Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 227 (2008) 3595–3624
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the Regional Ocean Modeling System

D.B. Haidvogel [a,*], H. Arango [a], W.P. Budgell [b], B.D. Cornuelle [c], E. Curchitser [a], E. Di Lorenzo [d], K. Fennel [e], W.R. Geyer [f], A.J. Hermann [g], L. Lanerolle [h], J. Levin [a], J.C. McWilliams [i], A.J. Miller [c], A.M. Moore [j], T.M. Powell [k], A.F. Shchepetkin [i], C.R. Sherwood [l], R.P. Signell [l], J.C. Warner [l], J. Wilkin [a]

[a] *Institute for Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, United States*
[b] *Institute of Marine Research, Bergen, Norway*
[c] *Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, United States*
[d] *Georgia Institute of Technology, Atlanta, GA, United States*
[e] *Dalhousie University, Halifax, Nova Scotia, Canada*
[f] *Woods Hole Oceanographic Institution, Woods Hole, MA, United States*
[g] *Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, WA, United States*
[h] *National Oceanic and Atmospheric Administration, Silver Spring, MD, United States*
[i] *University of California, Los Angeles, Los Angeles, CA, United States*
[j] *University of California, Santa Cruz, Santa, United States*
[k] *University of California, Berkeley, Berkeley, CA, United States*
[l] *US Geological Survey, Woods Hole, MA, United States*

Received 12 December 2006; received in revised form 3 June 2007; accepted 7 June 2007
Available online 29 June 2007

Abstract

Systematic improvements in algorithmic design of regional ocean circulation models have led to significant enhancement in simulation ability across a wide range of space/time scales and marine system types. As an example, we briefly review the Regional Ocean Modeling System, a member of a general class of three-dimensional, free-surface, terrain-following numerical models. Noteworthy characteristics of the ROMS computational kernel include: consistent temporal averaging of the barotropic mode to guarantee both exact conservation and constancy preservation properties for tracers; redefined barotropic pressure-gradient terms to account for local variations in the density field; vertical interpolation performed using conservative parabolic splines; and higher-order, quasi-monotone advection algorithms. Examples of quantitative skill assessment are shown for a tidally driven estuary, an ice-covered high-latitude sea, a wind- and buoyancy-forced continental shelf, and a mid-latitude ocean basin. The combination of moderate-order spatial approximations, enhanced conservation properties, and quasi-monotone advection produces both more robust and accurate, and less diffusive, solutions than those produced in earlier terrain-following ocean models. Together with advanced methods of data

* Corresponding author. Tel.: +1 732 932 6555x256; fax: +1 732 932 8578.
E-mail address: dale@imcs.rutgers.edu (D.B. Haidvogel).

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2007.06.016

assimilation and novel observing system technologies, these capabilities constitute the necessary ingredients for multi-purpose regional ocean prediction systems.

Keywords: Regional ocean prediction; Incompressible Navier–Stokes equations; Split-explicit time stepping; Sea ice modeling; Biogeochemical cycles

1. Introduction

The modeling of variability within the marine environment is essential for societal concerns that span the entire space/time spectrum. Examples include the prediction of estuarine and coastal ocean processes (circulation, turbulent mixing, and the transport of sediment) on time scales of minutes to days, studies of the oceanic response to global climate change (sea ice growth/retreat, sea level rise, carbon cycling) on interannual to decadal time scales, as well as numerous associated issues related to navigation, fisheries, security and health. Although modeling systems designed to deal with these problems will inevitably take a variety of forms, the development and demonstration of ocean modeling systems capable of broad geographic application is desirable.

Any broadly portable marine modeling system will require significant dynamical complexity, however. In addition to the underlying hydro-dynamical engine – responsible for determining sea level height, and the three-dimensional circulation and transport of momentum, temperature and salt – application-specific submodels are required for, *e.g.*, turbulent mixing; sediment resuspension, transport and deposition; sea ice thermo/dynamics; and *in situ* biogeochemical responses. These requirements in turn constrain the numerical solution procedures to be highly flexible, robust and computationally efficient.

