Chapter 7

100 Years of the Ocean General Circulation

CARL WUNSCH
Massachusetts Institute of Technology, and Harvard University, Cambridge, Massachusetts

RAFFAELE FERRARI
Massachusetts Institute of Technology, Cambridge, Massachusetts

ABSTRACT

The central change in understanding of the ocean circulation during the past 100 years has been its emergence as an intensely time-dependent, effectively turbulent and wave-dominated, flow. Early technologies for making the difficult observations were adequate only to depict large-scale, quasi-steady flows. With the electronic revolution of the past 50 years, the emergence of geophysical fluid dynamics, the strongly inhomogeneous time-dependent nature of oceanic circulation physics finally emerged. Mesoscale (balanced), submesoscale oceanic eddies at 100-km horizontal scales and shorter, and internal waves are now known to be central to much of the behavior of the system. Ocean circulation is now recognized to involve both eddies and larger-scale flows with dominant elements and their interactions varying among the classical gyres, the boundary current regions, the Southern Ocean, and the tropics.

1. Introduction

In the past 100 years, understanding of the general circulation of the ocean has shifted from treating it as an essentially laminar, steady-state, slow, almost geological, flow, to that of a perpetually changing fluid, best characterized as intensely turbulent with kinetic energy dominated by time-varying flows. The space scales of such changes are now known to run the gamut from 1 mm (scale at which energy dissipation takes place) to the global scale of the diameter of Earth, where the ocean is a key element of the climate system. The turbulence is a mixture of classical three-dimensional turbulence, turbulence heavily influenced by Earth rotation and stratification, and a complex summation of random waves on many time and space scales. Stratification arises from temperature and salinity distributions under high pressures and with intricate geographical boundaries and topography. The fluid is incessantly subject to forced fluctuations from exchanges of properties with the turbulent atmosphere.

Although both the ocean and atmosphere can be and are regarded as global-scale fluids, demonstrating analogous physical regimes, understanding of the ocean until relatively recently greatly lagged that of the atmosphere. As in almost all of fluid dynamics, progress in understanding has required an intimate partnership between theoretical description and observational or laboratory tests. The basic feature of the fluid dynamics of the ocean, as opposed to that of the atmosphere, has been the very great obstacles to adequate observations of the former. In contrast with the atmosphere, the ocean is nearly opaque to electromagnetic radiation, the accessible (by ships) surface is in constant and sometimes catastrophic motion, the formal memory of past states extends to thousands of years, and the analog of weather systems are about 10% the size of those in the atmosphere, yet evolve more than an order of magnitude more slowly. The overall result has been that as observational technology evolved, so did the theoretical understanding. Only in recent years, with the advent of major advances in ocean observing technologies, has physical/dynamical oceanography ceased to be a junior partner to dynamical meteorology. Significant physical regime differences include, but are not limited to, 1) meridional continental boundaries that block the otherwise dominant zonal...
flows, 2) predominant heating at the surface rather than at the bottom, 3) the much larger density of seawater (a factor of $10^3$) and much smaller thermal expansion coefficients (a factor of less than $1/10$), and 4) overall stable stratification in the ocean. These are the primary dynamical differences; many other physical differences exist too: radiation processes and moist convection have great influence on the atmosphere, and the atmosphere has no immediate analog of the role of salt in the oceans.

What follows is meant primarily as a sketch of the major elements in the evolving understanding of the general circulation of the ocean over the past 100+ years. Given the diversity of elements making up understanding of the circulation, including almost everything in the wider field of physical oceanography, readers inevitably will find much to differ with in terms of inclusions, exclusions, and interpretation. An anglophone bias definitely exists. We only touch on the progress, with the rise of the computer, in numerical representation of the ocean, as it is a subject in its own right and is not unique to physical oceanography. All science has been revolutionized.

That the chapter may be both at least partially illuminating and celebratory of how much progress has been made is our goal. In particular, our main themes concern the evolution of observational capabilities and the understanding to which they gave rise. Until comparatively recently, it was the difficulty of observing and understanding a global ocean that dominated the subject.\(^1\)

### 2. Observations and explanations before 1945

Any coherent history of physical oceanography must begin not in 1919 but in the nineteenth century, as it sets the stage for everything that followed. A complete history would begin with the earliest seafarers [see, e.g., Cartwright (2001) who described tidal science beginning in 500 BCE, Warren (1966) on early Arab knowledge of the behavior of the Somali Current, and Peterson et al. (1996) for an overview] and would extend through the rise of modern science with Galileo, Newton, Halley, and many others. Before the nineteenth century, however, oceanography remained largely a cartographic exercise. Figure 7-1 depicts the surface currents, as inferred from ships' logs, with the Franklin–Folger Gulf Stream shown prominently on the west. Any navigator, from the earliest prehistoric days, would have been very interested in such products. Emergence of a true science had to await the formulation of the Euler and Navier–Stokes equations in the eighteenth and nineteenth centuries. Not until 1948 did Stommel point out that the intense western intensification of currents, manifested on the U.S. East Coast as the Gulf Stream, was a fluid-dynamical phenomenon in need of explanation.

Deacon (1971) is a professional historian’s discussion of marine sciences before 1900. Mills (2009) brings the story of general circulation oceanography to about 1960. In the middle of the nineteenth century, the most basic problem facing anyone making measurements of the ocean was navigation: Where was the measurement obtained? A second serious issue lay with determining how deep the ocean was and how it varied with position. Navigation was almost wholly based upon celestial methods and the ability to make observations of sun, moon, and stars, along with the highly variable skill of the observer, including the ability to carry out the complex reduction of such measurements to a useful position. Unsurprisingly, particularly at times and places of constant cloud cover and unknown strong currents, reported positions could be many hundreds of kilometers from the correct values. One consequence could be shipwreck.\(^2\) Water depths were only known from the rare places where a ship could pause for many hours to lower a heavy weight to the seafloor. Observers then had to compute the difference between the length of stretching rope spooled out when the bottom was hit (if detected), and the actual depth. An example of nineteenth-century North Atlantic Ocean water-depth estimates can be seen in Fig. 7-2. A real solution was not found until the invention of acoustic echo sounding in the post–World War I era.

Modern physical oceanography is usually traced to the British Challenger Expedition of 1873–75 in a much-told tale (e.g., Deacon 1971) that produced the first global-scale sketches of the distributions of temperature and salinity [for a modern analysis of their temperature data, see Roemmich et al. (2012)].

\(^1\) Physical oceanography, as a coherent science in the nineteenth century existed mainly in support of biological problems. A purely physical oceanographic society has never existed—most professional oceanographic organizations are inevitably dominated in numbers by the biological ocean community. In contrast, the American Meteorological Society (AMS) has sensibly avoided any responsibility for, for example, ornithology or entomology, aircraft design, or tectonics—the atmospheric analogs of biological oceanography, ocean engineering, or geology. This otherwise unmanageable field may explain why ocean physics eventually found a welcoming foster home with the AMS with the establishment of the Journal of Physical Oceanography in 1971.

\(^2\) Stommel (1984) describes the many nonexistent islands that appeared in navigational charts, often because of position errors. Reports of land could also be a spurious result of observing mirages and other optical phenomena.
One of the most remarkable achievements by the late-nineteenth-century oceanographers was the development of a purely mechanical system (nothing electrical) that permitted scientists on ships to measure profiles of temperature $T$ at depth with precisions of order $0.01^\circ$C and salinity content $S$ to an accuracy of about $0.05$ g kg$^{-1}$ (Helland-Hansen and Nansen 1909, p. 27), with known depth uncertainties of a few meters over the entire water column of mean depth of about 4000 m. This remarkable instrument system, based ultimately on the reversing thermometer, the Nansen bottle, and titration chemistry, permitted the delineation of the basic three-dimensional temperature and salt distributions of the ocean. As the only way to make such measurements required going to individual locations and spending hours or days with expensive ships, global exploration took many decades. Figures 7-2 and 7-3 display the coverage that reached to at least 2000 and 3600 m over the decades, noting that the average ocean depth is close to 4000 m. [The sampling issues, including seasonal aliasing, are discussed in Wunsch (2016).] By good fortune, the large-scale structures below the very surface of $T$ and $S$ appeared to undergo only small changes on time scales of decades and spatial scales of thousands of kilometers, with “noise” superimposed at smaller scales. Measurements led to the beautiful hand-drawn property sections and charts that were the central descriptive tool.
Mechanical systems were also developed to measure currents. Ekman's current meter, one lowered from a ship and used for decades, was a purely mechanical device, with a particularly interesting method for recording flow direction (see Sandström and Helland-Hansen 1905). Velocity measurements proved much more challenging to interpret than hydrographic ones, because the flow field is dominated by rapidly changing

FIG. 7-2. Known depths in the North Atlantic, from Maury (1855). Lack of knowledge of water depths became a major issue with the laying of the original undersea telegraph cables (e.g., Dibner 1964). Note the hint of a Mid-Atlantic Ridge. Maury also shows a topographic cross section labeled “Fig. A” in this plot.
small-scale flows and not by stable large-scale currents. Various existing reviews permit us to provide only a sketchy overview; for more details of observational history, see particularly the chapters by Warren, Reid, and Baker in Warren and Wunsch (1981), Warren (2006), the books by Sverdrup et al. (1942) and Defant (1961), and chapter 1 of Stommel (1965).

