**, 33–54.

P. R. Gent and J. C. McWilliams (1990), Isopycnal mixing in ocean circulation models, *J. Physical Oceanography* **20**, 150–155.

P. R. Gent, J. Willebrand, T. J. McDougall and J. C. McWilliams (1995), Parameterizing eddy-induced tracer transports in ocean circulation models, *J. Physical Oceanography* **25**, 463–474.

A. E. Gill (1982), *Atmosphere-Ocean Dynamics*, Academic Press, San Diego.

S. M. Griffies (2004), *Fundamentals of Ocean Climate Models*, Princeton University Press, Princeton, NJ.

S. M. Griffies, C. Bönig, F. O. Bryan, E. P. Chassignet, R. Gerdes, H. Hasumi, A. Hirst, A.-M. Treguier and D. Webb (2000*a*), Developments in ocean climate modeling, *Ocean Modelling* **2**, 123–192.

S. M. Griffies, R. C. Pacanowski and R. W. Hallberg (2000*b*). Spurious diapycnal mixing associated with advection in a z-coordinate ocean model, *Monthly Weather Review* **128**, 538–564.

S. M. Griffies, M. J. Harrison, R. C. Pacanowski and A. Rosati (2004), A technical guide to MOM4. NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, available on-line.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

S. M. Griffies, A. Gnanadesikan, K. W. Dixon, J. P. Dunne, R. Gerdes, M. J. Harrison, A. Rosati, J. L. Russell, B. L. Samuels, M. J. Spelman, M. Winton, and R. Zhang (2005), Formulation of an ocean model for global climate simulations, *Ocean Science* **1**, 45–79.

D. B. Haidvogel and A. Beckmann (1999), *Numerical Ocean Circulation Modeling*, Imperial College Press, London.

R. Hallberg (1997), Stable split time stepping schemes for large-scale ocean modeling, *J. Comput. Phys.* **135**, 54–65.

R. Hallberg (2000), Time integration of diapycnal diffusion and Richardson number-dependent mixing in isopycnal coordinate ocean models, *Monthly Weather Review* **128**, 1402–1419.

R. Hallberg (2005), A thermobaric instability of Lagrangian vertical coordinate ocean models, *Ocean Modelling* **8**, 279–300.

M. W. Hecht (2006), Forward-in-time upwind-weighted methods in ocean modelling, *Int. J. Numer. Meth. Fluids* **50**, 1159–1173.

R. L. Higdon (2002), A two-level time-stepping method for layered ocean circulation models, *J. Comput. Phys.* **177**, 59–94.

R. L. Higdon (2005), A two-level time-stepping method for layered ocean circulation models: Further development and testing, *J. Comput. Phys.* **206**, 463–504.

R. L. Higdon and A. F. Bennett (1996), Stability analysis of operator splitting for large-scale ocean modeling, *J. Comput. Phys.* **123**, 311–329.

R. L. Higdon and R. A. de Szoeke (1997), Barotropic–baroclinic time splitting for ocean circulation modeling, *J. Comput. Phys.* **135**, 30–53.

C. W. Hirt, A. A. Amsden and J. L. Cook (1974), An arbitrary Lagrangian–Eulerian computing method for all flow speeds, *J. Comput. Phys.* **14**, 227–253.

J. R. Holton (1992), *An Introduction to Dynamic Meteorology*, 3rd edn, Academic Press, San Diego.

Y.-J. Hsu and A. Arakawa (1990), Numerical modeling of the atmosphere with an isentropic vertical coordinate, *Monthly Weather Review* **118**, 1933–1959.

Intergovernmental Panel on Climate Change (2001), *IPCC Third Assessment Report: Climate Change 2001*, on-line, `http://www.ipcc.ch/`

M. Iskandarani, D. B. Haidvogel and J. C. Levin (2003), A three-dimensional spectral element model for the solution of the hydrostatic primitive equations, *J. Comput. Phys.* **186**, 397–425.

M. Iskandarani, J. C. Levin, B.-J. Choi and D. B. Haidvogel (2005), Comparison of advection schemes for high-order h-p finite element and finite volume methods, *Ocean Modelling* **10**, 233–252.

P. D. Killworth, D. Stainforth, D. J. Webb and S. M. Paterson (1991), The development of a free-surface Bryan–Cox–Semtner ocean model, *J. Physical Oceanography* **21**, 1333–1348.

J. R. Ledwell, A. J. Watson and C. S. Law (1993), Evidence for slow mixing across the pycnocline from an open-ocean tracer-release experiment, *Nature* **364**, 701–703.

