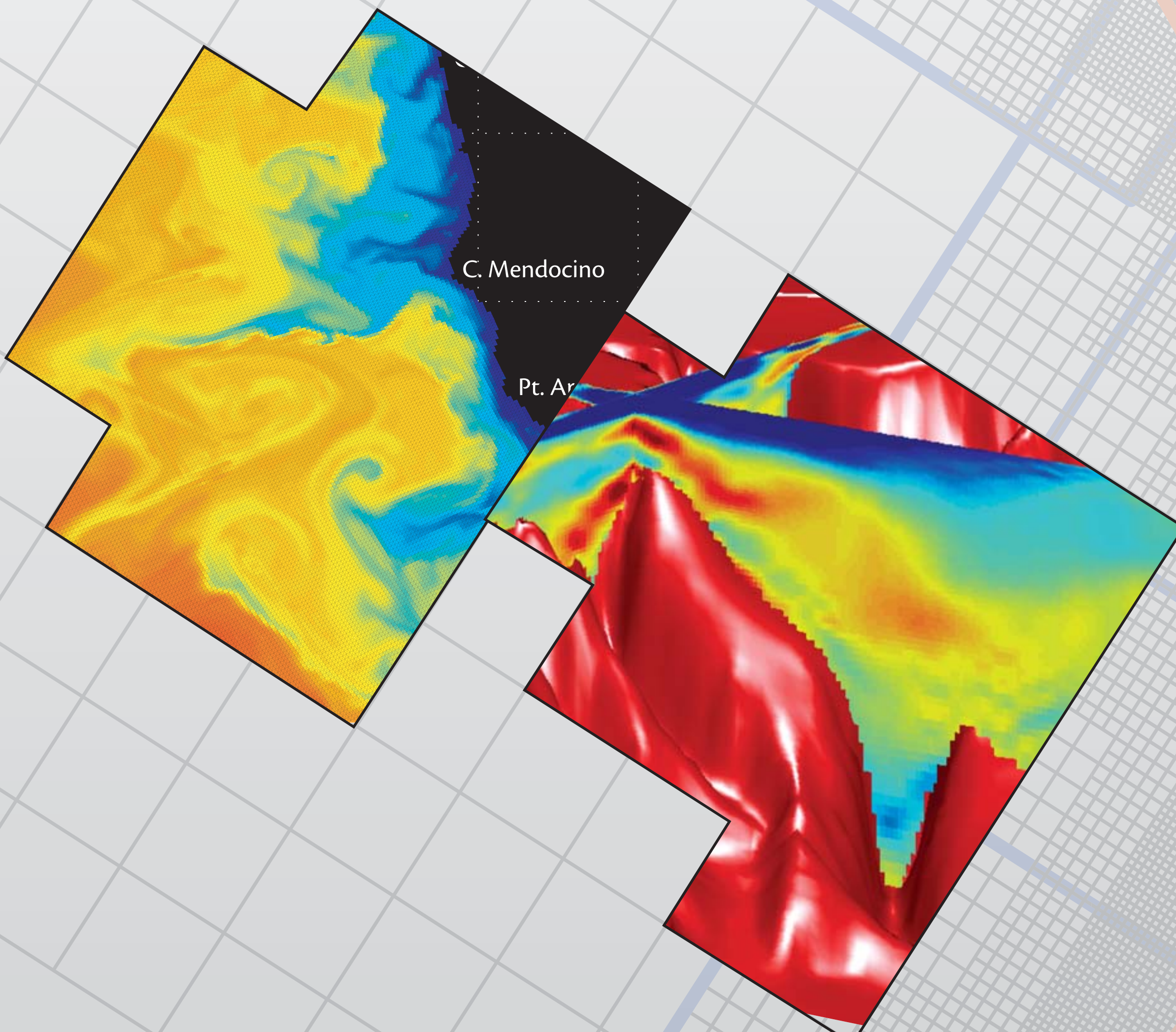
BY OLIVER B. FRINGER, JAMES C. MCWILLIAMS, AND ROBERT L. STREET

A New Hybrid Model FOR COASTAL SIMULATIONS



PHYSICAL PROCESSES IN the world's oceans span an enormous range of spatio-temporal scales, from ocean circulation at the global scale down to dissipation and mixing at the centimeter scale. It is impossible to simulate all of these processes using one model due to the shear computing power that would be required. Thus, different ocean models focus on specific processes and hence different spatio-temporal scales. At the largest scales are the global ocean circulation models, such as the Parallel Ocean Program (POP[1]), which is used primarily to simulate seasonal or climactic processes (i.e., the El Niño Southern Oscillation [ENSO] or thermohaline circulation). At the intermediate scales are the regional ocean models, such as the Princeton Ocean Model (POM[2]), the Regional Ocean Modeling System (ROMS[3,4]), or the MIT General Circulation Model (MITgcm[5]), which are well suited to modeling regional processes such as the Gulf Stream or the California Coastal Current. Indeed, these models are not necessarily limited to regional circulation studies; they have been applied extensively to simulate global circulation and, at the smaller scale, coastal processes, such as upwelling and internal waves. Ocean models that are specifically designed to simulate this end of the spatio-temporal spectrum are referred to as "coastal ocean models," which, like their regional modeling counterparts, are not necessarily limited to, but are best suited for, the coastal region, which includes estuaries. Models that are specifically designed for coastal and estuarine processes include the Stanford Unstructured Nonhydrostatic Terrain-Following Adaptive Navier-Stokes Simulator (SUNTANS[6]), DELFT3D[7], or the semi-implicit TRIM model of Casulli (1999). Coastal and estuarine processes distinguish themselves most significantly from regional and global processes in that they result from the interaction of ocean currents with complex boundaries and steep bathymetry. Models that simulate even smaller-scale processes, such as wind-waves, do not compute the three-dimensional circulation, but instead focus on computing time-averaged effects that can be included in larger-scale coastal circulation models, such as Simulating WAves Nearshore (SWAN[8]), a two-dimensional wind-wave model.

A hybrid model is generated when two or more models that focus on distinct spatio-temporal scales are coupled to form a simulation tool that can capture a larger range of spatio-temporal scales. With the advent of high-performance computing, the hybrid concept has gained popularity in recent years, combining the expertise behind different model development programs. In this article, a new hybrid simulation tool that is under development is presented. It attempts to tackle the problem of simulating regional circulation using ROMS while correctly incorporating the effects of smaller-scale coastal processes with SUNTANS.

[1] http://climate.lanl.gov/Models/POP; [2] http://www.aos.princeton.edu/WWWPUBLIC/htdocs.pom; [3] http://www.atmos.ucla.edu/cesr/ROMS_page.html; [4] http://marine.rutgers.edu/po/index.php?model=roms; [5] http://mitgcm.org; [6] http://suntans.stanford.edu; [7] http://www.wldelft.nl/soft/d3d/intro; [8] http://www.wldelft.nl/soft/swan

THE IMPORTANCE OF THE GRID

All ocean models represent discrete solutions of an approximate form of the Navier-Stokes equations. These equations govern the motion of fluid within the oceans at all scales, ranging from global circulation scales down to mixing and dissipative scales for processes like double-diffusion and turbulence in breaking internal waves. In some regions of the ocean, these processes can be as small as a few centimeters. Because fluid motions do not vary much at such small scales (they are smeared out by viscosity), we could represent, in principle, the full range of scales of motion in the oceans if we were to discretize them with cells that were one cubic centimeter in volume. A discrete solution of the flow field represented by these finite volumes using the Navier-Stokes equations would be a so-called direct numerical simulation (DNS) of all of the world's oceans, and it would eliminate the need for a turbulence model because all scales of motion would be computed. This type of simulation would require a total of 10^{22} grid cells, which would in turn require time-step sizes of 0.1 sec if we would expect to resolve the temporal behavior of motions at such small scales. A year-long simulation of the ocean, then, would require 315 million time steps. Because average simulation codes require roughly 100 mathematical operations per grid cell per time step, a DNS of the ocean would require a total of 3×10^{30} operations. If we had access to the world's fastest supercomputer, Blue Gene, which is located at Lawrence Livermore National Laboratory and is capable of performing roughly 300 trillion operations per second, then our simulation would take 333 million years! This, of course, is at the present rate of computing power, which, according to Moore's law, is doubling roughly every 18 months. Although Moore's prediction has stood the test of time over the past two decades, it is likely that computing power may begin to taper off soon. Assuming that it does not, however, and assuming that the simulation started today and was continuously transferred to improved computer systems as they emerged, the estimate of 333 million years surprisingly reduces down to a much more manageable 40 years.