The most basic feature found almost everywhere was a combined permanent “thermocline”/“halocline,” a depth range typically within about 800 m of the surface.
over which, in a distance of several hundred meters, both the temperature and salinity changed rapidly with depth. It was also recognized that the abyssal ocean was very cold, so cold that the water could only have come from the surface near the polar regions (Warren 1981).

The most important early advance in ocean physics was derived directly from meteorology—the development of the notion of “geostrophy” (a quasi-steady balance between pressure and Coriolis accelerations) from the Bergen school. Bjerkenes’s circulation theorem as simplified by Helland–Hansen for steady circulations (see Vallis 2017) was recognized as applicable to the ocean. To this day, physical oceanographers refer to the “thermal wind” when using temperature and salinity to compute density and pressure, and hence the geostrophically (and hydrostatically) balanced flow field. Even with this powerful idea, understanding of the perhaps steady ocean circulation lagged behind that of the atmosphere as oceanographers confronted an additional complication that does not exist in meteorology. By assuming hydrostatic and thermal wind balance, the horizontal geostrophic velocity field components—call them \( u_g \) and \( v_g \) (east and north)—can be constructed from measurements of the water density \( \rho(T, S, p) \) (where \( p \) is hydrostatic pressure) on the basis of the relationship

\[
\frac{\partial(\rho u_g)}{\partial z} = g \frac{\partial \rho}{\partial y} \quad \text{and} \quad \frac{\partial(\rho v_g)}{\partial z} = -g \frac{\partial \rho}{\partial x} \tag{7-1a}
\]

\[
\frac{\partial(\rho u_g)}{\partial z} = g \frac{\partial \rho}{\partial y} \quad \text{and} \quad \frac{\partial(\rho v_g)}{\partial z} = -g \frac{\partial \rho}{\partial x} \tag{7-1b}
\]

where \( x, y, \) and \( z \) are used to represent local Cartesian coordinates on a sphere, \( f = 2\pi \cos \phi \) is the Coriolis parameter as a function of latitude \( \phi \), and \( g \) is the local gravity. Either of these equations [e.g., Eq. (7-1a)], can be integrated in the vertical direction (in practice as a sum over finite differences and with approximations related to density \( \rho \)):

\[
\rho u_g(x, y, z) = \int_{z_0}^z g \frac{\partial \rho}{\partial y} \, dz + \rho u_0(x, y, z_0), \tag{7-2}
\]

and in a similar way for \( v_g(x, y, z) \). The starting depth of the integration \( z_0 \) is arbitrary and can even be the sea surface; \( u_0 \) is thus simply the horizontal velocity at \( z_0 \). These equations constitute the “dynamic method” and were in practical oceanographic use as early as Helland-Hansen and Nansen (1909, p. 155). The constant of integration \( u_0 \), as is the equivalent \( v_0 \), is missing. In the atmosphere, the surface pressure is known and thus the \( u_0 \) and \( v_0 \) can be estimated using geostrophic balance. Various hypotheses were proposed for finding a depth \( z_0 \) at which \( u_0 \) and \( v_0 \) could be assumed to vanish (a “level of no motion”). It is an unhappy fact that none of the hypotheses proved demonstrable, and thus oceanic flows were only known relative to unknown velocities at arbitrary depths. Estimated transports of fluid and their important properties such as heat could easily be dominated by even comparatively small, correct, values of \( u_0 \) and \( v_0 \). Physical oceanography was plagued by this seemingly trivial issue for about 70 years. It was only solved in recent years through mathematical inverse methods and by technologies such as accurate satellite altimetry and related work, taken up later.

The earliest dynamical theory of direct applicability to the ocean is probably Laplace’s (1775) discussion of tidally forced motions in a rotating spherical geometry, using what today we would call either the Laplace tidal or the shallow-water equations. Laplace’s equations and many of their known solutions are thoroughly described in Lamb (1932) and tidal theory per se will not be pursued here (see Cartwright 1999; Wunsch 2015, chapters 6 and 7). Those same equations were exploited many years later in the remarkable solutions of Hough (1897, 1898), and by Longuet-Higgins (1964) in his own and many others’ following papers. As with most of the theoretical ideas prior to the midtwentieth century, they would come to prominence and be appreciated only far in the future. [The important ongoing developments in fluid dynamics as a whole are described by Darrigol (2005).]

Probably the first recognizable dynamical oceanographic theory arises with the celebrated paper of Ekman (1905). In another famous story (see any textbook) Ekman produced an explanation of Fridtjof Nansen’s observation that free-floating icebergs tended to move at about 45° to the right of the wind (in the Northern Hemisphere). His solution, the Ekman layer, remains a cornerstone of oceanographic theory. [See Faller (2006) for discussion of its relationship to Prandtl’s contemporaneous ideas about the boundary layer.]

Ekman and others devoted much time and attention to developing instruments capable of making direct measurements of oceanic flow fields with depth. Much of the justification was the need to determine the missing integration constants \( u_0 \) and \( v_0 \) of the dynamic method.

---

3 We use “physics” in the conventional sense of encompassing both dynamics and all physical properties influencing the fluid ocean.

4 According to Gill (1982), the first use of the terminology was in 1916 by Napier Shaw—the expression does not appear at all in Sverdrup et al. (1942). The notion of geostrophic balance, however, appears earlier in the oceanographic literature through the work of Sandstroem and Helland-Hansen (1905) as inspired by the new dynamical approach to meteorology and oceanography introduced in the Bergen school by Bjerkenes (Mills 2009).
These instruments were lowered on cables from a ship. Unfortunately, ships could stay in the same place for only comparatively short times (typically hours) owing to the great costs of ship time, and with navigational accuracy being wholly dependent upon sun and star sights. Early on, it was recognized that such measurements were extremely noisy, both because of ship movements but also because of the possible existence of rapidly fluctuating internal waves, which was already apparent (see Nansen 1902), that would contaminate the measurements of slowly evolving geostrophic velocities $u_g$ and $v_g$.

Absent any method for direct determination of water motions over extended periods of time, and the impossibility of obtaining time series of any variable below the surface, theory tended to languish. The most notable exceptions were the remarkable measurements of Pillsbury (1891) in the Straits of Florida who managed to keep anchored ships in the Gulf Stream for months at a time. Stommel (1965) has a readable discussion of Pillsbury’s and other early measurements. These data, including the direct velocities, were used by Wüst (1924) to demonstrate the applicability of the thermal wind/dynamic method. Warren (2006), describing Wüst’s methods, shows that the result was more ingenious than convincing.

Of the theoretical constructs that appeared prior to the end of the World War II (WWII), the most useful ones were the development of internal wave theory by Stokes (1847), Rayleigh (1883), and Fjeldstad (1933) among others and its application to the two-layer “dead-water” problem by Ekman (1906). Building on the work of Hough and others, the English mathematician Goldsbrough (1933) solved the Laplace tidal equations on a sphere, when subjected to mass sources, a development only exploited much later for the general circulation by Stommel (1957) and then Huang and Schmitt (1993). One might also include Rossby’s “wake stream” theory of the Gulf Stream as a jet, although that idea has had little subsequent impact. The use of three-dimensional eddy diffusivities (“Austausch” coefficients), as employed in one dimension by Ekman, acting similarly to molecular diffusion and friction but with far-larger values, was the focus of a number of efforts summarized by Defant (1961), following the more general discussions of fluid turbulence.5

In the nineteenth century, controversy resulted over the question of whether the ocean was primarily wind driven or thermally forced—a slightly irrational, noisy, dispute that is typical of sciences with insufficient data (Croll 1875; Carpenter 1875). Sandström (1908; see the English translation in Kuhlbrodt 2008) showed that convection in fluids where the heating lay above or at the same surface (as in the ocean) would be very weak relative to fluids heated below the level of cooling (the atmosphere). Bjerknes et al. (1933) labeled Sandström’s arguments as a “theorem” and thus attracted to it some considerable later misinterpretation. Jefreys (1925), in an influential paper, had argued that Sandström’s inferences (“‘principles’”) had little or no application to the atmosphere but were likely relevant to the oceans. There the matter rested for 50+ years.6

The highly useful summary volume by Sverdrup et al. (1942) appeared in the midst of WWII. It remained the summary of the state of all oceanography, and not just the physical part, for several decades. Emphasis was given to water-mass volumes (basically varying temperatures and salinities), the dynamic method, and local (Cartesian coordinate) solutions to the shallow-water equations. The Ekman layer is the only recognizable element of “dynamical oceanography” relevant to the general circulation. In his condensed version directed specifically to meteorologists (Sverdrup 1942), Sverdrup concluded the monograph with the words: “It is not yet possible to deal with the system atmosphere-ocean as one unit, but it is obvious that, in treating separately the circulation of the atmosphere, a thorough consideration of the interaction between the atmosphere and the oceans is necessary” (p. 235), a statement that accurately defines much of the activity today in both atmospheric and oceanic sciences.

