1

Introduction

This book derives and classifies the most common dynamic equations used in physical oceanography, from the planetary geostrophic equations that describe the wind and thermohaline driven circulations to the equations of small-scale motions that describe three-dimensional turbulence and double diffusive phenomena. It does so in a systematic manner and within a common framework. It first establishes the basic dynamic equations that describe all oceanic motions and then derives reduced equations, emphasizing the assumptions made and physical processes eliminated.

The basic equations of oceanic motions consist of:

- the thermodynamic specification of sea water;
- the balance equations for mass, momentum, and energy;
- the molecular flux laws; and
- the gravitational field equation.

These equations are well established and experimentally proven. However, they are so general and so all-encompassing that they become useless for specific practical applications. One needs to consider approximations to these equations and derive equations that isolate specific types or scales of motion. The basic equations of oceanic motion form the solid starting point for such derivations.

In order to derive and present the various approximations in a systematic manner we use the following concepts and organizing principles:

- distinction between properties of fluids and flows;
- distinction between prognostic and diagnostic variables;
- adjustment by wave propagation;
- modes of motion;
- Reynolds decomposition and averaging;
- asymptotic expansion;
- geometric, thermodynamic, and dynamic approximations; and
- different but equivalent representations,

which are discussed in the remainder of this introduction.

First, we distinguish between properties of the fluid and properties of the flow. The equation of state, which determines the density of sea water, is a fluid property. To understand the impact that the choice of the equation of state has on the fluid flow we consider, for ideal fluid conditions, five different equations of state:

- a two-component fluid (the density depends on pressure, specific entropy and salinity);
- a one-component fluid (the density depends on pressure and specific entropy);
- a homentropic fluid (the density depends on pressure only);
- an incompressible fluid (the density depends on specific entropy only); and
- a homogeneous fluid (the density is constant).

Sea water is of course a two-component fluid, but many flows evolve as if sea water were a one-component, homentropic, incompressible, or homogeneous fluid, under appropriate conditions.

We further distinguish between:

- prognostic variables; and
- diagnostic variables.

A prognostic variable is governed by an equation that determines its time evolution. A diagnostic variable is governed by an equation that determines its value (at each time instant).

When a fluid flow is disturbed at some point it responds or adjusts by emitting waves. These waves communicate the disturbance to other parts of the fluid. The waves have different restoring mechanisms: compressibility, gravitation, stratification, and (differential) rotation. We derive the complete set of linear waves in a stratified fluid on a rotating sphere, which are:

- sound (or acoustic) waves;
- surface and internal gravity waves; and
- barotropic and baroclinic Rossby waves.

A temperature–salinity wave needs to be added when both temperature and salinity become dynamically active, as in double diffusion. The assumption of instantaneous adjustment eliminates certain wave types from the equations and forms the basis of many approximations. We regard these waves as linear manifestations of acoustic, gravity, Rossby, and temperature–salinity *modes of motion* though this concept is not well defined. A general nonlinear flow cannot uniquely be decomposed into such modes of motion. A well-defined property of a general nonlinear flow is, however, whether or not it carries potential vorticity. We thus distinguish between:

- the zero potential vorticity; and
- the potential vorticity carrying or vortical mode of motion.

This is a generalization of the fluid dynamical distinction between irrotational flows and flows with vorticity. Sound and gravity waves are linear manifestations of the zero potential vorticity mode and Rossby waves are manifestations of the vortical mode. We follow carefully the structure of the vorticity and potential vorticity equations and the structure of these two modes through the various approximations of this book.

The basic equations of oceanic motions are nonlinear. As a consequence, all modes and scales of motion interact, not only among themselves but also with the surrounding atmosphere and the solid earth. Neither the ocean as a whole nor any mode or scale of motion within it can be isolated rigorously. Any such isolation is approximate at best. There are two basic techniques to derive approximate dynamical equations:

- Reynolds averaging; and
- asymptotic expansions.

Reynolds averaging decomposes all field variables into a mean and a fluctuating component, by means of a space-time or ensemble average. Applying the same average to the equations of motion, one arrives at a set of equations for the mean component and at a set of equations for the fluctuating component. These two sets of equations are not closed. They are coupled. The equations for the mean component contain eddy (or subgridscale) fluxes that represent the effect of the fluctuating component on the mean component. The equations for the fluctuating component contain background fields that represent the effect of the mean component on the fluctuating component. Any attempt to decouple these two sets of equations must overcome the closure problem. To derive a closed set of equations for the mean component one must parametrize the eddy fluxes in terms of mean quantities. To derive a closed set for the fluctuating component one must prescribe the background fields. Prescribing background fields sounds less restrictive than parametrizing eddy fluxes, but both operations represent closures, and closures are approximations.