A computation time of 40 years might suggest that a direct numerical simulation of the oceans is not too far off in the distant future. However, even if a system were devised that made such a tremendous simulation feasible, it would be impossible to provide initial conditions such as salinity, temperature, and velocity and wave-height fields with such high resolution. In addition, it would not be possible to provide high-resolution boundary conditions such as bathymetry (we would need to know every detail about every sand ripple and every coral reef!) or heat-flux and wind-stress fields. Furthermore, the shear volume of data that would result would be unmanageable. The alternative, of course, is what

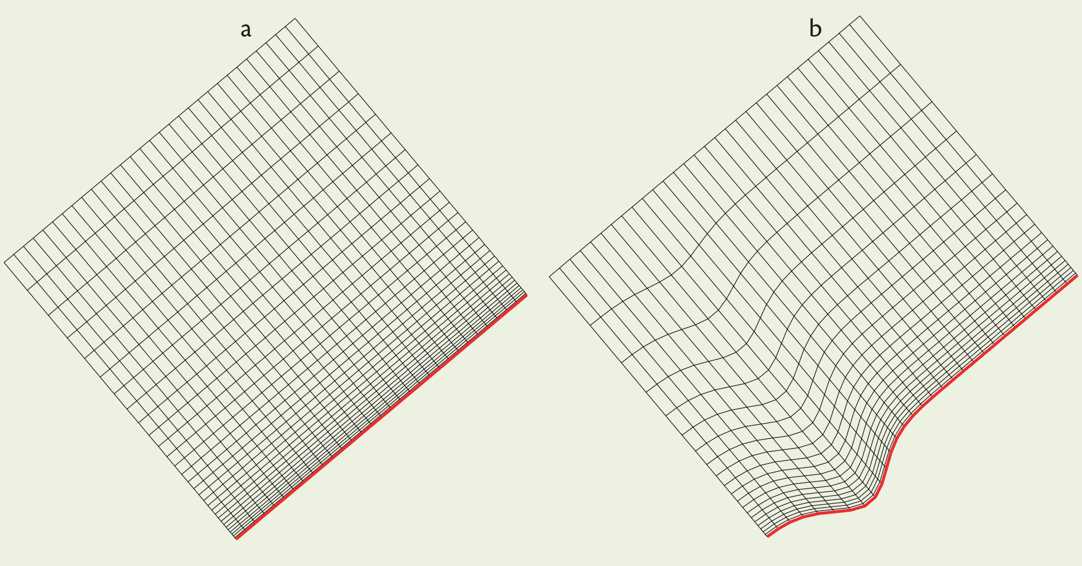
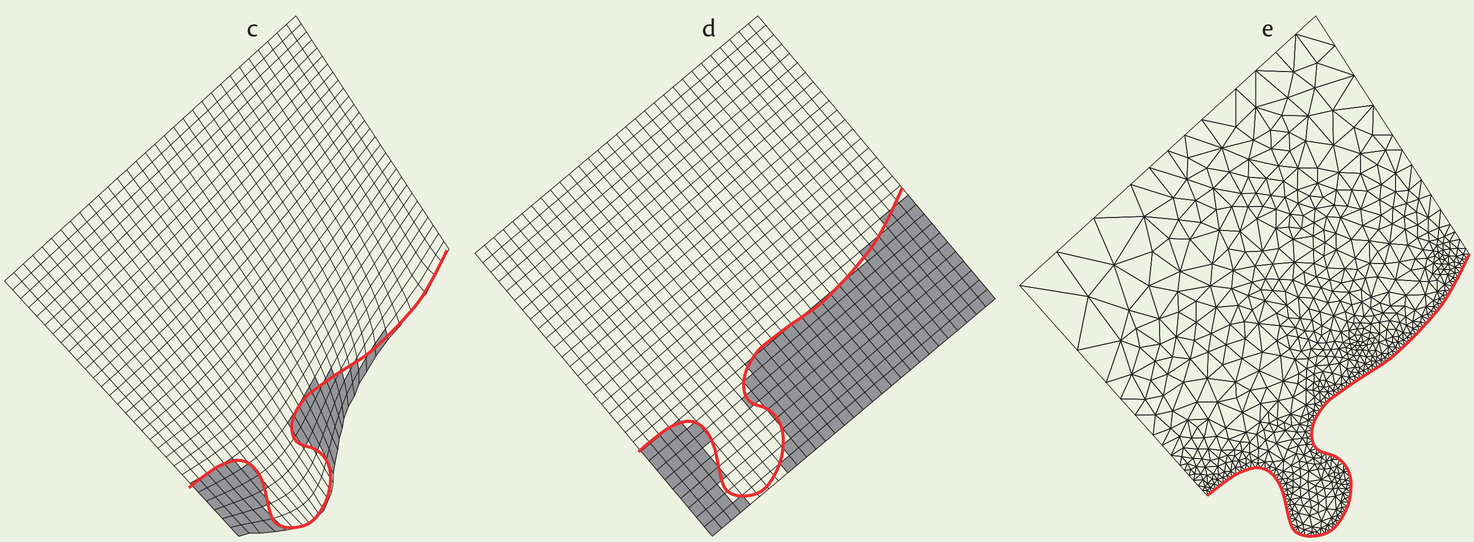


Figure 1. Different grid arrangements designed to capture multi-scale physics in the vicinity of a boundary (red line). (a) Grid stretching to resolve a straight boundary. (b) Grid stretching with a curvilinear grid near a relatively smooth boundary. (c) A curvilinear grid near sharply varying topography, which requires masking (gray cells) since the topography varies too quickly. (d) The masking procedure can also be performed with Cartesian grids, but the boundary is not followed as closely as in (c). (e) An unstructured grid allows for high resolution near irregular topography without the use of masking.

members of the ocean and atmosphere modeling communities have been doing for over half a century. Instead of computing all the scales of motion, simplifications of the Navier-Stokes equations are devised such that their discrete solutions are more tractable. The basic premise behind all numerical simulations is that because computing power dictates the number of points that will be used to discretize a given problem of interest, it sets the range of scales that will be simulated by a discretization of the equations themselves, while the effects of the remaining set of motions (often referred to as subgrid-scale motions) are parameterized, or modeled, based on values on the simulation grid that are computed.