3. Post-WWII developments and the emergence of GFD

An informal sense of the activities in physical oceanography in WW II and the period immediately following, with a focus on the United Kingdom, can be found in Laughton et al. (2010). Shor (1978) is another history, focused on Scripps Institution of Oceanography, and Cullen (2005) described the Woods Hole Oceanographic Institution. Mills (2009) covered the early-twentieth-century evolution specifically of dynamical oceanography in Scandinavia, France, Canada, and Germany. Other national quasi histories probably exist for

5 Welander (1985) noted that Ekman in 1923 could have produced the Sverdrup/Stommel results decades earlier than those authors did, having written down one form of Stommel’s equation. He speculated that, among other reasons, it was Ekman’s dislike of “approximations” that restrained him—a perhaps fatal problem for a fluid dynamicist.

6 The controversy reemerged in recent years under the guise of the physics of “horizontal convection,” its relevance to the study of the “thermohaline” circulation, and the still-opaque study of oceanic energetics in general.
other countries, including the Soviet Union, but these are not known to us.

A simple way to gain some insight into the intellectual flavor of physical oceanography in the interval from approximately 1945 to 1990 is to skim the papers and explanatory essays in the collected Stommel papers (Hogg and Huang 1995). The more recent period, with a U.S. focus, is covered in Jochum and Murtugudde (2006). The edited volume by Warren and Wunsch (1981) gives a broad overview of how physical oceanography stood as of approximately 1980—reflecting the first fruits of the electronics revolution.

The advent of radar and its navigational offspring such as loran and Decca greatly improved the navigational uncertainties, at least in those regions with good coverage (North Atlantic Ocean). This period also saw the launch of the first primitive navigational satellites (U.S. Navy Transit system), which gave a foretaste of what was to come later.

Because of the known analogies between the equations thought to govern the dynamics of the atmosphere and ocean, a significant amount of the investigation of theoretical physical oceanographic problems was carried out by atmospheric scientists (e.g., C.-G. Rossby, J. G. Charney, and N. A. Phillips) who were fascinated by the oceans. The field of geophysical fluid dynamics (GFD) emerged, based initially on oceanic and atmospheric flows dominated by Earth’s rotation and variations of the fluid density (see Fig. 7-4). Present-day GFD textbooks (e.g., Pedlosky 1987; McWilliams 2006; Cushman-Roisin and Beckers 2011; Vallis 2017) treat the two fluids in parallel. When it came to observations, however, Gill’s (1982) textbook was and is a rare example of an attempt to combine both the theory and

![Fig. 7-4. As in Fig. 7-3, but for stations reaching at least 3600 m by decade. The challenge of calculating any estimate of the heat content (mean temperature) or salinity in early decades will be apparent.](image-url)
observations of atmosphere and ocean in a single treatment. Although a chapter describes the atmospheric general circulation, however, he sensibly omitted the corresponding chapter on the ocean general circulation. GFD might be defined as the reduction of complex geophysical fluid problems to their fundamental elements, for understanding, above realism. The emergence of potential vorticity (a quasi-conserved quantity derived from the oceanic vorticity and stratification), in various approximations, as a fundamental unifying dynamical principle emerged at this time (see Stommel 1965, chapter 8). Vallis (2016) has written more generally about GFD and its applications.

a. Steady circulations

In the United States and United Kingdom at least, WWII brought a number of mathematically adept professionally trained scientists into close contact with the problems of the fluid ocean. Before that time, and apart from many of the people noted above, physical oceanography was largely in the hands of people (all men) who can reasonably be classified as “natural philosophers” in the older tradition of focusing on description, rather than physics. (In English, the very name “oceanography”—from the French—evokes the descriptive field “geography” rather than the explicitly suppressed “oceanology” as a parallel to “geology.”) Seagoing physical oceanography had, until then, been primarily a supporting science for biological studies. The start of true dynamical oceanography was provided in two papers (Sverdrup 1947; Stommel 1948) neither of whom would have been regarded as fluid dynamics experts. But those two papers marked the rise of GFD and the acceleration of dynamical oceanography. Sverdrup derived a theoretical relationship between the wind torque acting at the ocean surface and the vertically integrated meridional (north–south) transport of upper ocean waters. Stommel’s (1948) paper treated a linear, homogeneous flat-bottom ocean, but succeeded in isolating the meridional derivative of the Coriolis acceleration as the essential element in producing western boundary currents like the Gulf Stream in the North Atlantic—a prototype of GFD reductionism.

Closely following on the Sverdrup/Stommel papers were those of Munk (1950), who effectively combined the Sverdrup and Stommel solutions, Munk and Carrier (1950), Charney (1955), Morgan (1956), and a host of others.8 Following the lead of Munk and Carrier (1950) the Gulf Stream was explicitly recognized as a form of boundary layer, and the mathematics of singular perturbation theory was then enthusiastically applied to many idealized versions of the general circulation (Robinson 1970). Stommel with his partner, Arnold Arons, developed the so-called Stommel–Arons picture of the abyssal ocean circulation (Stommel 1957)—probably the first serious attempt at the physics of the circulation below the directly wind-driven regions near surface. A few years later, Munk (1966) produced his “abyssal recipes” paper that, along with the Stommel–Arons schematic, provided the framework for the next several decades of the understanding the deep ocean circulation, thought of as dynamically relatively spatially uniform. This subject will be revisited below.

Attempts at a theory of the thermocline that would predict the stratification and baroclinic flows forced by the surface winds started with linear perturbation methods (Stommel 1957; cf. Barcilon and Pedlosky 1967). But because the goal was explaining the basic oceanic stratification, rather than assuming it as part of the background state, the problem resulted in highly nonlinear equations (e.g., Needler 1967). Ingenious solutions to these equations were found by Robinson and Stommel (1959) and Welander (1959) using analytic similarity forms. These solutions looked sufficiently realistic (Fig. 7-5) to suggest that the basic physics had been appropriately determined. [See the textbooks by Pedlosky (1996); Vallis (2017); Huang (2010); Olbers et al. (2012).] Large-scale solutions that assumed vertical mixing of temperature and salinity in the upper ocean was a leading-order process (e.g., Robinson and Stommel 1959) were so similar to those that ignored mixing altogether (e.g., Welander 1959) that the immediate hope of deducing a vertical eddy diffusivity $K_v$ from hydrographic measurements alone proved unavailing. The puzzle ultimately led to a decades-long effort, much of it driven by C. S. Cox and continuing today, to measure $K_v$ directly (see Gregg 1991) and to its inference from a variety of chemical tracer observations.

b. Observations

Until about 1990, the chief observational tool for understanding the large-scale ocean circulation remained the shipboard measurement of hydrographic properties.

---

7 Not coincidentally, the Geophysical Fluid Dynamics Program arose soon afterward at the Woods Hole Oceanographic Institution. This program, continuing more than 50 years later, has provided a focus, and a kind of welcoming club, for anyone interested in what is now known worldwide as “GFD.”

8 A rigorous derivation of the Laplace tidal equations did not appear until the work of Miles (1974) who showed that they could only be justified if the fluid was actually stratified—even though they are used to describe unstratified (homogeneous) fluid flows.
leading to the calculation of density and the use of the
dynamic method, often still employing assumed levels
of no motion. Even as the technology evolved (Baker
1981) from reversing thermometers and Nansen bottles
to the salinity–temperature–depth (STD), and conductivity–
temperature–depth (CTD) devices, and from mechanical
bathythermographs (MBTs) to expendable BTs (XBTs),
the fundamental nature of the subject did not change. The
major field program in this interval was the International
Geophysical Year (IGY), July 1957–December 1958. The
IGY Atlantic surveys were modeled on the R/V Meteor
Atlantic survey of the 1920s (Wüst and Defant 1936). Nota-
ble, the Atlantic Ocean atlases of Fuglister (1960) and
Worthington and Wright (1970) were based on these cruises
and emerged as the basic representation of the ocean
circulation.

