
dx.doi.org/10.1016/B978-0-12-409548-9.11410-1
© 2019 Elsevier Ltd. All rights reserved.
Introduction

A general circulation model (GCM) of the ocean is nothing more than that—a numerical model that represents the movement of water in the ocean. Models, and more particularly, numerical models, play an ever-increasing role in all areas of science; in geophysics broadly, and in oceanography specifically. The range of length scales and timescales in oceanography is considerable. Important dynamics, such as that which creates "salt fingers" and hence influences the dynamically significant profile of density versus depth, takes place on centimeter scales, while the dominant features in the average circulation cascade all the way to basin scales of several thousand kilometers. Timescales for turbulent events, like waves breaking on shore, are small fractions of seconds, while at the opposite end, scientists have reliably identified patterns in the ocean with characteristic evolution times on the order of several decades.

Ocean modeling is a relatively new discipline within the field of oceanography. The underlying principles of the algorithmic formulation of ocean circulation models were first proposed in the 1960s by Bryan (1969) (see also McWilliams (1996) for an historical review). In the decades that followed, new technologies and exponential increases in computer processing capabilities have enabled ocean circulation models to deliver ever more meaningful and complete descriptions of the ocean circulation. Once designed simply to simulate ocean circulation as a stand-alone program, today’s ocean circulation models can also be integrated into data assimilation frameworks where they are used for ocean reanalysis and to produce short-range ocean forecasts. They can be coupled to atmospheric circulation models in order to produce seasonal to decadal forecasts. And when fully integrated into Earth System Models, they are essential to climate modeling. Ocean circulation models are also the preferred experimental method for increasing our ability to understand the ocean’s complex and diverse physical mechanisms.

Ocean circulation models must account for a variety of oceanic physical processes, and the number of processes that can be included in the models has broadened over past decades. When first introduced, ocean circulation models were only able to describe oceanic properties and physical processes at scales significantly larger than the mesoscale (horizontal scales on the order of 100 km and time scales on the order of 3 months). Today, ocean circulation models routinely resolve oceanic flows down to the submesoscale (horizontal scales on the order of 10 km or less; Chassignet and Xu (2017)) and to describe internal wave and internal tides on global scales (Shriver et al., 2012). This is due, in large part, to increases in computing power and the improved physical consistency of their formulation now possible.

However, model resolution and the propagation of uncertainty in forecasts remain significant challenges in the field. For example, increased model resolution has not solved the problem of subgrid scale closures (unresolved processes). In fact, some of the underlying assumptions and algorithms of current ocean circulation models (hydrostatic vs. non hydrostatic, for example) have been called into question using higher model resolutions. Furthermore, the modularity of modern geoscientific models requires the design of robust and physically rational approaches for coupling ocean circulation models with other model components.

Definition

Ocean modeling is a type of numerical modeling that focuses on the representation of the physical mechanisms governing the evolution of ocean physical properties, namely $T$, $S$, $u$, $v$, and $w$, where $T$ is the temperature, $S$ the salinity, and $u$, $v$, and $w$, the horizontal and vertical components of the velocity $V$. Because a numerical model is constructed using discretized equations of motion, building one capable of realistically representing the ocean circulation requires an understanding of a wide variety of physical processes. Without this foundational knowledge, one could easily draw erroneous conclusions. However, given the current
capacity of computers, direct numerical simulations (DNS) of the ocean (i.e., numerical representation of the smallest turbulent scale, i.e., the Kolmogorov length scale, which is on the order of 1 cm (Smyth et al., 2001)), are not yet possible. The largest direct numerical simulation attainable today is on scales of 10 m (Yeung et al., 2015).