Historically, ocean circulation models were developed for particular classes of applications (climate modeling, coastal prediction, *etc.*) and made specific choices as deemed best (at that time) for such considerations as vertical coordinate treatment and horizontal grid representation [1]. Recently, several of these classes of ocean models have been enriched with more general algorithmic options to allow a broader range of effective application. We review one such model here.

The Regional Ocean Modeling System (ROMS) benefits from a lengthy history of development within the terrain-following ocean modeling community (*e.g.* [2–4]). ROMS is maintained and applied by a community of users, now numbering in the many hundreds, who in turn have supplied key enhancements in the dynamical, geochemical and algorithmic areas mentioned further below. The result of this synergistic interaction has been the emergence of a truly multi-purpose marine modeling system having demonstrable skill across a variety of space/time scales and problem classes.

The purpose of this review is threefold. First, we provide a description of the formulation of the ROMS dynamical core and several of its important component sub-models. Second, we review the combination of novel algorithmic treatments that together allow effective application of ROMS across a variety of geographical and dynamical environments. Lastly, we note recent examples of quantitative skill assessment in four distinctive oceanic settings. Further information is available in the cited literature, as well as at the ROMS community web site [5].

2. The ROMS dynamical core, turbulence closures and selected sub-models

2.1. Hydrodynamic core

ROMS is a member of a general class of three-dimensional, free-surface, terrain-following numerical models that solve the Reynolds-averaged Navier–Stokes equations using the hydrostatic and Boussinesq assumptions [6,7]. The governing dynamical equations – in flux form, Cartesian horizontal coordinates and sigma vertical coordinates – take the traditional form:

$$\frac{\partial(H_z u)}{\partial t}+\frac{\partial(uH_z u)}{\partial x}+\frac{\partial(vH_z u)}{\partial y}+\frac{\partial(\Omega H_z u)}{\partial s}-fH_z v=-\frac{H_z}{\rho_0}\frac{\partial p}{\partial x}-H_z g\frac{\partial \zeta}{\partial x}-\frac{\partial}{\partial s}\left(\overline{u'w'}-\frac{v}{H_z}\frac{\partial u}{\partial s}\right), \tag{1}$$

$$\frac{\partial(H_z v)}{\partial t}+\frac{\partial(uH_z v)}{\partial x}+\frac{\partial(vH_z v)}{\partial y}+\frac{\partial(\Omega H_z v)}{\partial s}+fH_z u=-\frac{H_z}{\rho_0}\frac{\partial p}{\partial y}-H_z g\frac{\partial \zeta}{\partial y}-\frac{\partial}{\partial s}\left(\overline{v'w'}-\frac{v}{H_z}\frac{\partial v}{\partial s}\right), \tag{2}$$

$$0=-\frac{1}{\rho_0}\frac{\partial p}{\partial s}-\frac{g}{\rho_0}H_z\rho, \tag{3}$$

with the continuity equation:

$$\frac{\partial \zeta}{\partial t}+\frac{\partial(H_z u)}{\partial x}+\frac{\partial(H_z v)}{\partial y}+\frac{\partial(H_z \Omega)}{\partial s}=0 \tag{4}$$

and scalar transport given by:

$$\frac{\partial(H_z C)}{\partial t}+\frac{\partial(uH_z C)}{\partial x}+\frac{\partial(vH_z C)}{\partial y}+\frac{\partial(\Omega H_z C)}{\partial s}=-\frac{\partial}{\partial s}\left(\overline{c'w'}-\frac{v}{H_z}\frac{\partial C}{\partial s}\right)+C_{\text{source}}. \tag{5}$$

Here u, v, and Ω are the components of velocity in the horizontal (x and y) and vertical (scaled sigma coordinate, s) directions respectively; ζ is the wave-averaged free-surface elevation; h is the depth of the sea floor below mean sea level; H_z is a vertical stretching factor; and f is the Coriolis parameter. An over-bar represents a time average, and a prime ($'$) represents turbulent fluctuations. Pressure is p; ρ and ρ_0 are total and reference densities; g is the acceleration due to gravity; v and v_θ are molecular viscosity and diffusivity; C represents a tracer quantity (for example, salt, temperature, and suspended-sediment); C_{source} are tracer source/sink terms. Finally, a function $[\rho = f(C,p)]$ is required to specify the equation of state.