B. P. Leonard (1979), A stable and accurate convective modeling procedure based on quadratic upstream interpolation, *Comput. Methods Appl. Mech. Engrg.* **19**, 59–98.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

R. J. LeVeque (2003), *Finite Volume Methods for Hyperbolic Problems*, Cambridge University Press, Cambridge.
S. Levitus (1982), *Climatological Atlas of the World Ocean*, US Department of Commerce, National Oceanic and Atmospheric Administration, Rockville, MD.
S.-J. Lin and R. R. Rood (1996), Multidimensional flux-form semi-Lagrangian transport schemes, *Monthly Weather Review* **124**, 2046–2070.
W. H. Lipscomb and E. C. Hunke (2004), Modeling sea ice transport using incremental remapping, *Monthly Weather Review* **132**, 1341–1354.
W. H. Lipscomb and T. D. Ringler (2005), An incremental remapping transport scheme on a spherical geodesic grid, *Monthly Weather Review* **133**, 2335–2350.
T. J. McDougall (1987), Neutral surfaces, *J. Physical Oceanography* **17**, 1950–1967.
T. J. McDougall and W. K. Dewar (1998), Vertical mixing and cabbeling in layered models, *J. Physical Oceanography* **28**, 1458–1480.
J. Marshall and F. Schott (1999), Open-ocean convection: Observations, theory, and models, *Reviews of Geophysics* **37**, 1–64.
J. Marshall, C. Hill, L. Perelman and A. Adcroft (1997), Hydrostatic, quasi-hydrostatic, and nonhydrostatic ocean modeling, *J. Geophysical Research* **102**(C3), 5733–5752.
R. N. Miller (2006), *Numerical Modeling of Ocean Circulation*, Cambridge University Press, to appear.
R. B. Montgomery (1937), A suggested method for representing gradient flow in isentropic surfaces, *Bull. Amer. Meteorological Soc.* **18**, 210–212.
R. J. Murray (1996) Explicit generation of orthogonal grids for ocean models, *J. Comput. Phys.* **126**, 251–273.
D. Nechaev and M. Yaremchuk (2004), On the approximation of the Coriolis terms in C-grid models, *Monthly Weather Review* **132**, 2283–2289.
C. C. Pain, M. D. Piggott, A. J. H. Goddard, F. Fang, G. J. Gorman, D. P. Marshall, M. D. Eaton, P. W. Power and C. R. E. de Oliveira (2005), Three-dimensional unstructured mesh ocean modelling, *Ocean Modelling* **10**, 5–33.
J. Pedlosky (1987), *Geophysical Fluid Dynamics*, 2nd edn, Springer, New York.
J. Pedlosky (1996), *Ocean Circulation Theory*, Springer, Berlin.
J. Pedlosky (2003), *Waves in the Ocean and Atmosphere*, Springer, Berlin.
M. Rancic, R. J. Purser and F. Mesinger (1996), A global shallow-water model using an expanded spherical cube: Gnomonic versus conformal coordinates, *Quarterly J. Royal Meteorological Soc.* **122**, 959–982.
T. D. Ringler and D. A. Randall (2002), A potential enstrophy and energy conserving numerical scheme for solution of the shallow-water equations on a geodesic grid, *Monthly Weather Review* **130**, 1397–1410.
J. A. Rossmanith (2006), A wave propagation method for hyperbolic systems on the sphere, *J. Comput. Phys.* **213**, 629–658.
R. Sadourny (1975), The dynamics of finite-difference models of the shallow-water equations, *J. Atmospheric Sci.* **32**, 680–689.
C. Schär and P. K. Smolarkiewicz (1996), A synchronous and iterative flux-correction formalism for coupled transport equations, *J. Comput. Phys.* **128**, 101–120.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press

P. Schopf and A. Loughe (1995), A reduced gravity isopycnal ocean model, *Monthly Weather Review* **123**, 2839–2863.

A. J. Semtner (1997), Introduction to 'A numerical method for the study of the circulation of the world ocean', *J. Comput. Phys.* **135**, 149–153.

A. J. Semtner and R. M. Chervin (1992), Ocean general circulation from a global eddy-resolving model, *J. Geophysical Research* **97**, 5493–5550.

A. F. Shchepetkin and J. C. McWilliams (2005), The regional oceanic modeling system (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model, *Ocean Modelling* **9**, 347–404.

R. D. Smith, S. Kortas and B. Meltz (1995), Curvilinear coordinates for global ocean models. Technical Report LA-UR-95-1146, Los Alamos National Laboratory.

P. K. Smolarkiewicz and L. G. Margolin (1998), MPDATA: A finite-difference solver for geophysical flows, *J. Comput. Phys.* **140**, 459–480.

S. Sun, R. Bleck, C. Rooth, J. Dukowicz, E. Chassignet and P. Killworth (1999), Inclusion of thermobaricity in isopycnic-coordinate ocean models, *J. Physical Oceanography* **29**, 2719–2729.

F. J. Winninghoff (1968), On the adjustment toward a geostrophic balance in a simple primitive equation model with application to the problems of initialization and objective analysis, PhD thesis, Department of Meteorology, University of California, Los Angeles.

S. T. Zalesak (1979), Fully multidimensional flux-corrected transport algorithms for fluids, *J. Comput. Phys.* **31**, 335–362.

https://doi.org/10.1017/S0962492906250013 Published online by Cambridge University Press