Without the ability to represent smaller scales, we only have the ability to generate a truncated representation of the ocean. Of course, our capacity will expand as technology advances, but direct numerical simulations of the ocean will not be possible in the foreseeable future. The spatial and temporal scales that can currently be represented depends strongly on the model configuration and application. For example, coastal models require accurate sea level representation (response to wind, tides, and surface pressure) and operational oceanography requires accurate depiction of upper ocean structure and of mesoscale features such as eddies and meandering fronts, that is, "ocean weather." High horizontal resolution (1/10°–1/25° and very rarely to 1/50°; Chassignet and Xu (2017)) is mostly used to investigate ocean processes on seasonal to decadal time scales. The emphasis is on relatively short integrations (years to decades) and most models of this class are stand-alone (often coupled to a sea-ice model) and use prescribed atmospheric fields. Coarser resolution (1/4°–1°) is principally used for climate applications (Griffies et al., 2000). This class of models emphasizes long integrations, decades to millennia, with fully coupled ocean-ice-atmosphere models.

In practice, ocean models consist of a numerical solution to a set of partial differential equations (PDEs) describing the ocean dynamics. These PDEs are based on an approximated version of Navier-Stokes equations adapted to our regimes of interest (see the review papers by Griffies (2004) and Griffies and Treguier (2013) for details). Ocean numerical models are typically written in FORTRAN, consisting of 20,000–200,000 lines of code, and they can take up to a decade in community development before they are fully functional.

Numerical modeling of geophysical fluid started approximately 70 years ago with numerical weather prediction. Bryan (1969) introduced the first ocean numerical model 20 years later. The first discretized equations for an atmospheric application were solved manually by Richardson (1922) and numerically by Charney et al. (1950). The latter consisted of a 15 by 18 grid (Δx = 736 km) and demonstrated that numerical weather prediction was possible. The latest global- and basin-scale ocean configurations use horizontal resolution Δx on the order of 1 km (Chassignet and Xu, 2017). Ocean model development lags behind that of atmospheric models, primarily because of the societal needs for meteorological forecasts, but also because of the inherently greater complexity of circulation systems in closed basins and a nonlinear equation of state for seawater. Furthermore, the computing power required to resolve the relevant physical processes (such as baroclinic instabilities) is far greater for the ocean than the atmosphere since these processes occur on much smaller scale in the ocean. Historically, ocean models have been used primarily to numerically simulate the dominant space-time scales that characterize the ocean system. Simulations of physical integrity require an ability to both accurately represent the various phenomena that are resolved, and to parameterize those scales of variability that are not resolved (Chassignet and Verron, 1998). For example, numerical advection schemes address the problems associated with transport representation, whereas the parameterization of subgrid scale transport is linked to turbulence closure considerations. Although there are often areas of overlap between representation and parameterization, the distinction is useful to make and generally lies at the heart of various model development issues.

Ocean numerical models have traditionally been one of the biggest users of computer resources and, in order to deal with the complexity of ocean mechanisms, they will continue require the latest state-of-the-art supercomputers. Computing architecture is constantly changing, which means that ocean modelers must collaborate closely computer scientists to ensure that numerical codes run efficiently and maximize the computing capabilities. The main limitation is not computational speed of the processors, but access to memory and latency in reading/writing on disk drives (I/O). This is not likely to change in the near future since supercomputer development is closely linked to the performance of commodity chips, which are not well adapted to ocean applications (i.e., GPUs) and do not facilitate memory access.

It is important to note that, while numerical models are necessary to understand ocean dynamics, they cannot completely simulate reality because of limitations in computational power, incomplete understanding of subgrid scale parameterizations, poorly known forcing fields, and poorly understood interactions with other components of the earth’s system such as the atmosphere and sea ice. Observations are also not an accurate representation of reality because of the many space and time gaps in the observations that give us only limited information about the ocean’s state, its variability, trends, and possible instabilities and regime shifts. In summary, while numerical models allow for hypothesis testing and experimentation, it is important to know their limitations, with the primary challenge being the quantification of truncation errors introduced by the discretization of the Navier-Stokes equations.