These equations are closed by parameterizing the Reynolds stresses and turbulent tracer fluxes as:

$$\overline{u'w'}=-K_{\text{M}}\frac{\partial u}{\partial z};\quad \overline{v'w'}=-K_{\text{M}}\frac{\partial v}{\partial z};\quad \overline{c'w'}=-K_{\text{H}}\frac{\partial \rho}{\partial z}; \tag{6}$$

where K_{M} is the eddy viscosity for momentum and K_{H} is the eddy diffusivity for tracers. This results in the standard (harmonic) form for the vertical viscous/diffusive terms. Analogous operators may be added to represent sub-gridscale mixing in the horizontal, although they will not be used in many of the applications reviewed below (see Section 3.7). The horizontal Cartesian coordinates (x,y) may also be replaced by a more general curvilinear coordinate (ξ,η), in which case additional metric terms appear in these equations (see [6]).

2.2. Turbulence closures

In typical regional applications, the ROMS equation set is solved on a computational grid whose spacing is too large to adequately resolve small-scale turbulent processes at the dissipation level. Therefore these sub-gridscale processes for vertical mixing of momentum (eddy viscosity K_{M}) and mass (eddy diffusivity K_{H}) must be parameterized using a turbulence closure model. ROMS provides five methods for turbulence closure: (i) by user-defined analytical expressions for K_{H} and K_{M}; (ii) Brunt-Vaisala frequency mixing, in which the level of mixing is determined based upon the stability frequency; and by (iii) the K-profile parameterization, (iv) the Mellor–Yamada Level 2.5, and (v) the Generic Length Scale methods. The *K-profile parameterization* (KPP) is based on Monin–Obukov similarity theory and provides an estimate of the vertical mixing with both local and non-local sources. The closure scheme is based on the boundary-layer formulation presented in [8], and has been expanded to include both surface and bottom boundary layers in [9]. The *Mellor–Yamada level 2.5* (MY25) scheme is one of a hierarchy of closures proposed in [10]. The method is a two-equation model in which the turbulent kinetic energy and length scale are used to quantify mixing rates. The implementation of the method in ROMS is also described in [9]. Finally, the *Generic Length Scale* (GLS) approach [11] is a generalized two-equation model that takes advantage of the similarities among several two-equation turbulence closure models that have been used for turbulence modeling. These include the Mellor–Yamada Level 2.5, the k–ε model [12], and the k–ω (where omega is a turbulence frequency). The GLS formulation implemented in ROMS is described in [13].

2.3. Bottom boundary layer (BBL) dynamics

Bottom boundary layer dynamics determine the stress exerted on the flow by the bottom, which enters the Reynolds-averaged Navier–Stokes equations as boundary conditions for momentum. Likewise, treatment of the BBL is important for the transport of sediment, because bottom stress determines the transport rate for bedload and the resuspension rate for suspended sediment. ROMS implements either of two sub-models for representing BBL processes: (i) simple drag-coefficient expressions, or (ii) more complex formulations that represent the interactions of wave and currents over a moveable bed. The *drag-coefficient methods* implement formulae for linear bottom friction, quadratic bottom friction, or a logarithmic profile. The more complex methods implement a few of the many *wave-current BBL models*, and couple them with calculations of bottom roughness based on bed properties. In particular, ROMS offers three alternative subroutines that implement somewhat different combinations of algorithms for the wave-current interactions and moveable bed roughness. One is based on the wave-current algorithm of [14] and the ripple geometry and moveable bed roughness of [15]. The second subroutine uses the efficient wave-current BBL computations developed by [16] in combination with sediment and bedform roughness estimates of [17–19]. The third subroutine implements either the wave-current BBL model of [20] or that of [14] along with moveable bed routines proposed by [21].