Asymptotic expansions are at the core of approximations. Formally, they require the scaling of the independent and dependent variables, the non-dimensionalization of the dynamical equations, and the identification of the non-dimensional parameters that characterize these equations. The non-dimensional equations can then be studied in the limit that any of these dimensionless parameters approaches zero by an asymptotic expansion with respect to this parameter. Often, these asymptotic expansions are short-circuited by simply neglecting certain terms in the equations. High-frequency waves are not affected much by the Earth's rotation and can be studied by simply neglecting the Earth's rotation in the equations, rather than by going through a pedantic asymptotic expansion with respect to a parameter that reflects the ratio of the Earth's rotation rate to the wave frequency. Similarly, when encountering a situation where part of the flow evolves slowly while another part evolves rapidly one often eliminates the fast variables by setting their time derivative to zero. The fast variables adjust so rapidly that it appears to be instantaneous when viewed from the slowly evolving part of the flow. Nevertheless, any of these heuristic approximations is justified by an underlying asymptotic expansion.

The basic equations of oceanic motions are characterized by a large number of dimensionless parameters. It is neither useful nor practical to explore the complete multi-dimensional parameter space spanned by the parameters. Only a limited area of this space is occupied by actual oceanic motions. To explore this limited but still fairly convoluted subspace in a somewhat systematic manner we distinguish between:

- geometric approximations;
- thermodynamic approximations; and
- dynamic approximations.

Geometric approximations change the underlying geometric space in which the oceanic motions occur. This geometric space is a Euclidean space, in the non-relativistic limit. It can be represented by whatever coordinates one chooses, Cartesian coordinates being the simplest ones. Moreover, the dynamic equations can be formulated in coordinate-invariant form using vector and tensor calculus, and we use such invariant notation wherever appropriate. It is, however, often useful (and for actual calculations necessary) to express the dynamic equations in a specific coordinate system. Then metric coefficients (or scale factors) appear in the equations of motions. These coefficients are simply a consequence of introducing a specific coordinate system. A geometric approximation is implemented when these metric coefficients and hence the underlying Euclidean space are altered. Such geometric approximations include the spherical, *beta*-plane, and f-plane approximations. They rely on the smallness of parameters such as the eccentricity of the geoid, the ratio of the ocean depth to the radius of the Earth, and the ratio of the horizontal length scale to the radius of the Earth. If the smallness of these parameters is only exploited in the metric coefficients, but not elsewhere in the equations, then one does not distort the general properties of the equations. The pressure force remains a gradient; its integral along a closed circuit and its curl vanish exactly.

Thermodynamic approximations are approximations to the thermodynamic properties of sea water. Most important are approximations to the equation of state. They determine to what extent sea water can be treated as a one-component, incompressible, homentropic, or homogeneous fluid, with profound effects on the dynamic evolution and permissible velocity and vorticity fields. Auxiliary thermodynamic approximations assume that other thermodynamic (and phenomenological) coefficients in the basic equations of oceanic motions are not affected by the flow and can be regarded as constant. This eliminates certain nonlinearities from the equations.

Dynamical approximations affect the basic equations of oceanic motions more directly. They eliminate certain terms such as the tendency, advection, or friction term and the associated processes. A common thread of these approximations is that flows at large horizontal length scales and long time scales are approximately in hydrostatic and geostrophic balance and two-dimensional in character, owing to the influence of rotation and stratification. These balanced flows are characterized by small aspect ratios and small Rossby and Ekman numbers (which are ratios of the advective and frictional time scales to the time scale of rotation). As the space and time scales become smaller and the Rossby and Ekman numbers larger, the flow becomes less balanced and constrained and more three-dimensional in character. At the smallest scales the influence of rotation and stratification becomes negligible and the flow becomes fully three-dimensional.