To be precise, computing power sets the number of grid cells that can be used to simulate a problem of interest. However, it does not set the size and distribution of these grid cells. As a result, numerous methods have been devised to develop advanced simulation tools that enable the analysis of a broad range of scales by employing complex gridding techniques and advanced methods of discretizing the governing continuous equations most accurately, such as high-order advection schemes or high-order time-stepping techniques. Given an advanced discretization technique, the premise behind the accurate simulation of problems involving a broad range of length scales is that most multi-scale flows of interest are well represented by larger-scale motions, while a smaller percentage of the flows contain small-scale physics that require computationally expensive high resolution. The simplest example of a multi-scale flow is a parallel flow near a no-slip wall where a boundary layer exists. In this flow, the most logical way to handle the disparate length scales is to employ grid stretching, whereby the grid is refined near the wall where turbulence leads to small-scale eddies, and it is coarsened far from the wall where the flow is parallel and undisturbed by the motions of interest (Figure 1a). Of course, because most boundary layers occur near complex topography, a bulk of ocean models, in fact all of those mentioned in the introduction except for SUNTANS and TRIM, employ curvilinear grids that can follow topographic features while employing grid stretching to resolve boundary layers in the vicinity of coastal features (Figure 1b). This practice of employing

Oliver B. Fringer *(fringer@stanford.edu) is Assistant Professor, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA.* **James C. McWilliams** *is Louis B. Slichter Professor of Earth Sciences, Institute of Geophysics and Planetary Physics and Department of Atmospheric and Oceanic Sciences, University of California-Los Angeles, Los Angeles, CA, USA.* **Robert L. Street** *is William Alden and Martha Campbell Professor, Emeritus, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA.*



curvilinear coordinates for bottom-following grids to resolve bottom-boundary layers away from coastlines is referred to as sigma- or s-coordinates. Curvilinear coordinates have the disadvantage that the topography or coastlines that they follow must be relatively smooth. For sharply varying topography, it is possible to employ a curvilinear grid that follows the coastline as closely as possible, but only computing the flow field in the "wet'" regions (the unshaded regions in Figure 1c). This so-called "masking" strategy can also be employed for Cartesian, or rectangular grids, but the smoothly varying features of the coastline are not captured as well as they are with curvilinear grids (Figure 1d). The alternative is to employ unstructured grids, such as in SUNTANS and unTRIM (the unstructured version of TRIM; see Casulli and Zanolli, 2002), which have the advantage—they allow for high resolution in localized regions of flows in the vicinity of complex coastlines or embayments (Figure 1e). As with all aspects of computational fluid dynamics, a benefit associated with employing one seemingly advantageous methodology always comes with an associated cost, and it is probably not surprising that the more complex the grid, the more costly it is to generate. The structured grid in Figure 1a is trivial to generate, while the unstructured grid in Figure 1e requires the use of advanced grid-generation algorithms that are not even guaranteed to generate "good" grids. That is, depending on the geometry and other grid qualities, some cells in the grid may be so skewed that they render the entire grid useless unless specific measures are taken to deal with each particular grid cell.

A hybrid model is generated when two or more models that focus on distinct spatio-temporal scales are coupled to form a simulation tool that can capture a larger range of spatio-temporal scales.

CAPTURING MULTI-SCALE PHYSICS

Although a particular numerical technique to discretize the continuous equations of motion determines the accuracy of a simulation to a great extent, the choice of the grid is of crucial importance, because it determines which scales of motion can actually be simulated versus those that must be parameterized. Although simulating a large range of scales of motion comes at the obvious expense of additional computing power, the increased scales of motion also require an increase in complexity of the actual physics that must be computed. Consider the problem of the transport of effluent from a sewage outfall in the coastal ocean that is some distance from the surf zone. Employing a grid resolution of one cubic centimeter would enable the direct computation of turbulence that would allow for an accurate estimate of the rate of mixing of the pollutant field. However, such small grid resolution also requires an actual simulation of the effect of wind-waves on the surface of the ocean because these waves form, propagate, and break at scales that are much larger than 1 cm. In this case, the added expense associated with developing a simulation code that is able to compute surface-wave breaking would far outweigh the improvements in the prediction of the mixing and transport of the effluent plume of interest. It is therefore sometimes advantageous (although rare) to limit the size of the smallest grid spacing in order to prevent the appearance of additional physics that would add an expensive simulation component that could otherwise be modeled quite effectively, rather than directly computed.

Although the wind-wave example is an extreme case, a more common component of ocean models that is left out of the equations of motion because its inclusion adds a significant computational expense relative to its actual physical effect is the nonhydrostatic pressure. Of the models mentioned in the introduction, at present SUNTANS, TRIM, DELFT3D, and MITgcm are nonhydrostatic. In these models, the pressure field is split into a sum of its hydrostatic and nonhydrostatic components. Hydrostatic pressure is that which is associated with the weight of fluid above a specific point in the water column, and generally increases with depth. It also changes as a result of oscillations in the free surface as well as the density field, because a higher free surface implies a larger hydrostatic pressure while a denser fluid implies the same. The remaining component of the pressure field is the nonhydrostatic, or hydrodynamic pressure field, or that which is associated with fluid motion. Generally speaking, the effects of the nonhydrostatic pressure field become large in relation to the hydrostatic pressure field when the horizontal length scales of motion are less than or equal to the depth. Because a vast majority of length scales in the ocean are much greater than its depth, ocean models, such as POP and POM, do not compute the effect of the nonhydrostatic pressure on the flow field. This hydrostatic approximation is justified because the effect of the nonhydrostatic pressure is usually only felt in small regions of a flow, such as near steep bathymetry or where there may be short length-scale motions; this approximation has been a blessing to the ocean modeling community because computation of the effects of the nonhydrostatic pressure field typically takes three to five times longer than the effects of the hydrostatic pressure field.

A consequence of the hydrostatic approximation is that it imposes a restriction on the minimum horizontal grid resolution of the flow, because the hydrostatic approximation assumes that the horizontal length scales of motion are much larger than the vertical length scales. As a result, increasing the grid resolution in a hydrostatic model may result in an erroneous solution unless the nonhydrostatic pressure is incorporated into the model. An excellent example of what happens when a hydrostatic model is employed to compute a flow that is predominantly nonhydrostatic is depicted in Figure 2. The figure depicts the evolution of the internal lock exchange problem, whereby a hypothetical lock separates a dense, salty fluid from a light, fresh fluid, which represents the initial condition in Figure 2. Upon removal of the lock that separates the two fluids, the heavy fluid flows underneath the light fluid as a consequence of the hydrostatic pressure field, which is greater in the heavy fluid and hence forces that fluid