2.4. Ecosystem modules

ROMS offers a range of ecosystem sub-models. In order of increasing ecological complexity these include an NPZD-type model, a Fasham-type model, a two-phytoplankton-class model, and a multiple-phytoplankton-class model. The *NPZD-type model* has four state variables representing a limiting nutrient (most commonly inorganic nitrogen), phytoplankton, zooplankton and detritus, all of which are measured in units of the limiting nutrient. The NPZD version available in ROMS is coded based on [22]. The *Fasham-type model* is a modified version of Fasham's model [23]. In the ROMS implementation of this model the inorganic nitrogen species nitrate and ammonium are treated as separate state variables, chlorophyll is included as a prognostic variable in addition to phytoplankton biomass, and two size-classes of detritus are distinguished to allow for different settling rates [24]. The *two-phytoplankton-class model* differs from the Fasham-type model by the presence of two phytoplankton groups, one representing larger diatomaceous phytoplankton and one representing small phytoplankton, and the inclusion of silicic acid as an additional nutrient. A further difference lies in the explicit characterization of internal nitrogen, carbon and, in the case of diatoms and large detritus, silica concentrations within the plankton and detrital pools allowing time-evolving variability in the elemental composition of different functional groups. This model is described in detail in [25]. The most complex of the ROMS suite of ecosystem models is the *multiple-phytoplankton-class ECOSIM model* [26]. It explicitly represents four groups of phytoplankton including their internal carbon and nitrogen concentrations and dissolved organic matter; however, it does not explicitly include zooplankton. All of these models focus on the lower trophic levels of the ecosystem which are assumed to be described reasonably well by Eulerian concentrations. Characteristics such as individual behavior and the capacity of some organisms to move against the currents are assumed to be negligible in these models. Individual-based models that include such behaviors are under development [27].

2.5. Sea ice

The sea ice component of ROMS is a combination of the elastic–viscous-plastic rheology and simple one-layer ice and snow thermodynamics with a molecular sublayer under the ice. It is tightly coupled, having the same grid (Arakawa-C) and timestep as the ocean and sharing the same parallel coding structure for use with MPI or OpenMP. A description of the implementation for the Barents and Norwegian Seas is given in [28]. The ice dynamics are based upon an elastic–viscous-plastic (EVP) rheology after [29,30]. The EVP scheme is based on a time-splitting approach whereby short elastic time steps are used to regularize the solution when the ice exhibits nearly rigid behavior. Because the time discretization uses explicit time-stepping, the ice dynamics are readily parallelizable and thus computationally efficient. Employing linearization of viscosities about ice velocities at every elastic (short) time step, as recommended by [30], has the desirable property of

maintaining the ice internal stress state on or in the plastic yield curve. That is, the ice deforms as a plastic material unless it is in a rigid state. Another desirable property of this linearization is that the EVP ice dynamics are found to provide a good transient response to rapidly varying winds as well as to inertial and tidal dynamics, particularly in the marginal ice zone. The ice thermodynamics are based on those of [31,32]. Two ice layers and a single snow layer are used in solving the heat conduction equation. The snow layer possesses no heat content, but is, in effect, an insulating layer. Surface melt ponds are included in the ice thermodynamics. A molecular sub-layer [31] separates the bottom of the ice cover from the upper ocean. The inclusion of the molecular sub-layer was found to produce much more realistic freezing and melting rates than if the ice-ocean heat flux is based purely on the ice bottom to upper ocean temperature difference. Ice thickness distribution is improved by including the transport of ice enthalpy, or internal ice heat content, in determining the internal ice temperature.