The various approximations are depicted in more detail in Figure 1.1. We first separate acoustic and non-acoustic motions. The acoustic mode could be obtained by considering irrotational motion in a homentropic ocean. Similarly, the non-acoustic modes could be isolated by assuming sea water to be an incompressible fluid. Both these assumptions are much too strong for oceanographic purposes. Instead, one separates acoustic and non-acoustic modes by assuming that the (Lagrangian) time scale of the acoustic mode is much shorter than the (Lagrangian) time scale of the non-acoustic mode. Acoustic motions are fast motions and non-acoustic motions are slow motions in a corresponding two-time scale expansion. Such an expansion implies that when one considers the evolution of the acoustic pressure field one may neglect the slow temporal changes of the background non-acoustic pressure field. This, together with some ancillary assumptions, leads to the acoustic wave equation that forms the basis for acoustic studies of the ocean.

When one considers non-acoustic motions, the two-time scale expansion implies that one can neglect the fast temporal changes of the acoustic pressure in the pressure equation. The pressure field adjusts so rapidly that it appears to be instantaneous when viewed from the slow non-acoustic motions. This elimination of sound waves from the equations is called the anelastic approximation. The resulting equations do not contain sound waves but sea water remains compressible. A consequence of the anelastic approximation is that the pressure is no longer determined prognostically but diagnostically by the solution of a three-dimensional Poisson equation. The anelastic approximation is augmented by approximations that utilize the facts that the density of the ocean does not vary much at a point and from the surface to the bottom. Together, these approximations comprise the Boussinesq approximation, which is at the heart of all that follows. Its major result is that the velocity field is non-divergent or solenoidal. The flow (as opposed to the fluid) is now incompressible.

Next we introduce the shallow water approximation. It assumes that the aspect ratio, i.e., the ratio between the vertical and horizontal length scale of the flow,