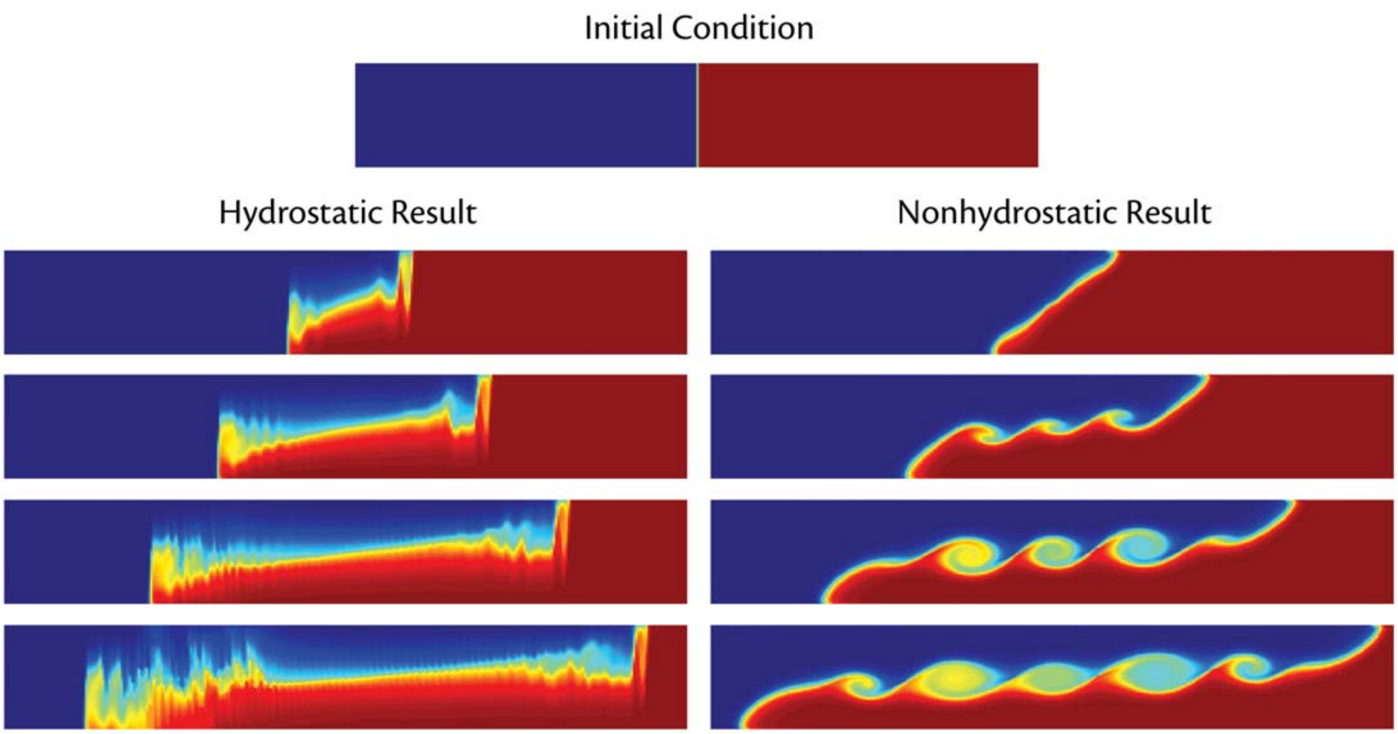


Figure 2. Effect of computing the lock-exchange problem with a hydrostatic and a nonhydrostatic simulation code. The figure shows how the hydrostatic code captures the large-scale dynamics by correctly computing the position of the front while it fails to capture the formation of the billows because their horizontal length scales are less than the depth, and therefore highly nonhydrostatic. The red fluid represents the salty or heavy fluid, while the blue fluid represents the freshwater or light fluid. Gravitational circulation and the hydrostatic pressure field enforce a recirculating flow field in which the salty, heavy fluid (red) flows beneath the fresh, light fluid (blue). For details on these simulations please refer to Fringer et al. (submitted).

underneath the light fluid. Although the acceleration of the two masses of water is driven predominantly by the hydrostatic pressure, the differences between the hydrostatic and the nonhydrostatic results are striking in that the hydrostatic result does not capture the billows that form at the interface between the two fluids. These billows are highly nonhydrostatic because the horizontal length scale of the motion is less than the depth. This example shows that the large-scale motion of the flow is captured quite well by the hydrostatic model, in that the position of the front is similar to that in the nonhydrostatic approximation. However, the hydrostatic model does not capture the finer-scale features of the flow and hence it is not justifiable to include such high horizontal resolution in the hydrostatic simulation because it cannot compute the fine-scale, nonhydrostatic physics of the flow.

With the advent of high-performance computing, the hybrid concept has gained popularity in recent years, combining the expertise behind different model development programs.

BASICS OF NESTING IN NUMERICAL SIMULATIONS

Although the implementation of different grids allows for a simulation to capture a wide range of spatial scales, especially when using unstructured grids, the minimum computation time required for the simulation is still limited by the smallest time scale involved in the simulation. That is, in a simulation that employs a grid with disparate grid sizes, the time-step size with which the flow over the entire grid advances forward in time is limited by the shortest time scale of motion on the grid. Therefore, very fine grid resolution comes at the penalty of requiring a large number of time steps to advance the solution over a specified time interval. Although a small time step is necessary to accurately compute flow in the regions where the grid is highly resolved, small time steps are not necessary in other regions of the flow where the grid is coarse. Ideally, then, a simulation that employs a multi-scale grid would also incorporate a spatially variable time-step size that would advance the solution in the fine grid with smaller time steps than in the coarse grid. In this manner, the unnecessary cost associated with computing large-scale flows with a small time step would be eliminated.

Unfortunately, the use of spatially variable time-step sizes in ocean models is not possible because, while the actual equations could, in principle, be developed to incorporate spatially variable time-step sizes, the equations would be too complex to solve on one single grid. However, it is feasible and computationally effective to employ several grids in a simulation and have the simulation on each grid advance forward at a time-step size that is set by the smallest grid spacing, and hence the smallest physical time scale, on that particular grid. The use of multiple grids requires that the solution on each grid is coupled with the solution on neighboring grids to maintain a continuous solution over the entire domain of interest. This concept is known as grid nesting, whereby different grids are employed in a simulation in a way that allows the grids that cover the fine-scale motions to advance their solutions forward in time with much smaller time-step sizes than the larger grids, thereby eliminating the need to advance long-time-scale motions forward in time with small time steps.

The decision on where to place the nested grids with different resolutions can be decided initially, based on a knowledge of where the flow is likely to encompass smaller spatio-temporal scales of motion, or it can be decided dynamically. The former is known as static mesh refinement, whereas the latter is known as adaptive mesh refinement (AMR), whereby refined grids are nested dynamically when more grid resolution is required because of the development of fine-scale motions in particular regions of the flow. Typically, a refinement ratio is specified that determines the ratio between the grid resolution on a parent grid and that on its child grid; one parent grid can have multiple children, each of which can have their own child grids (Figure 3). This grid could result from a calculation with adaptive mesh refinement, or it could be created statically given the knowledge that small, turbulent length scales are likely to develop in the vicinity of the boundary. Either way, this nested grid arrangement provides a great deal of savings because, although the smallest grid spacing is 1/16 the size of the largest grid spacing, high resolution exists only where it is needed. If this same domain were discretized with such high resolution over the entire domain, then 102,400 cells would be required, as opposed to 21,200 with the nested grid arrangement. Furthermore, because a small percentage of the flow contains fine grid resolution, then only a small percentage of the flow is advanced forward in time using small time steps, while the rest of the flow containing the coarse grids is advanced using larger time steps.

Transfer of information between grids in a nested simulation can either be one-way, whereby child grids receive information from their parents, but the parent grids are not in turn affected by information on child grids, or it can be two-way, whereby parent grids are directly affected by the information on the child grids. One-way nesting is probably the most common nesting strategy in ocean modeling and relies on the assumption that the solution on the refined grid likely does not affect the large-scale motions on the coarse grid. In two-way nesting, this assumption does not hold and information from the fine grid is transferred back onto the coarse grid as the simulation progresses. Figure 4 depicts a schematic of a one-dimensional simulation using one- and two-way nesting and two levels of refinement. The primary difference between one-way and two-way nesting is that after the simulation on the fine grid is updated to yield its solution at time t + Δt (Step 2a in Figure 4a), the solution progresses independently on the coarse grid for one-way nesting, while in two-way nesting, the solution on the fine grid is transferred back onto the coarse grid at the beginning of each coarse-grid time step (Steps 4b and 8b in Figure 4b). In addition to ease of implementation, the advantage of one-way nesting is that data at all time steps for the coarse-grid solution can be computed first, and the fine-grid solution can be obtained later, because its solution does not affect the coarse-grid solution. For example, predictions from global-scale tidal models are often used to drive the boundareis of smaller-scale coastal models.The idea is that the tides drive the motions in the smaller-scale simulation, but the smaller-scale simulation does not alter the tides enough to justify feeding the small-scale solution back into the tidal model.