3. Discretization and numerical algorithms

The ROMS computational kernel utilizes consistent temporal averaging of the barotropic mode to guarantee both exact conservation and constancy preservation properties for tracers and therefore more accurately resolves barotropic processes, while preventing aliasing of unresolved barotropic signals into the slow baroclinic motions (see Sections 3.3 and 3.4). Accuracy of the mode-splitting is further enhanced with redefined barotropic pressure-gradient terms to account for the local variations in the density field (*i.e.*, the pressure-gradient truncation error that has previously plagued terrain-following coordinate models is greatly reduced) without sacrificing the efficiency of the split-explicit formulation (Section 3.5). Vertical interpolation is performed using either centered fourth-order schemes or an interpolation based on conservative parabolic splines. The combination of moderate-order spatial approximations, quasi-monotone advection operators (Section 3.7), and enhanced conservation properties produces both more robust and accurate, and less diffusive, solutions than those produced in earlier terrain-following ocean models.

Shchepetkin and McWilliams describe the algorithms that comprise the ROMS computational kernel in [7,33,34]. The following brief summary is extracted from these earlier works, and the reader is referred to them for the complete details.

3.1. Generalized topography-following coordinate

In a topography-following vertical coordinate system there is a transformation,

$$z = z(x, y, \sigma), \tag{7}$$

where z is the Cartesian height and σ is the vertical distance from the surface measured as the fraction of the local water column thickness; *i.e.*, $-1 \leqslant \sigma \leqslant 0$, where $\sigma = 0$ corresponds to the free surface, $z = \zeta$, and $\sigma = -1$ corresponds to the oceanic bottom, $z = -h(x, y)$. In the case of the classical σ-coordinate, (7) reduces to

$$z = \sigma \cdot h(x, y). \tag{8}$$

This may be combined with nonlinear stretching, $S(\sigma)$,

$$z(x, y, \sigma) = S(\sigma) \cdot h(x, y) \tag{9}$$

and further generalized into the S-coordinate of [4] which in essence behaves like Eq. (8) in shallow regions and Eq. (9) in the deep ocean.

3.2. Spatial discretization

In the horizontal (x, y), the state variables are arranged as shown in Fig. 1; this is equivalent to the well known Arakawa "C" grid. Except where noted otherwise below (*e.g.*, Section 3.7), a centered, second-order finite difference approximation is adopted in the horizontal.

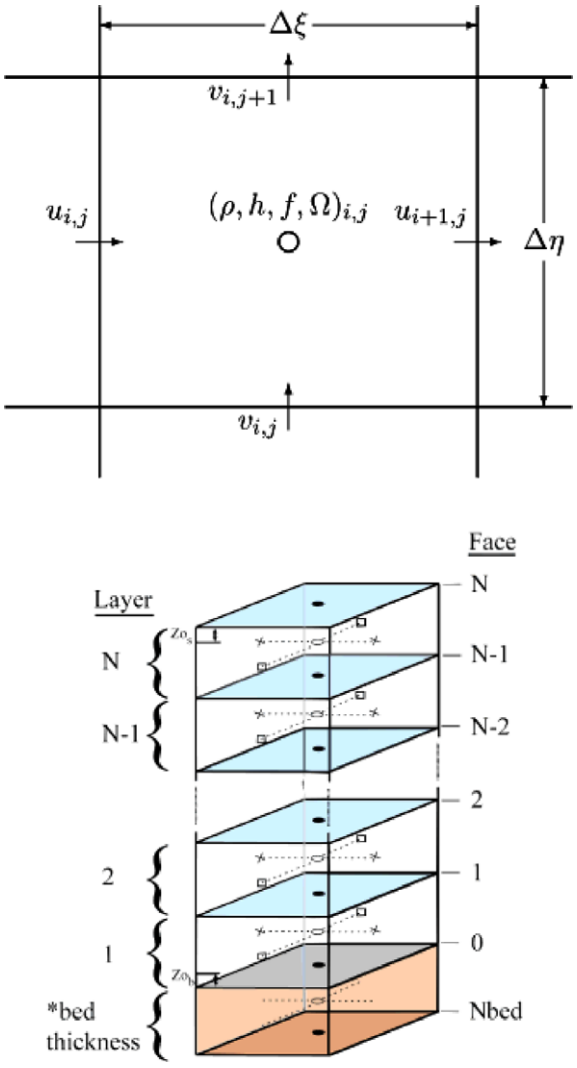

Fig. 1. Upper: horizontal stencil of the Arakawa C grid. Lower: a vertical section of the ROMS grid showing placement of variables.