Apart from the Atlantic Ocean, hydrographic surveys
to the bottom remained extremely rare, with the so-
called Scorpio sections in the mid-1960s in the South
Pacific Ocean (Stommel et al. 1973), the R/V Eltanin
survey of the Southern Ocean (Gordon and Molinelli
1982), and an isolated trans–Indian Ocean section
(Wyrtki et al. 1971) being late exceptions. This rarity
reflected a combination of the great difficulty and ex-
 pense of measurements below about 1000 m, coupled
with the very convenient supposition that the deep

---

**Fig. 7-5.** From Robinson and Stommel (1959, their Fig. 2) showing (top) a rendering of a solution to the ther-
mocline equations in comparison with (bottom) a sketch of the observed isotherms (°C). (© 1959 Allan Robinson
and Henry Stommel, published by Taylor and Francis Group; https://creativecommons.org/licenses/by/4.0/.)
The ocean was simple and boring (H. Stommel circa 1965, personal communication to C. Wunsch).

c. High latitudes

During this long period, observations were focused on the mid- to lower latitudes, with the difficult-to-reach Southern Ocean remaining comparatively poorly observed. Theoretical work was directed at the dynamics of the Antarctic Circumpolar Current (ACC). The absence of continuous meridional barriers in the latitude range of Drake Passage did not allow the development of the western boundary currents that were crucial in the theories of Stommel and Munk. Stommel (1957) argued that the Scotia Island Arc could act as a porous meridional barrier permitting the ACC to pass though, but be deflected north to join the meridional Falkland Current along the South American continent. Gill (1968) pointed out that the zonal ACC current could also result from a balance between the surface wind stress and bottom friction, without any need of meridional boundaries. However, he considered only models with a flat bottom that produced transports far in excess of observations for any reasonable value of bottom drag coefficients. Surprisingly both theories ignored Munk and Palmén’s (1951) work, which had identified topographic form drag (the pressure forces associated with obstacles) from ocean ridges and seamounts as a key mechanism to slow down the ACC and connect it to currents to the north. Development of a theory of the Southern Ocean circulation is taken up below. The ice-covered Arctic Sea was essentially unknown.

d. Tropical oceanography

Tropical oceanography was largely undeveloped until attention was directed to it by the rediscovery of the Pacific (and Atlantic) equatorial undercurrents. Buchanan (1888) had noted that buoys drogued at depth moved rapidly eastward on the equator in the Atlantic, but his results were generally forgotten (Stroup and Montgomery 1963). Theories of the steady undercurrent were almost immediately forthcoming (see Fig. 7-6) with perhaps the most important result being their extremely sensitive dependence on the vertical eddy diffusivity \( K_v \) (e.g., Charney and Spiegel 1971; Philander 1973). But the real impetus came with the recognition (see Wyrski 1975a; Halpern 1996) that El Niño, known from colonial times as a powerful, strange, occasional, event in the eastern tropical Pacific and regions of Ecuador and Peru, was in fact a phenomenon both global in scope and involving the intense interaction of atmosphere and ocean. Such physics could not be treated as a steady state.

e. Time-dependent circulation

Recognition of a very strong time dependence in the ocean dates back at least to Helland-Hansen and Nansen (1909) and is already implicit in Maury (1855). Fragmentary indications had come from the new Swallow floats (Crease 1962; Phillips 1966) and the brief direct current-meter measurements from ships had shown

---

9 Whether the Arctic is a sea or an ocean is not universally agreed on. Sverdrup et al. (1942) called it the “Arctic Mediterranean Sea,” both in acknowledgment of its being surrounded by land and because of its small size.
variability from the longest down to the shortest measurable time scales. Physical oceanographers in contact with the meteorological community were acutely aware of Starr’s (1948, 1968) demonstration that atmospheric “eddies” to a large extent controlled the larger-scale flow fields, rather than being a passive dissipation mechanism—in the sense of the Austausch coefficients of much theory. But because observational capabilities were still extremely limited, most of the contributions in the immediate postwar period tended to be primarily theoretical ones. Rossby et al. (1939) had produced a mathematical formulation of what came to be known as the “Rossby wave” and in Rossby (1945) he had made explicit its hypothetical application to the ocean. As Platzman (1968) describes in detail, the physics of those waves had been known for a long time in the work of Hough (1897, 1898)—who called them “tidal motions of the second class”—Rossby’s analysis produced the simplest possible waves dependent upon the variation of the Coriolis parameter, and the label has stuck. In a series of papers starting in 1964, Longuet-Higgins extended Hough’s analysis on the sphere and showed clearly the relationship to the approximations based upon Rossby’s beta plane. Many of the papers in Warren and Wunsch (1981) provided a more extended account of this period. Difficulties with observations vis-à-vis the emerging theories had led Stommel (see Hogg and Huang 1995, Vol. 1, p. 124) to famously assert that the theories “had a peculiar dreamlike quality.”

f. The level of no motion

The issue of the missing constant of integration when computing the thermal wind had attracted much attention over many decades, frustrating numerous oceanographers who were trying to calculate absolute flow rates. Although a number of methods had been proposed over the years [see the summary in Wunsch (1996)], none of them proved satisfactory. To a great extent, the steady ocean circulation was inferred by simply assuming that, at some depth or on some isopycnal or isotherm, both horizontal velocities, $u$ and $v$, vanished, implying $u_0$ and $v_0 = 0$ there. Choice of such a “level of no horizontal motion” $z_0(x, y)$, although arbitrary, did give qualitatively stable results, as long as a sufficiently deep value of $z_0$ was used; temporal stability was rarely ever tested. This apparent insensitivity of results (see Figs. 7-7 and 7-8) is understandable on the assumption that the magnitude of the horizontal flows diminished with depth—an inference in turn resting upon the hypotheses that flows were dominantly wind driven.

For quantitative use, however, for example in computing the meridional transport of heat or oxygen by the ocean as we mentioned above, differing choices of $z_0$ could lead to large differences. Ultimately Worthington (1976), in trying to balance the steady-state mass, temperature, salinity, and oxygen budgets of the North Atlantic Ocean, had come to the radical, and indefensible, conclusion that large parts of the circulation could not be geostrophically balanced by pressure gradients. (The inference was indefensible because no other term in the equations of motion is large enough to balance the inferred Coriolis force and Newton’s Laws are then violated.)

The problem was eventually solved in two, initially different-appearing ways: through the methods of inverse theory (Wunsch 1977) and the introduction of the β spiral (Stommel and Schott 1977). These methods and their subsequent developments employed explicit conservation rules that are not normally part of the dynamic method (heat, salt, volume, potential vorticity, etc.). Wunsch (1996) summarizes the methods—including Needler’s (1985) formal demonstration that, with perfect data in a steady state, the three components of steady velocity ($u$, $v$, and $w$) were fully determined by the three-dimensional density field. None of the methods was practical prior to the appearance of digital computers.

Ironically, the solution to the major weakness of the dynamic method emerged almost simultaneously with the understanding that the ocean was intensely time dependent: the meaning of the statically balanced ocean calculations was thus unclear. When accurate satellite altimetry and accurate geoids became available after 1992, it was possible to obtain useful direct measurements of the absolute pressure of the sea surface elevation [Fig. 7-8 and see Fu et al. (2019)]. Both the inverse methods and the absolute measurements showed that a level of no motion did not exist. That deep velocities are weaker generally than those near the surface is, however, generally correct (Reid 1961).

4. Steady-state circulations circa 1980+

The physics and mathematical challenges of deducing the nature of a hypothetical, laminar steady-state ocean continue to intrigue many GFD theoreticians and modellers. The most important of such theories was instigated by Luyten et al. (1983) who, backing away from the continuous ocean represented in very complicated equations, reduced the problem to one of a finite number of layers (typically 2–3). Following Welander’s (1959) model, the theory ignored mixing between layers and assumed that temperature, salinity, and potential vorticity were set at the surface in each density layer. This theory of the “ventilated thermocline” of the upper ocean, combined also with ideas about the effects of
eddies (Rhines and Young 1982), led to a renaissance in the theory. In the theories, the upper ocean is divided into a large region that is directly ventilated by the atmosphere and two or more special regions (the “shadow zone” and the unventilated “pool”). These theoretical ideas are well covered in the textbooks already noted and are not further discussed here except to mention that the theory has since been extended to connect it to the rest of the ocean interior (which requires addition of mixing at the base of the ventilated thermocline; Samelson and Vallis 1997) and to the tropical oceans (which alleviates the need of any mixing to explain the equatorial currents; Pedlosky 1996). Determining the extent to which these theories describe the upper ocean in the presence of intense time variability is a major subject of current activity in both theory and observation.

Theories for the deep ocean circulation lagged behind. Starting with an influential paper by Stommel (1961) that introduced a two-box model to describe the deep circulation as resulting from the density difference between the low- and high-latitude boxes, the idea gained ground that the deep circulation was driven by the density differences generated by heating and cooling/evaporation and precipitation at high latitudes, in contrast to the wind-driven circulation in the upper thermoclines. This deep “thermohaline circulation,” as it came to be called, consisted of waters sinking into the abyss in the North Atlantic and around Antarctica and rising back to the surface more or less uniformly in the rest of the ocean. Van Aken (2007) provides a good review of the theoretical progress until the end of the twentieth century. Beyond the Stommel–Arons model to describe the depth-integrated deep circulation, theory focused on the overturning circulation and the associated cross-equatorial heat transport because of its relevance for climate. The approach was much less formal than in theories of the upper ocean and relied largely on...
box models and simple scaling arguments. Indeed the most influential description of the supposed thermohaline circulation up to this time was the cartoon simplifications drawn by Gordon (1986) and Broecker (1987). These and other discussions led in turn to a heavy emphasis on the North Atlantic Ocean and its overturning in the guise of the Atlantic meridional overturning circulation (AMOC), but whose role in the climate state remains only a portion of the global story. As described below, it is inseparable from the mechanically driven circulations.