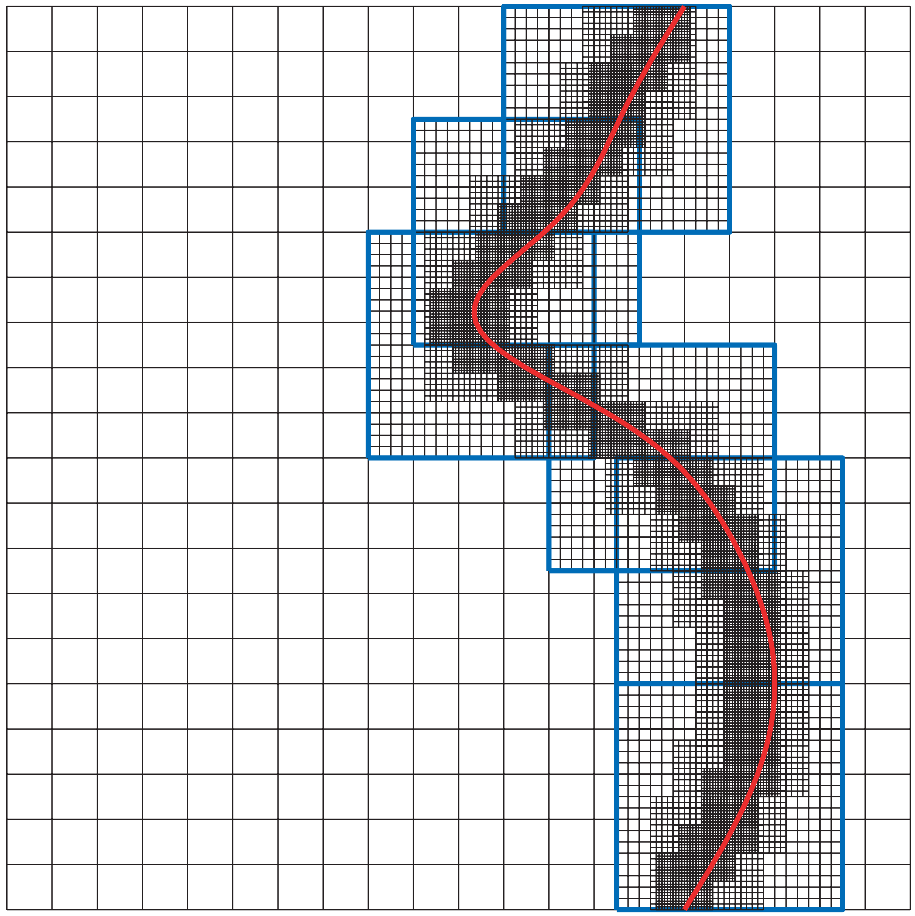


Figure 3. Example of a nested set of grids that allows for high resolution in the vicinity of the topography (red line) while employing coarse resolution further away from it. This grid employs four levels of refinement in which each refined level has twice the resolution of its parent, so that the finest grids are 16 times as resolved as the main coarse grid. The largest parent grid has seven child grids (outlined in blue) and 52 total offspring to make for a total of 21,200 grid cells. A grid the size of the largest grid with the same resolution as the finest grid over the entire domain would require 102,400 cells.

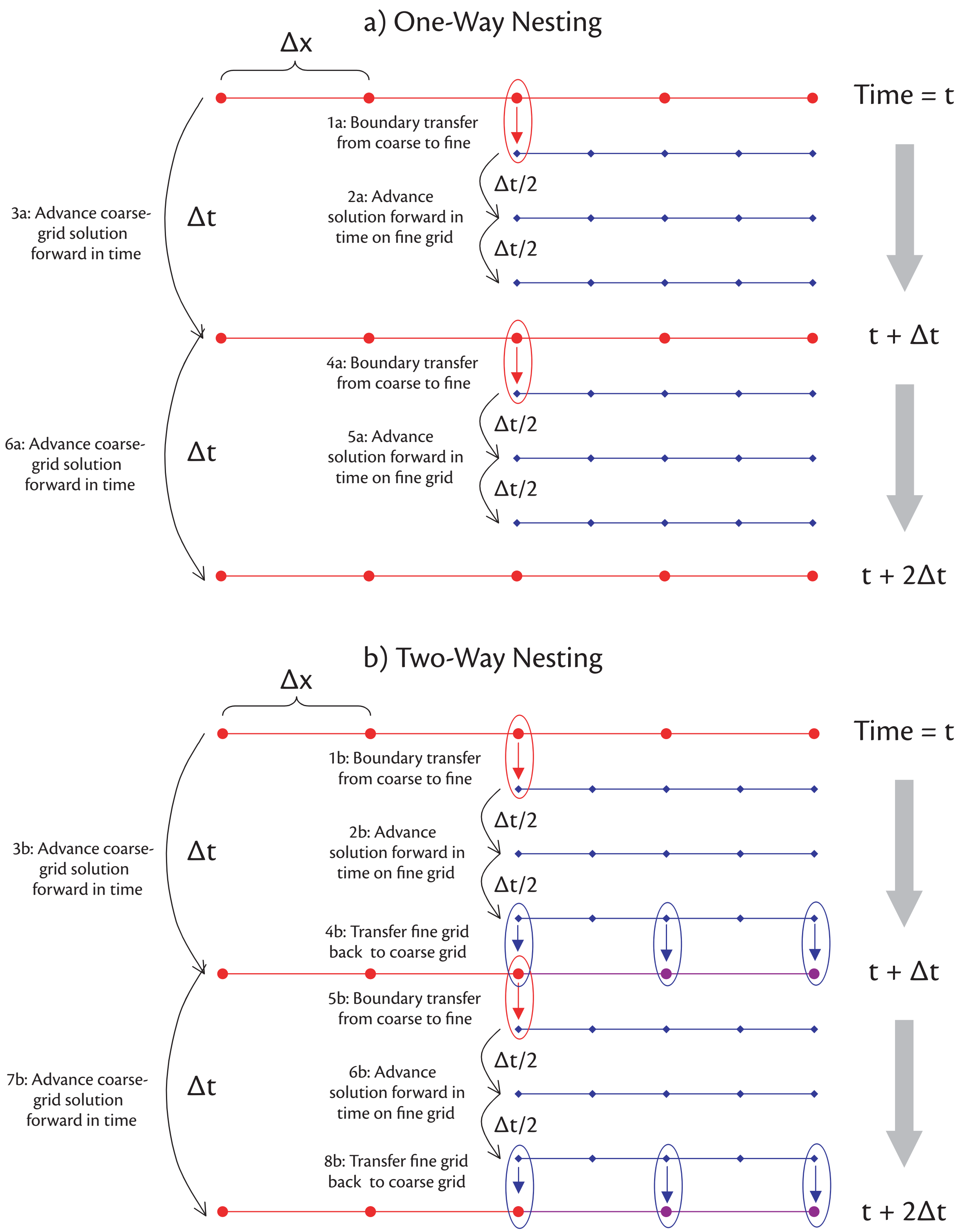


Figure 4. Depiction of a one-dimensional simulation involving one-way nesting (a) and two-way nesting (b). For one-way nesting, the procedure to obtain the solution at time t+2Δt proceeds as follows: (1a) Information from the simulation at time t on the coarse grid is transferred to the boundary of the fine grid. (2a) The solution then progresses on the fine grid over two time steps with time step size Δt/2. (3a) The solution on the coarse grid is advanced forward in time one time step. (4a) This solution is then transferred again to the boundary of the fine grid. The procedure then repeats itself in steps 5a and 6a. For two-way nesting, the procedure is altered by transferring the information from the fine grid onto the coarse grid in steps 4b and 8b.

The schematic in Figure 4 implies that the information is transferred back onto the coarse grid at the beginning of each coarse-grid time step in a one-to-one direct transfer. In practice, the fine-grid solution is filtered before it is transferred onto the coarse grid, or the transfer may include an average of the fine- and coarse-grid solutions, or it may be in the form of nudging, in which the solution on the coarse grid progresses as if it were being forced to have a tendency to follow the fine-grid solution. These represent a small subset of strategies designed to account for differences in the simulation between the fine- and coarse-grid solutions that would otherwise result in abrupt changes in the solution on the coarse grid if one-to-one transfers were employed. The alternative is to employ an iterative procedure at each time step in which the transfer is adjusted until the solution at the boundary of the fine grid matches the solution at the boundary of the coarse grid. This is not generally required unless the solutions on the two grids are tightly coupled, which is the case when detailed flow physics crosses the boundary between nested grids. For this reason, it is often necessary to place child grid boundaries in regions of the flow where it is unlikely that complex flow physics may cross over an inter-grid boundary.