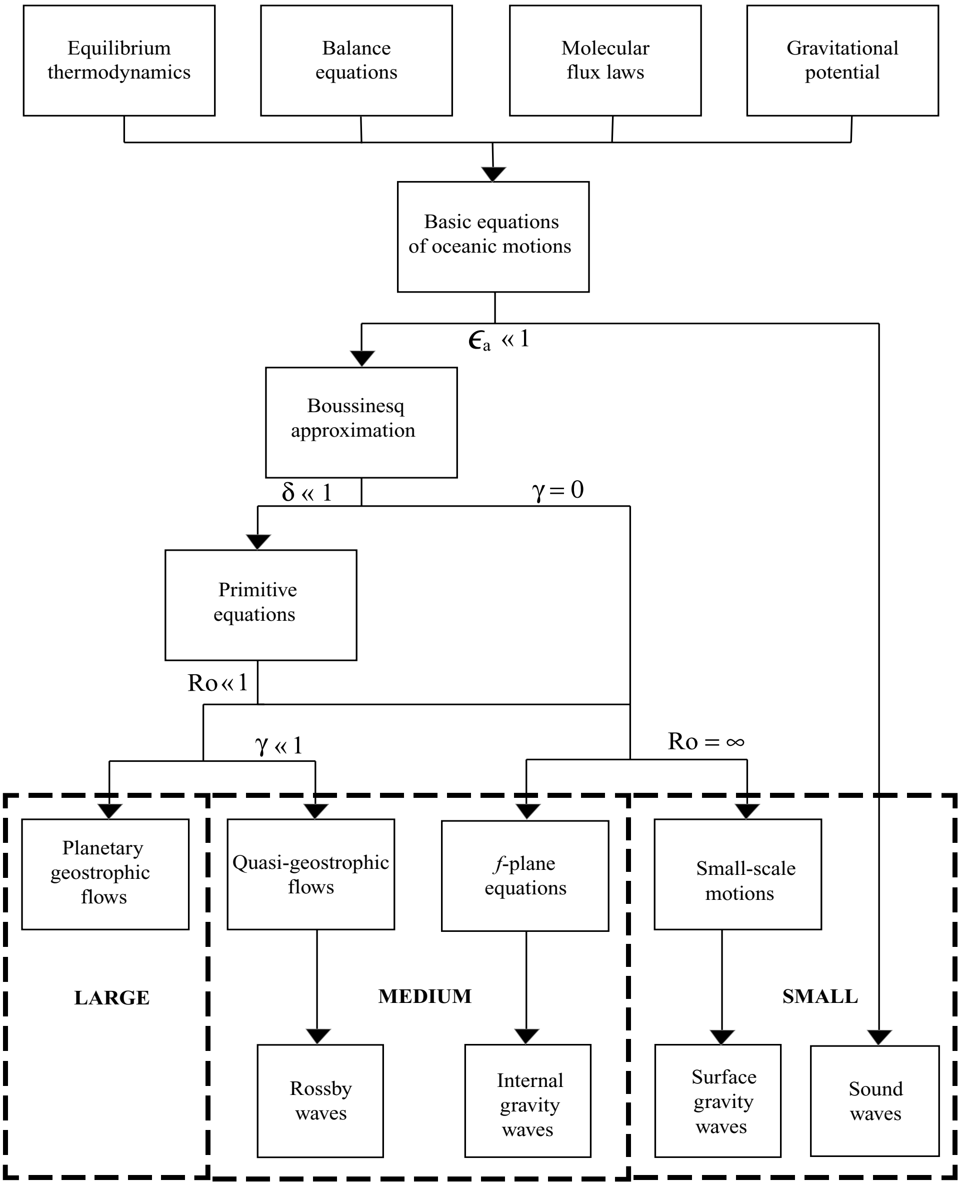

Figure 1.1. Diagram of the overall organization of this book. The parameter ϵ_a represents the ratio of the acoustic to the non-acoustic time scale, δ the aspect ratio, γ the ratio of the horizontal length scale to the radius of the Earth, and Ro the Rossby number.

is small. It leads to the primitive equations. The major simplification is that the vertical momentum balance reduces to the hydrostatic balance where the vertical pressure gradient is balanced by the gravitational force. The pressure is thus no longer determined by the solution of a three-dimensional Poisson equation but by an ordinary differential equation. The shallow water approximation also eliminates the local meridional component of the planetary vorticity, a fact referred to as the traditional approximation.

For small Rossby and Ekman numbers, the horizontal momentum balance can be approximated by the geostrophic balance, the balance between Coriolis and pressure force. This geostrophic approximation has far-reaching consequences. It eliminates gravity waves or the gravity mode of motion. The velocity field adjusts instantaneously. Geostrophic motions carry potential vorticity. They represent the vortical mode of motion. Their evolution is governed by the potential vorticity equation. One has to distinguish between planetary or large-scale geostrophic flows and quasi-geostrophic or small-scale geostrophic flows. For planetary geostrophic flows the potential vorticity is given by f/H, where f is the Coriolis frequency and H the ocean depth. Quasi-geostrophic flows employ two major additional assumptions. One is that the horizontal length scale of the flow is much smaller than the radius of the Earth. This assumption is exploited in the beta-plane approximation. The second assumption is that the vertical displacement is much smaller than the vertical length scale (or that the vertical strain is much smaller than 1). This assumption implies a linearization of the density equation and the flow becomes nearly two-dimensional. The quasi-geostrophic potential vorticity consists of contributions from the relative vorticity, the planetary vorticity, and the vertical strain.

As the length and time scales of the flow decrease further (or as the Rossby number, aspect ratio, and vertical strain all increase) one arrives at a regime where rotation and stratification are still important but not strong enough to constrain the motion to be in hydrostatic and geostrophic balance and nearly two-dimensional. The horizontal length scales become so small that the f-plane approximation can be applied. These f-plane equations contain all modes of motion except the acoustic one. The f-plane motions offer the richest variety of dynamical processes and phenomena. The zero potential vorticity and vortical mode can be isolated in limiting cases.

At the smallest scales one arrives at the equations of regular fluid dynamics. The fact that the motion actually occurs on a rotating sphere becomes inconsequential. The motions may be directly affected by molecular friction, diffusion, and conduction. The temperature–salinity mode (which is at the heart of double diffusive phenomena) again emerges since the molecular diffusivities for heat and salt differ. A major distinction is made between laminar flows for low Reynolds numbers and turbulent flows for high Reynolds numbers. If all buoyancy effects are neglected, one obtains the Navier–Stokes equations that are used to study motions such as three-dimensional isotropic turbulence and nonlinear water waves.

Tidal motions fall somewhat outside this classification since they are defined not by their scales but by their forcing. They are caused by the gravitational potential of the Moon and Sun. The tidal force is a volume force and approximately constant throughout the water column. It only affects the barotropic component of the flow. Tidal motions are thus described by the one-layer shallow water equations, generally called Laplace tidal equations. Since the tidal force is the gradient of the tidal potential it does not induce any vorticity into tidal flows. Tidal flows represent the zero potential vorticity mode though this fact has not been exploited in any systematic way. The tidal potential also causes tides of the solid but elastic earth, called Earth tides. The moving tidal water bulge also causes an elastic deformation of the ocean bottom, called the load tides. The moving bulge modifies the gravitational field of the ocean, an effect called gravitational self-attraction. All these effects need to be incorporated into Laplace tidal equations.