A theory for the deep circulation more grounded in basic GFD has only started to emerge in the last twenty years, after the crucial role of the Southern Ocean in the global overturning circulation was fully appreciated. We review the emergence of this paradigm in the section on the Southern Ocean below. Here suffice it to say that the role of the Southern Hemisphere westerlies took central stage in the theory of the deep overturning circulation—rendering obsolete the very concept of a purely thermohaline circulation. The deep ocean is as sensitive to the winds as the upper thermoclines, and both circulations are strongly affected by the distinct patterns of heating and cooling, and of evaporation and precipitation.

5. Era of the time-dependent ocean

The most important advance in physical oceanography in the last 50 years, as with so many other fields, was the invention of the low-power integrated circuit, making possible both the remarkable capability of today’s observational instruments, and the computers necessary to analyze and model the resulting data. This revolution began to be apparent in the early 1970s as the purely
mechanical systems such as the Nansen bottle/reversing thermometer, the bathythermograph, Ekman current meter, and so on gave way to their electronic counterparts (see e.g., Baker 1981; Heinmiller 1983) and with the parallel capabilities of spaceborne instrumentation [e.g., Martin (2014) and Fu et al. (2019)]. True time series, both Eulerian and near-Lagrangian (employing floats), became available for the first time, spanning months and then years—capabilities unimaginable with shipborne instruments. Equally important was the revolution in navigational accuracy that built on the development of radar, loran, and other radiometric methods during WWII. The present culmination is the global positioning system (GPS). Today, a push button on a cellular phone or equivalent produces, with zero user skill, much higher accuracies than the ingenious, painstaking, methods of celestial navigation that required years of experience to use.

Much of the ocean-bottom topography has been described, but many details remain uncertain (e.g., Wessel et al. 2010). Very small-scale topography, presumed to be of importance in oceanic boundary layer physics, remains unknown, and determinable at present only with the limited and expensive shipboard multibeam surveys (see Fig. 7-9).

As new instrumentation gradually evolved from the 1970s onward (self-contained moored instruments operating successfully for months and years, neutrally buoyant floats tracked in a variety of ways, rapid chemical analysis methods, sea surface temperature pictures from the new satellites, etc.), the attention of much of the seagoing and theoretical communities turned toward the problems of understanding the newly available, if fragmentary, time series. In the background was the knowledge of much of the community of the importance of large-scale meteorological patterns known as weather, and in particular the book by Starr (1968) and the preceding papers. Some of Starr’s students (e.g., Webster 1961) had already tried employing limited ocean data in meteorological analogies.

In what became known as the International Decade of Ocean Exploration (IDOE; see Lambert 2000), largely funded in the United States by the National Science Foundation and the Office of Naval Research, much of the oceanographic community focused for the first time on documenting the time variability in the hope of understanding those elements of the ocean that were not in steady state.

A convenient breakdown can be obtained from the various physically oriented IDOE elements: the Mid-Ocean Dynamics Experiment (MODE), POLYGON+MODE (POLYMODE; the U.S.–Soviet follow-on to “POLYGON” and MODE), North Pacific Experiment (NORPAX), International Southern Ocean Studies (ISOS), Climate Long-range Investigation, Mapping, and Prediction Study (CLIMAP), and Coastal Upwelling Ecosystems Analysis (CUEA). MODE (see MODE Group 1978) was an Anglo–U.S. collaboration in the western Atlantic south of Bermuda involving moored current meters, temperature–pressure recorders, bottom pressure sensors, Swallow floats, and SOFAR floats.
eventually became MODE.\textsuperscript{10} Despite some instrumental problems (the new U.S. current meters failed after approximately a month), the “experiment”\textsuperscript{11} showed beyond doubt the existence of an intense “mesoscale” eddy field involving baroclinic motions related to the baroclinic radii of about 35 km and smaller, as well as barotropic motions on a much larger scale. In oceanography, the expression mesoscale describes the spatial scale that is intermediate between the large-scale ocean circulation and the internal wave field and is thus very different from its meteorological usage (which is closer to the ocean “submesoscale”). A better descriptor is “balanced” or “geostrophic” eddies, as in the meteorological “synoptic scale.” [The reader is cautioned that an important fraction of the observed low-frequency oceanic motion is better characterized as a stochastic wave field—internal waves, Rossby waves, etc.—and is at least quasi linear, with a different physics from the vortexlike behavior of the mesoscale eddies. Most of the kinetic energy in the ocean does, however, appear to be in the balanced eddies (Ferrari and Wunsch 2009)]. Understanding whether the MODE area and its physics were typical of the ocean as a whole then became the focus of a large and still-continuing effort with in situ instruments, satellites, and numerical models.

Following MODE and a number of field programs intended to understand 1) the distribution of eddy energy in the ocean as a whole and 2) the consequences for the general circulation of eddies, a very large effort, which continues today, has been directed at the eddy field and now extending into the submesoscale (i.e., scales between 100 m and 10 km where geostrophic balance no longer holds but rotation and stratification remain important). Exploration of the global field by moorings and floats was, and still is, a slow and painful process that was made doubly difficult by the short

\textsuperscript{10} From a letter addressed to the Massachusetts Institute of Technology Lincoln Laboratory (unpublished document, 11 August 1969).

\textsuperscript{11} Physical oceanographers rarely do “experiments” [the purposeful tracer work of Ledwell et al. (1993) and later is the major exception], but the label has stuck to what are more properly called field observations.
spatial coherence scales of eddies, and the long-measuring times required to obtain a meaningful picture. The first true (nearly) global view became possible with the flight of
the high-accuracy TOPEX/Poseidon12 altimeter in 1992 and successor satellite missions. Although limited to measurements of the sea surface pressure (elevation), it became obvious that eddies exist everywhere, with an enormous range in associated kinetic energy (Fig. 7-11). The spatial variation of kinetic energy by more than two orders of magnitude presents important and interesting obstacles to simple understanding of the influence on the general circulation of the time-dependent components.

In association with the field programs, the first fine resolution (grid size <100 km) numerical models of ocean circulation were developed to examine the role of mesoscale eddies in the oceanic general circulation [see the review by Holland et al. (1983)]. Although idealized, the models confirmed that the steady solutions of the ocean circulation derived over the previous decades were hydrodynamically unstable and gave rise to a rich time-dependent eddy field. Furthermore, the eddy fields interacted actively with the mean flow, substantially affecting the time-averaged circulation.

a. Observing systems

As the somewhat unpalatable truth that the ocean was constantly changing with time became evident, and as concern about understanding of how the ocean influenced climate grew into a public problem, efforts were undertaken to develop observational systems capable of depicting the global, three-dimensional ocean circulation. The central effort, running from approximately 1992 to 1997, was the World Ocean Circulation Experiment (WOCE) producing the first global datasets, models, and supporting theory. This effort and its outcomes are described in chapters in Jochum and Murtugudde (2006), and in Siedler et al. (2001, 2013). Legacies of this program and its successors include the ongoing satellite altimetry observations, satellite scatterometry and gravity measurements, the Argo float program, and continuing ship-based hydrographic and biogeochemical data acquisition.

Having to grapple with a global turbulent fluid, with most of its kinetic energy in elements at 100-km spatial scales and smaller, radically changed the nature of observational oceanography. The subsequent cultural change in the science of physical oceanography requires its own history. We note only that the era of the autonomous seagoing chief scientist, in control of a single ship staffed by his own group and colleagues, came to be replaced in many instances by large, highly organized international groups, involvement of space and other government agencies, continual meetings, and corresponding bureaucratic overheads. As might be expected, for many in the traditional oceanographic community the changes were painful ones (sometimes expressed as “we’re becoming too much like meteorology”).

b. The turbulent ocean

A formal theory of turbulence had emerged in the 1930s from G. I. Taylor, a prominent practitioner of GFD. Taylor (1935) introduced the concept of homogeneous–isotropic turbulence (turbulence in the absence of any large-scale mean flow or confining boundaries), a concept that became the focus of most theoretical research. Kolmogorov (1941) showed that in three dimensions homogeneous–isotropic turbulence tends to transfer energy from large to small scales. [The book by Batchelor (1953) provides a review of these early results.] Subsequently, Kraichnan (1967) demonstrated that in two dimensions the opposite happens and energy is transferred to large scales. Charney (1971) realized that the strong rotation and stratification at the mesoscale acts to suppress vertical motions and thus makes ocean turbulence essentially two dimensional at those scales.

A large literature developed on both two-dimensional and mesoscale turbulence, because the inverse energy cascade raised the possibility that turbulence spontaneously generated and interacted with large-scale flows. Problematically the emphasis on homogeneous–isotropic turbulence, however, eliminated at the outset any large-scale flow and shifted the focus of turbulence research away from the oceanographically relevant question of how mesoscale turbulence affected the large-scale circulation. A theory of eddy–mean flow interactions was not developed for another 30 years, until the work of Bretherton (1969a,b) and meteorologists Eliassen and Palm (1961) and Andrews and McIntyre (1976).