SUNTANS+ROMS: A NEW HYBRID NESTED MODELING TOOL

The primary advantage of model nesting is that it allows a simulation to capture a broad range of length scales while enabling the simulation of the small-scale physics in the flow to advance forward in time with smaller time steps than the simulation of the larger-scale physics on the coarser grids. The implication behind different grids capturing different spatio-temporal length scales is that they also capture different flow physics. Therefore, an ideal simulation would not only employ different time step sizes on different grids but also different approximations based on the particular physics being resolved by the scales on each grid. This is the idea behind nested hybrid simulations, which employ different simulation codes on the different grids for an optimal multi-scale, multi-physics calculation.

An example of a hybrid coupling strategy is the technique of including the effect of wind-waves on coastal circulation models. Because wind-wave spatio-temporal scales are much too short to include in reasonably sized coastal circulation simulations, the strategy is to compute the effects of wind-waves using two-dimensional wave models that return the wave-induced current, which can then be included in the circulation model. In a one-way nested approach, the wave models drive the currents, which in turn do not affect the surface waves. In the two-way nested approach, the currents from the circulation model are fed back into the wave model so that it calculates the wave field that has been altered by the currents. Because the time scale of variation of the wave-induced currents is typically much slower than the smallest temporal scale in the circulation model, the models are coupled over a longer time scale, which involves on the order of tens to hundreds of circulation-model time steps. Among the most popular implementations of a coupled wave-current hybrid model is the incorporation of SWAN into DELFT3D.

Nesting of wave models with circulation models is an example of a hybrid tool that couples a two-dimensional model with a three-dimensional circulation model. Recent work in coastal ocean simulations has focused on the development of coupled three-dimensional multiphysics codes, where the larger-scale codes compute the regional-scale, predominantly hydrostatic circulation, and the smaller-scale codes compute the littoral, coastal-scale, nonhydrostatic circulation. Among these hybrid codes is the nested simulation tool that combines ROMS (Shchepetkin and McWilliams, 2005) and SUNTANS (Fringer et al., submitted). ROMS simulates the regional-scale flow on a curvilinear grid. It employs state-of-the art numerical methods to discretize the equations of ocean circulation and builds upon decades of research in the large-scale ocean modeling community to develop a simulation tool to simulate virtually dissipation-free flows in the open ocean. SUNTANS, on the other hand, simulates small-scale nonhydrostatic physics that are based on a discretization of the Navier-Stokes equations in the absence of physical approximations. Like the nonhydrostatic MITgcm, SUNTANS is written in parallel using MPI (message-passing interface) and is designed for high-performance computers; however, unlike MITgcm, which incorporates curvilinear grids, SUNTANS incorporates unstructured grids in a manner similar to unTRIM, thereby enabling it to resolve a disparate range of length scales within the same grid.

The idea behind the ROMS-SUN-

TANS coupling tool is that the large-scale flow field is well represented on curvilinear grids with relatively smooth topography, while the small-scale flow field is dominated by complex topography near the coast and hence it requires an unstructured grid. Additionally, energy and momentum conservation are much more important for the larger-scale flow; the discretization in ROMS is able to capture this conservation on the curvilinear coordinate grid. In particular, ROMS captures the fine balance between Coriolis acceleration and pressure gradients, which dominates the energetics for large-scale flows. These characteristics are not as important for fine-scale flow calculations, particularly near coastal features, because physically there is more dissipation and mixing for the small-scale physics captured by SUNTANS.

In both the coastal and open ocean, and hence for both ROMS and SUNTANS, fast free-surface gravity waves represent the shortest time-scale physics in each simulation tool. However, for the coastal ocean, only the tidal and regional time scales associated with the free surface are of interest. Therefore, the algorithms in ROMS and SUNTANS are designed to allow for larger time step sizes without incurring numerical instabilities that might arise when such large time steps are used to advance short-time-scale physics. In ROMS, this problem is handled with a technique called mode-splitting, whereby the free-surface waves, which represent the fast, or external-mode physics, are advanced forward in time with time steps that are typically 100 times smaller than those for the other, longer time scale, or slow, internal-mode physics. In SUNTANS, the fast free-surface waves are computed by employing an implicit technique as used in TRIM by Casulli (1999) to advance the equations of motion forward in time, thereby removing the time-step limit on the free-surface motions.

Among the unique challenges in developing the ROMS-SUNTANS hybrid tool is that, although the nature of the one- and two-way coupling strategy is similar to that depicted in Figure 4, the information transfer between the two simulation codes is complicated by the fact that ROMS employs curvilinear-coordinate grids, while SUNTANS employs unstructured grids. Complex interpolation procedures are thus required to transfer information back and forth between the grids both in the horizontal as well as the vertical, because ROMS employs curvilinear coordinates in the vertical while SUNTANS employs z-level grids (Figure 5). In this coupled schematic, the coupling of ROMS and SUNTANS grids occurs in the horizontal, with both grids extending throughout the water column and exchanging information only across vertical faces. Although this type of exchange is the initial focus of the coupling tool, the long-term goal will allow for SUNTANS grids to be nested in the vertical; fine-scale SUNTANS grids will resolve processes in the mixed layer while allowing ROMS to compute the larger-scale features beneath it. In this case, information will be exchanged across both vertical and horizontal grid faces.

Further complicating the nesting of SUNTANS into ROMS is the fact that both simulation codes are written for parallel, high-performance computers. Of critical importance to optimal parallel computations is the notion of load balancing, which implies that a group of computers involved in a parallel calculation are all performing the same amount of work so that one computer is not performing calculations while the others stand idle. Both ROMS and SUNTANS are optimally load-balanced on their own, and therefore load balancing is straightforward to achieve using static mesh refinement. However, the creation of new adaptive SUNTANS (or ROMS) grids over the course of the simulation requires a rebalancing of the workload as the simulation progresses. This rebalancing is accomplished by redistributing all of the grids involved in the simulation over the available computers each time a new SUNTANS or ROMS grid is adaptively added to the simulation.

It is clear that the coupling of two vastly different multi-physics simulation tools like ROMS and SUNTANS is quite a complex task. Fortunately, tools exist that are specifically designed to couple several multi-physics tools under one common framework, thus allowing the implementer to focus on the physics rather than the computer science of model coupling. Among the most advanced of these tools is the Earth System Modeling Framework (ESMF[9]), which has been employed to couple over a dozen multi-physics codes to form the NASA GEOS-5 Atmospheric General Circulation Model, and is similar to the PRISM[10] project. For oceanic applications, the

[9] Earth System Modeling Framework (ESMF, http://www.esmf.ucar.edu)

[10] Programme for Integrated Earth System Modeling (PRISM, http://www.prism.enes.org)

Model Coupling Environment Library (MCEL[11]) has been employed to couple wave models with three-dimensional circulation models. The ROMS-SUNTANS coupling is handled by a combination of ROMS-SUNTANS specific tools and those available in the Model Coupling Toolkit (MCT[12]), which handles a host of the parallel intercommunication and load-balancing features essential to the ROMS-SUNTANS coupled tool.