Truncation errors arise from the discretization of PDEs derived from the Navier-Stokes equations, which usually assumes that the ocean is incompressible and makes the spherical approximation (assumes the earth is a sphere), the thin-shell approximation (allows to neglect variations of the local rotation rate with respect to depth, therefore simplifying the treatment of the Coriolis acceleration), the Boussinesq approximation (neglects variation of relative density in the horizontal momentum equations), and the hydrostatic approximation (allows for neglecting the vertical acceleration). Before the Navier-Stokes differential equations can be solved numerically, they must be converted into an algebraic system, a conversion process that entails numerous approximations. Numerical modelers strive to achieve numerical accuracy. Otherwise, the discretization or "truncation" error introduced when approximating differentials by finite differences or Galerkin methods becomes detrimental to the numerical realization. Sources for truncation errors are plentiful, and many of these errors depend strongly on model resolution. Examples include horizontal coordinates (spherical and/or generalized orthogonal), vertical and horizontal grids, time-stepping schemes, representation of the surface and bottom boundary layers, bottom topography representation, equation of state, tracer and momentum transport, subgrid scale processes, viscosity, and diffusivity. Numerical models have improved over the years, not only because of better
physical understanding, but also because modern computers permit a more faithful representation of the differential equations by their algebraic analogs.

The dominance of lateral over vertical transport is a key characteristic of rotating and stratified fluids, including the ocean. Therefore, the customary approach in ocean modeling is to orient the two horizontal coordinates orthogonal to the local vertical direction as determined by gravity. The more difficult choice is how to specify the vertical coordinate. Indeed, as noted by various ocean modeling studies, the choice of a vertical coordinate system is the single most important aspect of an ocean model’s design. The practical issues of representation and parameterization are often directly linked to the vertical coordinate choice. Currently, there are three main vertical coordinates in use, none of which provide universal utility. Hence, many developers have been motivated to pursue research into hybrid or generalized coordinates approaches.

There are three regimes of the ocean that need to be considered when choosing an appropriate vertical coordinate (Griffies et al., 2000). First, there is the surface mixed layer. This region is generally turbulent and dominated by transfers of momentum, heat, freshwater, and tracers. It is typically very well mixed in the vertical through three-dimensional convective/turbulent processes. These processes involve non-hydrostatic physics, which requires very high horizontal and vertical resolution to be explicitly represented (i.e., a vertical to horizontal grid aspect ratio near unity). A parameterization of these processes is therefore necessary in primitive equation ocean models. In contrast, tracer transport processes in the ocean interior predominantly occur along constant density directions (more precisely, along neutral directions). Therefore, water mass properties in the interior tend to be preserved over large space and time scales (e.g., basin and decadal scales). Finally, there are several regions where density driven currents (overflows) and turbulent bottom boundary layer processes act as a strong determinant of water mass characteristics. Many such processes are crucial for the formation of deep water properties in the world ocean.

The simplest choice of vertical coordinate is \(z\), which represents the vertical distance from a resting ocean surface. Alternatively, one could use the potential density referenced to a given pressure. In a stably stratified adiabatic ocean, potential density is materially conserved and defines a monotonic layering of the ocean fluid. A third option would be to use the terrain-following \(\sigma\) coordinate. The depth or \(z\) coordinate provides the simplest and most established framework for ocean modeling. It is especially well-suited for situations with strong vertical/diapycnal mixing and/or low stratification, but has difficulty in accurately representing the ocean interior and bottom. The density coordinate, on the other hand, is well-suited to modeling the observed tendency for tracer transport to be along density (neutral) directions, but is inappropriate in unstratified regions. The \(\sigma\) coordinate provides a suitable framework in situations where capturing the dynamical and/or boundary layer effect associated with topography is important. Terrain-following \(\sigma\) coordinates are particularly well suited for modeling flows over the continental shelf, but are rarely used in a global modeling context. They have been used extensively for coastal engineering applications and prediction, as well as for regional and basin-wide studies.