The role of microscale (less than 10 m) turbulence in maintaining the deep stratification and ocean circulation was recognized in the 1960s and is reviewed below (e.g., Munk 1966.) A full appreciation of the role of geostrophic turbulence on the ocean circulation lagged behind. Even after MODE and the subsequent field programs universally found vigorous geostrophic eddies with scales on the order of 100 km, theories of the large-scale circulation largely ignored this time dependence,

12 Ocean Topography Experiment/Premier Observatoire Spatial Étude Intensive Dynamique Ocean et Nivosphere [sic], or Positioning Ocean Solid Earth Ice Dynamics Orbiting Navigator.
primarily for want of an adequate theoretical framework for its inclusion and the lack of global measurements.

That the ocean, like the atmosphere, could be unstable in baroclinic, barotropic, and mixed forms had been recognized very early. Pedlosky (1964) specifically applied much of the atmospheric theory [Charney (1947), Eady (1949), and subsequent work] to the oceanic case. Theories of the interactions between mesoscale turbulence and the large-scale circulation did not take center stage until the 1980s in theories for the midlatitude circulation (Rhines and Young 1982; Young and Rhines 1982) and the 1990s in studies of the Southern Ocean.
(Johnson and Bryden 1989; Gnanadesikan 1999; Marshall and Radko 2003).

Altimetric measurements, beginning in the 1980s, showed that ocean eddies with scales slightly larger than the first deformation radius dominate the ocean eddy kinetic energy globally (Stammer 1997) but with huge spatial inhomogeneity in levels of kinetic energy and spectral distributions (Fig. 7-12), and understanding their role became a central activity, including the rationalization of the various power laws in this figure. The volume edited by Hecht and Hasumi (2008) and Vallis (2017) review the subject to their corresponding dates.

Much of the impetus in this area was prompted by the failure of climate models to reproduce the observed circulation of the Southern Ocean. Their grids were too coarse to resolve turbulent eddies at the mesoscale. Because the effect of generation of mesoscale eddies is to flatten density surfaces without causing any mixing across density surfaces (an aspect not previously fully recognized), Gent and McWilliams (1990) proposed a simple parameterization, whose success improved the fidelity of climate models (Gent et al. 1995). It led the way to the development of theories of Southern Ocean circulation (Marshall and Speer 2012) and the overturning circulation of the ocean (Gnanadesikan 1999; Wolfe and Cessi 2010; Nikurashin and Vallis 2011).

Attention has shifted more recently to the turbulence that develops at scales below approximately 10 km—the so-called submesoscales (McWilliams 2016). Sea surface temperature maps show a rich web of filaments no more than a kilometer wide (see Fig. 7-13).

Unlike mesoscale turbulence, which is characterized by eddies in geostrophic balance, submesoscale motions become progressively less balanced as the scale diminishes—as a result of a host of ageostrophic instabilities (Boccaletti et al. 2007; Capecchi et al. 2008; Klein et al. 2008; Thomas et al. 2013). Unlike the mesoscale regime, energy is transferred to smaller scales and exchanged with internal gravity waves, thereby providing a pathway toward energy dissipation (Capet et al. 2008). Both the dynamics of submesoscale turbulence and their interaction with the internal gravity waves field are topics of current research and will likely remain the focus of much theoretical and observational investigation for at least the next few decades.

c. The vertical mixing problem

Although mesoscale eddies dominate the turbulent kinetic energy of the ocean, it was another form of turbulence that was first identified as crucial to explain the observed large-scale ocean state. Hydrographic sections showed that the ocean is stratified all the way to the bottom. Stommel and Arons (1960a,b) postulated that the stratification was maintained through diffusion of
temperature and salinity from the ocean surface. However, molecular processes were too weak to diffuse significant amounts of heat and salt.

Eckart (1948) had described how “stirring” by turbulent flows leads to enhanced “mixing” of tracers like temperature and salinity. Stirring is to be thought of as the tendency of turbulent flows to distort patches of scalar properties into long filaments and threads. Mixing, the ultimate removal of such scalars by molecular diffusion, would be greatly enhanced by the presence of stirring, because of the much-extended boundaries of patches along which molecular-scale derivatives could act effectively. [A pictorial cartoon can be seen in Fig. 7-14 (Welander 1955) for a two-dimensional flow. Three-dimensional flows, which can be very complex, tend to have a less effective horizontal stirring effect but do operate also in the vertical direction.] Munk (1966) in a much celebrated paper, argued that turbulence associated with breaking internal waves on scales of 1–100 m was the most likely candidate for driving stirring and mixing of heat and salt in the abyss—geostrophic eddies drive motions along density surfaces and therefore do not generate any diapycnal mixing.

Because true fluid mixing occurs on spatial scales that are inaccessible to numerical models, and with the understanding that the stirring-to-mixing mechanisms control the much-larger-scale circulation patterns and properties, much effort has gone into finding ways to “parameterize” the unresolved scales. Among the earliest such efforts was the employment of so-called eddy or Aus- tausch coefficients that operate mathematically like molecular diffusion but with immensely larger numerical diffusion coefficients (Defant 1961). Munk (1966) used vertical profiles of temperature and salinity and estimated that maintenance of the abyssal stratification required a vertical eddy diffusivity of $10^{-4}$ m$^2$ s$^{-1}$ (memorably 1 in the older cgs system) a value that is 1000 times as large as the molecular diffusivity of temperature and 100 000 times as large as the molecular diffusivity of salinity.

For technical reasons, early attempts at measuring the mixing generated by breaking internal waves were confined to the upper ocean and produced eddy diffusivity values that were an order of magnitude smaller than those inferred by Munk (see Gregg 1991). This led to the notion that there was a “missing mixing” problem. However, the missing mixing was found when
the technology was developed to measure mixing in the abyssal ocean—the focus of Munk’s argument [see the reviews by Wunsch and Ferrari (2004) and Waterhouse et al. (2014)]. Estimates of the rate at which internal waves are generated and dissipated in the global ocean [Munk and Wunsch (1998) and many subsequent papers] further confirmed that there is no shortage of mixing to maintain the observed stratification. The field has now moved toward estimating the spatial patterns of turbulent mixing with dedicated observations, and more sophisticated schemes are being developed to better capture the ranges of internal waves and associated mixing known to exist in the oceans. In particular, it is now widely accepted that oceanic boundary processes, including sidewalls, and topographic features of all scales and types dominate the mixing process, rather than it being a quasi-uniform open-ocean phenomenon (see Callies and Ferrari 2018).

6. ENSO and other phenomena

A history of ocean circulation science would be incomplete without mention of El Niño and the coupled atmospheric circulation known as El Niño–Southern Oscillation (ENSO). What was originally regarded as primarily an oceanic phenomenon of the eastern tropical Pacific Ocean, with implications for Ecuador–Peru rainfall, came in the 1960s (Bjerknes 1969; Wyrtki 1975b) to be recognized as both a global phenomenon and an outstanding manifestation of ocean–atmosphere coupling. As the societal impacts of ENSO became clear, a major field program [Tropical Ocean and Global Atmosphere (TOGA)] emerged. A moored observing system remains in place. Because entire books have been devoted to this phenomenon and its history of discovery (Philander 1990; Sarachik and Cane 2010; Battisti et al. 2019), no more will be said here.

The history of the past 100 years in physical oceanography has made it clear that a huge variety of phenomena, originally thought of as distinct from the general circulation, have important implications for the latter. These phenomena include ordinary surface gravity waves (which are intermediaries of the transfer of momentum and energy between ocean and atmosphere) and internal gravity waves. Great progress has occurred in the study of both of these phenomena since the beginning of the twentieth century. For the surface gravity wave field, see, for example, Komen et al. (1994).

For internal gravity waves, which are now recognized as central to oceanic mixing and numerous other processes, the most important conceptual development of the last 100 years was the proposal by Garrett and Munk (1972; reviewed by Munk 1981) that a quasi-universal, broadband spectrum existed in the oceans. Thousands of papers have been written on this subject in the intervening years, and the implications of the internal wave field, in all its generalities, are still not entirely understood.

7. Numerical models

Numerical modeling of the general ocean circulation began early in the postwar computer revolution. Notable early examples were Bryan (1963) and Veronis (1963). As computer power grew, and with the impetus from MODE and other time-dependent observations, early attempts [e.g., Holland (1978), shown in Fig. 7-15] were made to obtain resolution adequate in regional models to permit the spontaneous formation of balanced eddies in the model.

Present-day ocean-only global capabilities are best seen intuitively in various animations posted on the Internet (e.g., https://www.youtube.com/watch?v=CCmTY0PKGDs), although even these complex flows still have at best a spatial resolution insufficient to resolve all important processes. A number of attempts have been made at quantitative description of the space–time complexities in wavenumber–frequency space [e.g., Wortham and Wunsch (2014) and the references therein].
Ocean models, typically with grossly reduced spatial resolution, have, under the growing impetus of interest in climate change, been coupled to atmospheric models. Such coupled models (treated elsewhere in this volume) originated with one-layer "swamp" oceans with no dynamics. Bryan et al. (1975) pioneered the representation of more realistic ocean behavior in coupled systems.