Discretization of the vertical coordinate introduces a set of coordinate surfaces,

$$\left\{ z_{k+\frac{1}{2}} = z_{k+\frac{1}{2}}(x,y), k = 0, 1, \ldots, N \right\}, \tag{10}$$

where $z_{\frac{1}{2}} \equiv z^{(0)}_{\frac{1}{2}} \equiv -h$ and $z_{N+\frac{1}{2}} \equiv \zeta$. If the ocean is at rest, the free-surface elevation is $\zeta = 0$; hence $z_{N+\frac{1}{2}} = 0$. In the case of a non-zero ζ, all $z_{N+\frac{1}{2}}$ are displaced by a distance proportional to ζ and the distance from the bottom as the fraction of the unperturbed local depth,

$$z_{k+\frac{1}{2}} = z^{(0)}_{k+\frac{1}{2}} + \zeta \left(1 + \frac{z^{(0)}_{k+\frac{1}{2}}}{h} \right). \tag{11}$$

As a result, the perturbed grid-box height $\Delta z_k \equiv z_{k+\frac{1}{2}} - z_{k-\frac{1}{2}}$ is related to the unperturbed height $\Delta z^{(0)}_k \equiv z^{(0)}_{k+\frac{1}{2}} - z^{(0)}_{k-\frac{1}{2}}$ according to

$$\Delta z_k = \Delta z^{(0)}_k \left(1 + \frac{\zeta}{h} \right), \tag{12}$$

where the multiplier $(1+\zeta/h)$ is independent of the vertical coordinate. One consequence of (12) is the fact that vertical mass fluxes generated by a purely barotropic motion vanish identically at every interface, $z_{k+\frac{1}{2}}$.

3.3. Integral conservation and constancy preservation for tracers

Combining the adiabatic tracer equation in advective form,

$$\frac{\partial q}{\partial t}+(\mathbf{u}\cdot\nabla)q=0, \tag{13}$$

with the statement of non-divergence of the vector velocity field,

$$(\nabla\cdot\mathbf{u})=0, \tag{14}$$

the tracer equation in conservation form becomes

$$\frac{\partial q}{\partial t}+\nabla\cdot(\mathbf{u}q)=0. \tag{15}$$

As a consequence of the flux form of the tracer equations, if a tracer is spatially uniform at the initial time, it remains so regardless of the velocity field. On the other hand, as a consequence of (15), the volume integral of the tracer concentration is conserved in the absence of incoming and outgoing fluxes across the domain boundary. The continuity equation (14) provides the compatibility condition between these two properties.

The discretization of (15) may be obtained using a finite-volume discretization,

$$\begin{aligned}\Delta V^{n+1}_{i,j,k}q^{n+1}_{i,j,k}=\Delta V^{n}_{i,j,k}q^{n}_{i,j,k}-\Delta t\Big[&\tilde{q}_{i+\frac{1}{2},j,k}U_{i+\frac{1}{2},j,k}-\tilde{q}_{i-\frac{1}{2},j,k}U_{i-\frac{1}{2},j,k}+\tilde{q}_{i,j+\frac{1}{2},k}V_{i,j+\frac{1}{2},k}-\tilde{q}_{i,j-\frac{1}{2},k}V_{i,j-\frac{1}{2},k}\\&-\tilde{q}_{i,j,k+\frac{1}{2}}W_{i,j,k+\frac{1}{2}}+\tilde{q}_{i,j,k-\frac{1}{2}}W_{i,j,k-\frac{1}{2}}\Big],\end{aligned} \tag{16}$$

where $q_{i,j,k}$ is understood to be a volume-averaged concentration over the grid-box $\Delta V_{i,j,k}$,

$$q_{i,j,k}=\frac{1}{\Delta V_{i,j,k}}\int_{\Delta V^{n}_{i,j,k}} q\,\mathrm{d}V. \tag{17}$$