Ideally, an ocean model should retain its water mass characteristics for centuries of integration (a characteristic of density coordinates), have high vertical resolution in the surface mixed layer for proper representation of thermodynamical and biochemical processes (a characteristic of \(z\) coordinates), maintain sufficient vertical resolution in unstratified or weakly stratified regions of the ocean, and have high vertical resolution in coastal regions (a characteristic of terrain-following \(\sigma\) coordinates). This has led to the recent development of several generalized vertical coordinate numerical models that combine the advantages of the different types of vertical coordinates in optimally simulating coastal and open-ocean circulation features.

**Limits on Numerical Models**

Ocean circulation models do not account explicitly for the entire range of scale interactions that control ocean circulation. Because of the turbulent nature of oceanic flows, ocean circulation at a given scale is indeed fundamentally dependent on oceanic motions at scales ranging from global scale (of order 10,000 km) to dissipative scale (of order 1 cm). But the finite grid resolution of a particular ocean model configuration constrains the spectrum of scales of motions that are explicitly represented in the model solution. For instance, consider ocean mesoscale eddies which are known to play a fundamental role in shaping ocean circulation (McWilliams, 2008). Their representation in ocean models has motivated a large number of studies (see Hecht and Hasumi, 2008, for a review). Even the most high-end global ocean circulation models with grid resolutions down to just a few kilometers (Chassignet and Xu, 2017) are still not able to fully capture the dominant length scales of mesoscale variability at high latitudes (Hallberg, 2013). Scale interactions involving oceanic mesoscale eddies are therefore not explicitly represented at high latitudes in these models.

Most applications of ocean circulation models find describing the broadest possible spectrum of oceanic scales of motion at a given computational cost problematic. The overall computational budget of an application is constrained by the available computing resources and by the expected time-to-solution, the latter constraint being particularly limiting in operational applications that deliver near real time products. The model grid of ocean model components of operational systems is also usually constrained by other practical factors. The same model grid is often used for several years because of the effort required for tuning the physical model configuration, for calibrating the data assimilation components, and for preparing the downstream data production chains. Given these constraints, the optimal design of ocean model components in operational systems requires a robust and rational understanding of a priori what controls the spectrum of resolved process in ocean circulation models besides the model grid resolution. Developing this understanding of what controls the spectrum of resolved scales in an ocean model should rely on idealized process studies, model inter-comparison exercises, and confrontation of model solutions with observations.
A commonly accepted practical criteria to decide whether a flow feature is resolved in an ocean model is that there should be no less than $2 \pi \Delta x$ grid points spanning the feature (see for instance Griffies and Treguier (2013)). This is supported by the notion of effective resolution, defined by Skamarock (2004) as the smallest resolved scale that is not significantly affected by numerical dissipation due to discretization errors. The effective resolution of an ocean circulation model is about $6 \sim 10 \Delta x$, where $\Delta x$ is the horizontal grid resolution, but the validity of the concept of effective resolution is generally restricted to model configurations with small enough grid resolutions and using high order discretization schemes (Soufflet et al. 2016).

Improving the performance of ocean circulation models on modern high performance computing (HPC) platforms allows for the broadening the spectrum of resolved scales in ocean circulation models. The current trend in HPC is toward massively parallel machines with heterogeneous multicores (Giles and Reguly, 2014). But while modern HPC platforms can deliver a peak performance in the Petaflop/s range, existing ocean circulation models are unable to exploit this potential. The computational intensity, that is, floating point operations per memory access, of ocean models is very low because they are dominated by stencil operations (typical of discretized PDEs). So, they typically run at $\sim 5\%$ of the peak speed of the system. Also, ocean models have a very small vertical dimension, typically $O(10)$, so they scale more like 2-D domains than 3-D domains. In practice, $1/12^o$ and $1/25^o$ global ocean models might scale well to 8000 and 30,000 cores, respectively. Where scalability is eventually limited is by the communication overhead, load imbalance and latency inherent to spatial 2-D domain decomposition and by I/O overhead. Single processor performance depends on computational intensity, which can be improved by using higher order discretization schemes. However, scalability can be reduced by better single processor performance because relatively more time is spent in communications.