PLANNED APPLICATIONS OF THE ROMS-SUNTANS HYBRID TOOL

Both SUNTANS and ROMS have been applied extensively to the west coast of the United States. ROMS has been employed to study the regional circulation, with a particular emphasis on the California Coastal Current and the dynamics of upwelling fronts along the California coast, as depicted in Figure 6a. SUNTANS, on the other hand, has been applied to the much smaller domain of Monterey Bay to study the interaction of the tides with the bottom topography and the subsequent generation of internal waves of tidal frequency, or internal tides, as depicted in Figure 6b. These fine-scale simulations of Monterey using SUNTANS are being coupled with the larger-scale simulations of ROMS using the ROMS-SUNTANS hybrid tool. In the two-way nested simulations, the primary effect of the coupling is that the ROMS mesoscale currents interact with the internal tides and other tidal time-scale motions in SUNTANS and have a strong effect on the internal tidal wave field, which in turn affects the mixing and dissipation within coastal regions such as Monterey Bay. This information is fed back into the ROMS simulations to account for the effect of the tidal-scale motions on the mesoscale dynamics, in particular how they affect the dynamics of the upwelling fronts and their interaction with the California Current System. The

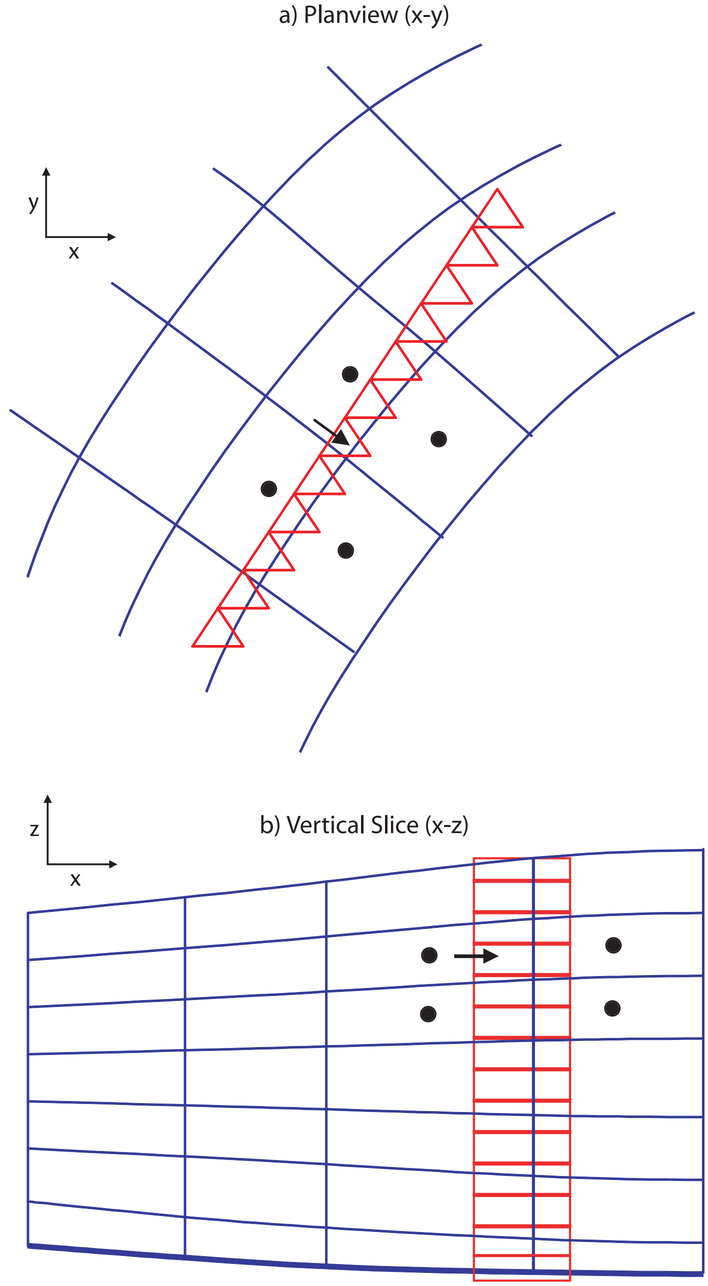


Figure 5. Planview (a) and vertical slice (b) of a ROMS-SUNTANS intergrid boundary, showing the four points on the ROMS grid (represented by the dark circles) that are used to interpolate to obtain the velocity (black arrow) at a cell face on the SUNTANS grid. The red lines represent the unstructured, z-level SUNTANS grid, while the blue lines represent the curvilinear-coordinate, bottom-following ROMS grid.

[11] Model Coupling Environment Libarary (MCEL, http://www.navo.hpc.mil/Navigator/fall02_Feature2.html)

[12] Model Coupling Toolkit (http://www-unix.mcs.anl.gov/mct)

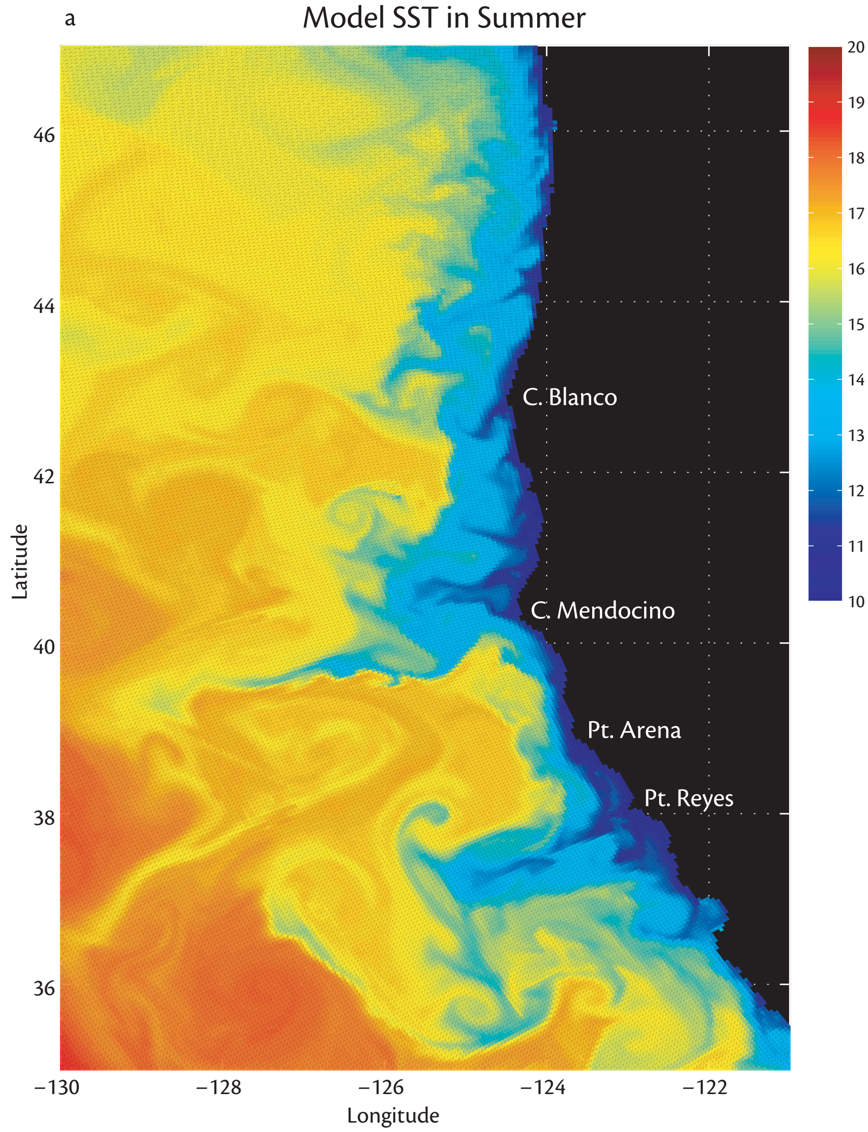
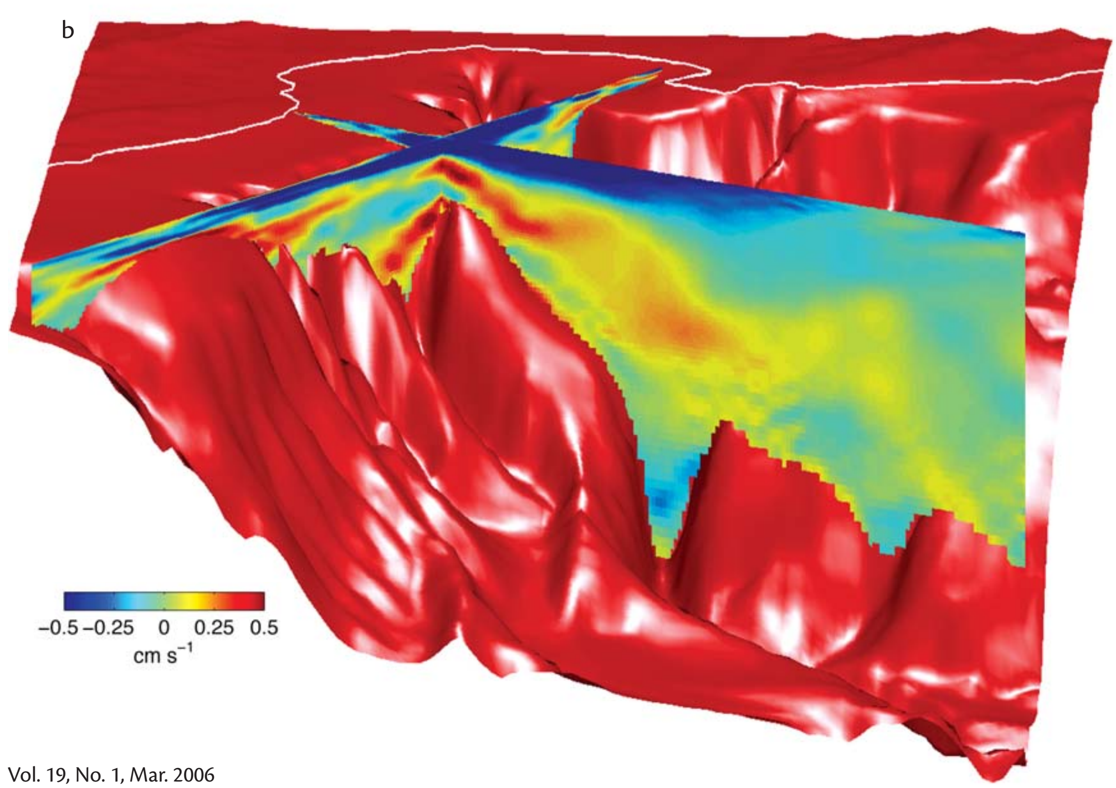