These approximations are overlaid with a triple decomposition into large-, medium-, and small-scale motions that arises from two Reynolds decompositions: one that separates large- from medium- and small-scale motions and one that separates large and medium motions from small-scale motions. For large-scale motions one must parametrize the eddy fluxes arising from medium- and small-scale motions; for medium-scale motions one must parametrize the eddy fluxes from small-scale motions and specify the large-scale background fields; for small-scale motions the subgridscale fluxes are given by the molecular flux laws and one must only specify the large- and medium-scale background fields. The medium-scale motions are the most challenging ones from this point of view. Eddy fluxes have to be parametrized and background fields have to be specified. In this book we only introduce the standard parametrizations of the eddy fluxes in terms of eddy diffusion and viscosity coefficients. Efforts are underway to improve these parametrizations but have not arrived yet at a set of canonical parametrizations.

Of course, the geometric, thermodynamic, and dynamic approximations do not match exactly the triple Reynolds decomposition since the Reynolds averaging scales are not fixed but can be adapted to circumstances. The same dynamical kind of motion might well fall on different sides of a Reynolds decomposition. Surface gravity waves might neglect rotation and the sphericity of the Earth and may be regarded as small-scale motions dynamically. Their subgridscale mechanisms may, however, be either molecular friction or turbulent friction in bottom boundary layers and wave breaking, depending on whether they are short or long waves. In Figure 1.1, we overlaid the Reynolds categories for general orientation only (and put surface gravity waves in the small-scale category).

The systematic representation of the equations of oceanic motions is further complicated by the fact that the equations can be represented in many different but equivalent forms. One can express them in different coordinate systems or in different but equivalent sets of state variables. Of particular relevance is the representation of the vertical structure of the flow. Here, we discuss:

- the decomposition into barotropic and baroclinic components;
- the representation in isopycnal coordinates;
- the representation in sigma coordinates;
- layer models; and
- the projection onto vertical normal modes.

All these representations fully recover the continuous vertical structure of the flow field, except for the layer models that can be obtained by discretizing the equations in isopycnal coordinates. These representations do not involve any additional approximations but their specific form often invites ancillary approximations.

These are the basic concepts and organizing principles to present the various equations of oceanic motions covered in this book. The basic concepts and formulae of equilibrium thermodynamics, of vector and tensor analysis, of orthogonal curvilinear coordinate systems, and of the kinematics of fluid motion and waves, and conventions and notation are covered in appendices.

2

Equilibrium thermodynamics of sea water

The basic equations of oceanic motion assume local thermodynamic equilibrium. The ocean is viewed as consisting of many fluid parcels. Each of these fluid parcels is assumed to be in thermodynamic equilibrium though the ocean as a whole is far from thermodynamic equilibrium. Later we make the continuum hypothesis and assume that these parcels are sufficiently small from a macroscopic point of view to be treated as points but sufficiently large from a microscopic point of view to contain enough molecules for equilibrium thermodynamics to apply. This chapter considers the equilibrium thermodynamics that holds for each of these fluid parcels or points. The thermodynamic state is described by thermodynamic variables. Most of this chapter defines these thermodynamic variables and the relations that hold among them. An important point is that sea water is a two-component system, consisting of water and sea salt. Gibbs' phase rule then implies that the thermodynamic state of sea water is completely determined by the specification of three independent thermodynamic variables. Different choices can be made for these independent variables. Pressure, temperature, and salinity are one common choice. All other variables are functions of these independent variables. In principle, these functions can be derived from the microscopic properties of sea water, by means of statistical mechanics. This has not been accomplished yet. Rather, these functions must be determined empirically from measurements and are documented in figures, tables, and numerical formulae. We do not present these figures, tables, and formulae in any detail. They can be found in books, articles, and reports such as Montgomery (1957), Fofonoff (1962, 1985), UNESCO (1981), and Siedler and Peters (1986). We list the quantities that have been measured and the algorithms that can, in principle, be used to construct all other thermodynamic variables. In the final section we discuss the mixing of water parcels at constant pressure. An introduction into the concepts of equilibrium thermodynamics is given in Appendix A.