### a. The resolution problem

With the growing interest in the effects of the balanced eddy field, the question of model resolution has tended to focus strongly on the need to realistically resolve both it and the even smaller submesoscale with Rossby numbers of order 1. Note, however, that many features of the quasi-steady circulation, especially the eastern and western boundary currents, require resolution equal to or exceeding that of the eddy field. These currents are very important in meridional property transports of heat, freshwater, carbon, and so on, but parameterization of unresolved transports has not been examined. Figure 7-16 shows the Gulf Stream temperature structure in the WOCE line at 67°W for the top 500 m (Koltermann et al. 2011). The very warmest water has the highest from-west-to-east \(u\) velocity here, and its structures, both vertically and horizontally, are important in computing second-order products \(\langle uC \rangle\) for any property \(C\)—including temperature and salinity. Apart from the features occurring at and below distances of about 1° of latitude, like the submesoscale, the vertical structure requires resolution of baroclinic Rossby radii much higher than the first one.

For the most part, ocean and climate modelers have sidestepped the traditional computational requirement of demonstrating numerical convergence of their codes, primarily because new physics emerges every time the resolution is increased. Experiments with regional, but very high resolution, models suggest, for example, that, in the vicinity of the Gulf Stream (and other currents), latitude and longitude resolution nearing 1/50° (2 km) is required (Lévy et al. 2010). In the meantime, the computational load has dictated the use of lower-resolution models of unknown skill—models sometimes labeled as "intermediate complexity" and other euphemisms. When run for long periods, the accuracy and precision in the presence of systematic and stochastic errors in under-resolved models must be understood.

### b. State estimation/data assimilation

The meteorological community, beginning in the early 1950s (Kalnay 2002), pioneered the combination of observational data with system dynamics encompassed in the numerical equations of general circulation models [numerical weather prediction (NWP) models]. Almost all of this work was directed at the urgent problem of weather forecasting and came to be known as "data assimilation." In the wider field of control theory, data assimilation is a subset of much more general problems of model–data combinations (see Brogan 1990). In particular, it is the subset directed at prediction—commonly for days to a couple of weeks.

When, in much more recent years, oceanographers did acquire near-global, four-dimensional (in space and time) datasets, the question arose as to how to make "best estimates" of the ocean using as much of the data as possible and the most skillful GCMs. The well-
developed methods from NWP were used by some (e.g., Carton and Giese 2008) to make estimates of the changing ocean, ignoring the important point that the methods were directed at prediction and were not general best estimators. Without the urgent demands for short-range forecasts that drove the meteorological methods, the oceanographic problem was (and still is) primarily that of scientific understanding of the time-evolving system. In that context, the now conventional data-assimilation methods were scrutinized for physical consistency, and it came to be recognized that prediction schemes, and as used in so-called reanalyses, failed to satisfy basic global conservation laws (heat, freshwater, vorticity, energy, etc.). The more general methods of control theory specifically distinguish the prediction problem from the “smoothing” problem whereby the results are intended to satisfy known equations of motion for all times, without the unphysical jumps in state variables. Such jumps have no detrimental effect on the prediction problem. They do raise fundamental questions when used for understanding the physics of a system.

With this recognition, some effort has gone into finding estimates of the time-evolving ocean circulation over months to decades that would be physically consistent. These early explorations (e.g., Stammer et al. 2003) generally began with the smoother problem in which the hypothetically optimal sequential prediction algorithm (the Kalman filter) was replaced by sequential methods producing a dynamically self-consistent solution.13

At the present time, although not yet in widespread use, multidecadal estimates of the ocean circulation employing Lagrange multipliers do exist (e.g., Fukumori et al. 2018) and are being analyzed to understand the mean and time-changing ocean. These solutions remain of coarser resolution than theory requires; the very great power of geostrophic balance governing the interior data is used to argue, as in the Stommel–Arons theory, that essentially passive boundary currents will transport the required mass and, at least to some extent, heat, salt, and so on that are necessary to reproduce the constraints of interior observations. With unresolved boundary dynamics, boundary currents will be dominated locally by dissipative processes.

c. **High-resolution models**

With the increase in computer power, it has recently become possible to run high-resolution numerical simulations of ocean regions a few hundred kilometer across. Such simulations have mesh grids of a few meters and resolve most of the whole of ocean physics down to the scale of breaking gravity waves. At this resolution the models can be configured to simulate ocean regions that are targets of field campaigns to fill the observational gaps. For example, many of the recent advances in understanding of the submesoscale dynamics have come from a careful coordination of numerical experiments and field campaigns [e.g., the Scalable Lateral Mixing and Coherent Turbulence (LatMix) process study described in Shcherbina et al. (2015)].

Ocean models have become an essential tool to interpret both the global ocean state and also its dynamics at the mesoscale and below. The challenge remains to bring all of these interacting scales into a unified picture.

8. **The Southern Ocean**

As already noticed, the Southern Ocean had long been recognized as having a physics that is distinct from the midlatitude and equatorial regions (Fig. 7-17). Without a supporting pressure gradient, ordinary Sverdrup dynamics cannot be applied to flows within the zonally unbounded latitude band of Drake Passage. Mesoscale

---

13 In a linear, or linearizable, discrete dynamical system, the optimal predictor is what is known as the Kalman filter. It is an elegant solution, and for linear systems it provides an optimal prediction if accurate covariances are used. Although its use has become a “buzzword” in the climate community, its computational cost is so great that in practice it is never used in weather or climate forecasting.
eddies are now believed to play a central role both in the
dynamics and thermodynamics, making the Southern
Ocean a turbulent fluid that cannot be understood with
linear models.

The basic features of the circulation in the Southern
Ocean had been identified as early as the mid-1930s
(Sverdrup 1933; Deacon 1937) from hydrographic mea-
urements collected primarily during the Challenger
(1872–76) and Meteor (1925–27) Expeditions. The strong
eastward flow of the ACC connects each of the ocean
basins and is part of a global overturning circulation, in-
ferred following property extrema such as an oxygen
minimum or salinity maxima, consisting of deep water
that spreads poleward and upward across the ACC and is
balanced by an equatorward flow in lighter and denser
layers as sketched in Fig. 7-18.

Deacon (1937) originally showed that this flow pat-
tern is consistent with the wind stress acting on the sea
surface. A deep return flow must develop to compen-
sate the northward Ekman flow and close the over-
turning circulation. This overturning cell was named
the Deacon cell by K. Bryan (Döös et al. 2008). Dea-
con, however, had only suggested that there ought to
be a subsurface return flow, and not a deep cell ex-
tending well below the thermocline. The full cell in-
stead appeared in the solutions computed much later
by Wyrtki (1961) using observed winds and available
hydrographic sections.

At first, it was assumed that the Deacon cell involved
deep downwelling of actual water parcels, which is dif-
ficult to reconcile with the prevailing strong stratification
because it would require very strong diapycnal flow—
especially in the downwelling branch. Furthermore the
Deacon cell consisted of a closed loop with no exchange
of waters with the basins to the north, again inconsistent
with the circulation inferred from measured property
extrema.

The problem came into focus in the 1990s as computer
power increased to the point that full global simulations
of the three-dimensional ocean circulation became fea-
sible (e.g., McWilliams 1998). Model solutions produced
a Deacon cell within the latitude band of the Drake
Passage, but they were characterized by vertical iso-
pycnals there, a result of resolution inadequate to pro-
duce the inevitable baroclinic instability. When the Gent
and McWilliams (1990) eddy-effects parameterization
was implemented in ocean models, it produced an
overturning circulation that crossed density surfaces
only in the surface mixed layer and that resulted in a
much more realistic stratification. The parameteriza-
tion, however, could not be tested easily because the
resulting isopycnal slope depended on an arbitrary Aus-
tausch coefficient in the representation.

It took another decade to realize fully that the lack of
lateral continental barriers in the latitude band of the
Drake Passage makes the Southern Ocean dynamically

Fig. 7-17. The 20-yr (1994–2013) average time-mean meridional flow at 5 m from the Estimating the Circulation
and Climate of the Ocean (ECCO) Consortium state estimate. Red values are positive and blue values are negative.
Intense structure in the Southern Ocean is an indicator of the strong topographic effects and the generally distinct
physics of that region. The strong (ageostrophic) divergence near the equator is also conspicuous in the Atlantic and
Pacific Oceans.
similar to the midlatitude atmosphere (Gent et al. 1995; Marshall and Radko 2003, 2006.) In particular, the “vanishing” of the Deacon cell is analogous to the atmospheric Ferrel cell, which is also cancelled by a counter-rotating eddy-driven cell. In both atmosphere and ocean a “residual” overturning circulation appears to exist that is composed of two counter-rotating cells stacked on top of each other. This residual circulation is the net circulation experienced by tracers in the ocean and sketched in Fig. 7-18—an estimate supported by modern observations and inverse models (Ganachaud and Wunsch 2000; Lumpkin and Speer 2007; Talley 2013). Marshall and Speer (2012) review the recent understanding of the Southern Ocean dynamics.