Figure 6. Example ROMS (a) and SUNTANS (b) simulations. The ROMS simulation represents the sea-surface temperature (in °C) during summer on the west coast of the United States, showing how the cold, upwelled water (in blue) interacts with the California Coastal Current to eject cold fluid out into the deep ocean in the form of so-called jets and squirts. The SUNTANS result depicts two cross sections through the complex Monterey Bay bathymetry (in red) that show the east-west velocity field induced by internal waves generated as a result of the interaction of the tidal currents with the bathymetry. These internal waves are known to be significant sources of dissipation and mixing in the coastal ocean. The white line on the red topography represents the Monterey Bay coastline. The time step in the SUNTANS simulation is 28 seconds and the average grid spacing is 250 m, while the time step in the ROMS simulation is 1 hour (30 seconds for the surface waves) and the average grid spacing is 3.5 km. Monterey Bay is located where the 122° E longitude intersects the coastline in the ROMS simulation.



feedback of the SUNTANS simulations into the ROMS simulations effectively acts as an advanced subgrid-scale model for the ROMS simulations. That is, because the grids in ROMS do not resolve subtidal-scale motions, the feedback of SUNTANS information into ROMS accounts for information not resolved by the ROMS grids. ROMS computes mesoscale dynamics, and SUNTANS computes the effect of the submesoscale dynamics on the ROMS simulations.

The overarching goals of these simulations and future simulations using the hybrid tool are two-fold. The first and most obvious goal is to enable the simulation of multi-scale problems in the coastal ocean in a way that very-high-resolution physics can be computed while obtaining an accurate representation of the large-scale flow features and their effects on the smaller, fine-scale flow features of interest in SUNTANS. The second goal is to understand the effect of submesoscale dynamics on mesoscale features to ultimately devise submesoscale parameterizations in ROMS that can accurately predict the effects of the unresolved features of the flow on the large-scale, resolved features of the flow without actually having to compute them. This submesocale parameterization enables accurate predictions of mesoscale currents over much longer time scales that would otherwise not be feasible with a hybrid tool because of the computational expense associated with computing a much wider range of scales of motion.

DISCUSSION

We have presented an overview of a hybrid approach in which two models that focus on distinct spatio-temporal scales are coupled to form a simulation tool that captures a larger range of scales: the coupling of ROMS (Shchepetkin and McWilliams, 2005) and SUNTANS (Fringer et al., submitted). Accurate simulation of problems involving a broad range of length scales is possible because most multi-scale flows of interest are well represented by the larger-scale motions, while a smaller percentage of the flow contains small-scale physics that require computationally expensive high resolution. ROMS simulates the regional-scale flow on a curvilinear grid, while SUNTANS simulates the small-scale physics on an unstructured grid.

Although this hybrid approach provides a tool that captures a broad range of scales, from the mesoscale down to tens of meters, it still captures a limited portion of the overall spectrum of fluid motions in the ocean. That is, mesoscale simulations still rely on boundary conditions taken from larger-scale global ocean circulation simulations, and the coastal simulations still require turbulence models and forcing with wind-wave models to account for unresolved motions not captured by the SUNTANS grids. And, as with all well-resolved simulations of the ocean, the fidelity of the tool is still highly dependent on the initial and boundary conditions that are obtained from oceanographic field measurements. Nevertheless, the hybrid tool brings us one step closer to coupling scales of motion that would otherwise not be feasible with one single code.

Development of this hybrid tool is an ongoing project (e.g., Kang et al., 2006). Applications of this system and completion of the two-way coupling are currently underway. For more information the reader is encouraged to visit http://suntans.stanford.edu/roms-suntans.

ACKNOWLEDGMENTS

Ongoing development of the SUNTANS-ROMS nested simulation tool is supported by the Office of Naval Research under grants N00014-05-1-0293 and N00014-05-1-0294 (Scientific officers: Dr. C. Linwood Vincent, Dr. Terri Paluszkiewicz, and Dr. Scott Harper). The authors have written this article on behalf of the project team at the University of California-Los Angeles (UCLA) and Stanford University, whose other members are Alexander Shchepetkin and Yuliya Kanarska (UCLA), and Margot Gerritsen, Ed Gross, Dujuan Kang, and Zhonghua Zhang (Stanford University). The comments of three anonymous reviewers who provided invaluable feedback for this article are greatly appreciated.

REFERENCES

Casulli, V. 1999. A semi-implicit finite difference method for non-hydrostatic, free-surface flows. *International Journal for Numerical Methods in Fluids* 30(4):425–440.

Casulli, V., and P. Zanolli. 2002. Semi-implicit numerical modeling of non-hydrostatic free-surface flows for environmental problems. *Mathematical and Computer Modeling* 36:1,131–1,149.

Fringer, O.B., M. Gerritsen, and R.L. Street. Submitted. An unstructured-grid, finite-volume, nonhydrostatic, parallel coastal ocean simulator. *Ocean Modelling*.

Kang, D., O.B. Fringer, M. Gerritsen, Y. Kanarska, J.C. McWilliams, A.F. Shchepetkin, and R.L. Street. 2006. High-resolution simulations of upwelling and internal waves in Monterey Bay using a coupled ROMS-SUNTANS multi-scale simulation tool. *Eos Transactions of the American Geophysical Union 87*(36-Suppl.):Abstract OS36G-07.

Shchepetkin, A.F., and J.C. McWilliams. 2005. The Regional Oceanic Modeling System: A split-explicit, free-surface, topography-following-coordinate ocean model. *Ocean Modelling* 9:347–404.