**Summary**

In this article, we have briefly described the fundamental principles that support the formulation of modern ocean circulation models and discussed how a modeling strategy proposed in the 1960s for solving the primitive equations was the impetus for ocean circulation models that are now used for a wide range of applications. Without data assimilation, they are primarily used to scientifically rationalize the observed ocean by testing, via idealized or realistic configurations, mechanisms underlying observations. With data assimilation, they are used to perform hindcasts to understand past evolution and to perform forecasts that can be used for societal applications (Chassignet and Verron, 2006; Dombrowsky et al., 2009; Schiller and Brassington, 2011; Bell et al., 2015). Observational data, via data assimilation, sets the stage for model state estimates and forecasts (Chassignet et al., 2009). The quality of the estimates and of the forecast will depend on the ability of the ocean numerical model to faithfully represent the resolved dynamics of the ocean and the parameterized subgrid scale physics. Thus, it is important to recognize that even the use of an infinite amount of data to constrain the initial conditions of an ocean model would not necessarily improve the forecast against persistence of a poorly performing ocean numerical model.

Ocean circulation model design is a very active field of research with entire scientific teams dedicated to developing or improving ocean circulation models. The maturity of the field is arguably a consequence of the high level of collaboration and merging of efforts among different groups involved in different applications and aspects of ocean circulation models. A key driver for this collaborative approach to ocean model development is this notion of seamless geoscientific modeling, which posits that the same numerical code can actually be used for a range of different applications covering a range of different resolutions and dynamical regimes (Hurrell et al., 2009). Overall, this approach has improved the robustness of ocean circulation models and the sustainability of ocean model development process over recent decades. The new frontiers in ocean circulation modeling are (i) the computational performance of ocean models, (ii) their ability to represent scale interactions either explicitly or through parameterizations, and (iii) the representation of model uncertainty with stochastic and ensemble approaches.

The field of ocean circulation model design has now reached a level of maturity and involves strong collaborations between different fields of expertise. It is also important to stress the crucial role that ocean observing networks play in improving ocean circulation models. Sustained ocean observations from satellites and in situ networks are key for routinely assessing the skills and the limitations of circulation ocean models over different timescales ranging from days to decades. There is much to be learned from targeted field observations geared toward documenting specific oceanic processes in order to improve their representation in models. For instance, recent experiments documenting fine-scale ocean processes in the ocean surface boundary layer provided a wealth of information that can be used for improving surface processes in ocean circulation models (Shecherbina et al., 2015; Buckingham et al., 2016).

Sustained collaborations between ocean modelers and computer scientists will be needed to overcome single processor and scalability issues. Practical approaches will likely involve more hybrid parallel programming in order to exploit more efficiently memory hierarchy and innovative algorithms for solving the set of PDEs that govern ocean dynamics as for instance parallelization in time in addition to spatial domain decomposition (Schreiber et al., 2017).

Looking to the future, ocean circulation models must be able to account explicitly for the broadest range of physical scales possible due to the multiscale nature of oceanic flows, as well as the evolving needs of end-users of operational oceanography. But more than just an expansion of computing power will be needed to broaden the spectrum of resolved scales. The efficiency of the algorithms used for translating ocean dynamical equations into practical computation will directly impact how efficiently ocean circulation models use. Anticipating what approach will be most used in the future for broadening the spectrum of resolved scales is not easy. But it is arguable that HPC and algorithmic aspects will become key in ocean model development in the future. New advances will most likely rely on a high level of collaborations between ocean modelers, applied mathematicians, and computer scientists.
Acknowledgments

This article is heavily influenced by and reiterates many of the key points found in Le Sommer et al. (2018).

References


Further Reading