Observations had since demonstrated that the Southern Ocean is a critical element in the global overturning circulation—blending waters from the salty Atlantic with the fresh Indian and Pacific Oceans, and producing dense waters along the shelves of Antarctica that filled the ocean abyss at rates comparable to the dense waters formed in the North Atlantic (see the discussion in Sloyan and Rintoul (2001)). It was not until a basic understanding of the dynamics of the Southern Ocean had emerged that the theoretical focus shifted to the interaction of the Southern Ocean with the ocean basins to the north, however. Biogeochemists were instrumental in bringing the question to the fore, because they realized that high latitudes appear to exert a strong control on the CO₂ concentration of the global ocean and hence the atmosphere (Knox and McElroy 1984; Sarmiento and Toggweiler 1984; Siegenthaler and Wenk 1984). Toggweiler and Samuels (1995) showed that in models the AMOC is very sensitive to the strength of the winds blowing over the Southern Ocean. Full theories of the overturning circulation that couples the Southern Ocean to the Atlantic, Indian, and Pacific Oceans are now under development [see, in particular, Gnanadesikan (1999), Wolfe and Cessi (2010), Nikurashin and Vallis (2011), and Ferrari et al. (2017)].

9. Other physics

a. Paleophysical oceanography

The study of paleoceanography, as a subset of paleoclimatic, gradually matured in the years after WWII, with most of the foundation built upon the measured isotopic ratio techniques developed during the war, in a geochemical/geological setting. Much of the most useful data, roughly corresponding to time series, came from drilling into deep-sea sediments and measuring quantities such as the chlorinity (salinity) of pore waters and the variations in isotopic ratios such as [18O]/[16O] and [3H]/[1H], with the square brackets denoting concentrations (see, e.g., Bradley 1999; Cronin 2009). The most robust inference is that the climate of the past—extending out to the age of Earth—has commonly been radically different from that of today, ranging from global glaciation episodes to far warmer climates with no
ice present at all. What the role of the ocean has been in these climatic changes and how the necessarily different ocean circulations would have influenced the isotopic ratio distributions and their interpretation became urgent problems.

In practice, paleoceanographic time scales span a range from the decades before global instrumental data became available to thousands and hundreds of millions of years ago. The problems of paleoceanographic circulation inference are manifold but are mainly due to the paucity of data (deep-sea core measurements are possible only in the limited regions of major, preserved, sedimentary structures on the seafloor, and properties within cores can undergo complex in situ transformations), and little information exists about the state of the overlying atmosphere. Combined with the uncertain relationships between the “proxies” measurable in cores and physical variables of interest such as water temperature, making inferences about the ocean circulation becomes an exercise in stacked assumptions. In particular, absent knowledge of the wind field, computing past circulations is especially difficult. Paleophysical oceanography was reviewed by Huybers and Wunsch (2010).

Coupled models of the atmosphere and oceans, and sometimes with land and sea ice, are also now being used (e.g., Muglia and Schmittner 2015) although lower-resolution versions exist from earlier efforts. All of the difficulties of modern coupled modeling are encountered, but these are greatly aggravated by the lack of data and the immense range of time scales encountered in both modeling and observations. Barring a breakthrough in the observational problem, these coupled models might be labeled a form of “geonovel” or “geonovel” (following Harry Hess on seafloor spreading) and will remain so for the foreseeable future.

b. Multitudes of physics

Ocean physics encompasses a wide variety of phenomena that are not discussed here. Among the topics omitted are the important subfields devoted to sea ice, the Arctic Seas, tidal dynamics, flow and sediment interactions, physics of the air–sea boundary, and the entire related fields of geochemistry and biology as well as the surface and internal wave problems to which we have already alluded. All of these have some bearing on the general circulation and the deduction of its properties. The time at which a one-volume compendium (Sverdrup et al. 1942) could encompass all of oceanography vanished long ago—a measure of how far we have collectively come.

10. Future of the subject

Prediction of the future ocean circulation—under global climate change—is a difficult subject in its own right, and the history of such efforts is beyond our present ambitions. One can nonetheless speculate about the science itself. This review is being written at a time of very great uncertainty about the future of U.S. science, particularly that part related to the environment and climate change in general. What the next few years will bring to U.S. science is unpredictable. A particular worry concerns the necessary ongoing observations of the ocean—observations that are almost wholly government supported. In the bleakest outcome, one or more generations of expert scientists will be lost, whose experience and interest it would take decades to recover, and data gaps will exist that can never be filled. Whether the rest of the world can or will compensate for a decrease in U.S. efforts also remains an enigma.

Setting aside this possibility of a new dark age, we can consider the trajectory of physical oceanography as it has existed for the past few decades and attempt to extrapolate it into the future. That the emphasis in any science generally depends also upon the supporting societal infrastructure renders fraught any prediction.

The earlier reasonable assumptions that a few basic principles applied almost everywhere (e.g., the Sverdrup relation or uniform upwelling) have become untenable: many regionally distinct “oceans” exist with differing physics and expected differences in future response. These oceans differ by geography, by time scale, and by which physical elements dominate on varying space and time scales.

a. Future instrumentation and observations

It is doubtful that anyone in 1919 could have foreseen the technologies of ocean observation that had developed by the end of the twentieth century. Whether the advances of the past 100 years have been unique—as they made available the fruits of the electromagnetic and quantum revolutions—is an imponderable.

Physical oceanography in all its generality is likely to remain an observational science for the foreseeable future: even as models become more capable, ever more detail will require testing and confirmation. For the global problem—our focus here—the move away from ships as the primary platform must continue. Growth can be expected in the autonomous float, drifter, autonomous vehicle, and animal-borne technologies, preferably to the point where the whole oceanic volume is covered sufficiently rapidly that the residual space–time aliasing is tolerable or is at least quantitatively bounded. Historically, the combination of Eulerian fluid dynamics along with near-Eulerian measurements has led to the deepest insights into basic fluid physics. The understandable growth of freely moving devices—with low production costs and easy deployments—has taken the
field into the complications of Lagrangian fluid mechanics and near-Lagrangian measurements. The fluid mechanics theory is much more difficult, and the measurements inextricably mix temporal and spatial structures and statistics. Revival of serious global-scale moored measurements using more modern technology is needed. History suggests that the appearance of some entirely unforeseen new technology should be expected, but "if" and "what" are shrouded in darkness. Perhaps the under-exploited methods of acoustics—for example, natural-sound tomography—or of biologically based sensors will develop into routine global-scale measurements.

For satellites, the main challenge is likely to be the maintenance and incremental improvement of the existing technologies—almost the whole accessible electromagnetic spectrum has already been explored. If entirely novel remote sensing methods can exist, they are unknown to us.

Physical oceanography is likely to remain distinct from physical/dynamical meteorology into the future. Weather forecasting, its gradual extension into seasonal time scales, and its improvements are likely to remain the focus of the atmospheric sciences. Although the ocean has a clear analog of "weather" in its balanced and near-balanced eddy field, detailed prediction in any particular region is likely to remain of most interest to narrow military, fishing, or shipping groups. Prediction of the basin-to-global oceanic state decades in the future will be of intense scientific and practical interest, but the long times required to test any such predictions mean that the focus of the science is likely to continue to be on understanding rather than on prediction. The infrastructure of weather and climate forecasting, largely built on major civilian government laboratories and funding resources around the world, has no counterpart in oceanography. Thus, the research flavor of the two fluids is likely to remain distinct.

b. Future theory and modeling

Much of the existing effort surrounding physical oceanography on the large scale is directed at the use and interpretation of model results. As has been true of ocean numerical modeling since its beginnings in the 1950s, much of the focus concerns the effects of inadequate resolution on the large-scale flows. Over the next decades, two possible routes to full success can be envisioned: 1) Continued growth in computer power, perhaps through further revolutionary (e.g., quantum) methods, will ultimately permit complete resolution of all space and time scales of importance. Whether that includes reaching the laminar (viscous/diffusive limit) scale and the smallest important topographic scales on a global basis is not known. The computing power and storage requirements are entirely forbidding, but the capability available today would have seemed miraculous in 1919. 2) Fully adequate parameterizations of all scales of oceanic turbulence (at least from the mesoscale through internal waves and submesoscale waves to the inertial subrange) will be developed—conceivably with machine-learning algorithms trained with observations and high-resolution numerical simulations—permitting the quantitatively accurate calculation of their influence on coarser resolved scales.

As seems likely, the physics governing the circulation and variability will be distinct, at least to a degree, at every spatial point. The intellectual challenge will then be to extract the major governing principles operating globally so that one will have come full circle from the initial global- and basin-scale theories to the full understanding of regionality—and then back to determining the universal, overall, governing principles. In a rational world, the next 100 years ought to be very interesting!

Acknowledgments. We had very helpful comments on an earlier draft from J. Pedlosky, G. Vallis, A. Gordon, R. Williams, and B. Owens. We thank C. Garrett for careful reviews and appreciate also some of the comments from an anonymous reviewer who thinks that the chapter should have been mainly about numerical models. No implication of consensus about our choices of topic, emphases, or omissions is intended. Preparations was supported in part by funds from Harvard University and the Massachusetts Institute of Technology.

Author RF acknowledges NSF support through Awards OCE-1536515 and OCE-1736109.

REFERENCES