The $\tilde{q}_{i+\frac{1}{2},j,k}$ are the interfacial values of tracer concentration. $U_{i+\frac{1}{2},j,k}$, $V_{i,j+\frac{1}{2},k}$, and $W_{i,j,k+\frac{1}{2}}$ are volumetric fluxes in the two horizontal and vertical directions. These are defined as velocity components multiplied by the contact area between two adjacent grid boxes,

$$\begin{aligned}U_{i+\frac{1}{2},j,k}&=u_{i+\frac{1}{2},j,k}\Delta z_{i+\frac{1}{2},j,k}\Delta\eta_{i+\frac{1}{2},j},\\V_{i,j+\frac{1}{2},k}&=v_{i,j+\frac{1}{2},k}\Delta z_{i,j+\frac{1}{2},k}\Delta\xi_{i,j+\frac{1}{2}},\end{aligned} \tag{18}$$

where $\Delta z_{i+\frac{1}{2},j,k}$, $\Delta\eta_{i+\frac{1}{2},j}$, and $\Delta z_{i,j+\frac{1}{2},k}$, $\Delta\xi_{i,j+\frac{1}{2}}$ are vertical and horizontal measures of the corresponding grid box interfaces ($\Delta\xi$, $\Delta\eta$ may be non-uniform in the case of curvilinear horizontal coordinates). The superscripts $n+1$ and n denote new and old time levels. The time step for the flux variables in (16) is not specified yet, but must be effectively at $n+1/2$ to achieve second-order temporal accuracy. However, the flux form by itself guarantees exact conservation of the global volume integral of the advected quantity as long as there is no net flux across the domain boundary.

Setting $q_{i,j,k}\equiv 1$ in (16) yields the discretized continuity equation,

$$\Delta V^{n+1}_{i,j,k}=\Delta V^{n}_{i,j,k}-\Delta t\cdot\Big[U_{i+\frac{1}{2},j,k}-U_{i-\frac{1}{2},j,k}+V_{i,j+\frac{1}{2},k}-V_{i,j-\frac{1}{2},k}-W_{i,j,k+\frac{1}{2}}+W_{i,j,k-\frac{1}{2}}\Big]. \tag{19}$$

Once (19) holds, the conservative form of the discrete tracer Eq. (16) also has the property of constancy preservation in addition to global content conservation.

In a hydrostatic model, the discrete continuity Eq. (19) is needed to compute vertical velocity rather than grid-box volume $\Delta V^{n+1}_{i,j,k}$. (The latter is entirely controlled by change of ζ via (12).) Hence,

$$W_{i,j,\frac{1}{2}} = 0, \quad \text{at the sea floor, and} \tag{20}$$

$$W_{i,j,k+\frac{1}{2}} = -\sum_{k'=1}^{k} \left\{ \frac{\Delta V^{n+1}_{i,j,k'} - \Delta V^{n}_{i,j,k'}}{\Delta t} + U_{i+\frac{1}{2},j,k'} - U_{i-\frac{1}{2},j,k'} + V_{i,j+\frac{1}{2},k'} - V_{i,j-\frac{1}{2},k'} \right\} \tag{21}$$

for all $k = 1, 2, \ldots, N$ which defines the meaning of $W_{i,j,k+\frac{1}{2}}$ as a finite-volume flux across the *moving* grid-box interface $z_{i,j,k+\frac{1}{2}}$. Vertical summation of (19) for different k leads to the equation for the free surface,

$$\zeta^{n+1}_{i,j} = \zeta^{n}_{i,j} - \frac{\Delta t}{\Delta A_{i,j}} \left[\overline{U}_{i+\frac{1}{2},j} - \overline{U}_{i-\frac{1}{2},j} + \overline{V}_{i+\frac{1}{2},j} - \overline{V}_{i+\frac{1}{2},j} \right], \tag{22}$$

where $\Delta A_{i,j}$ is the horizontal area of the grid box i, j;

$$\overline{U}_{i+\frac{1}{2},j} = \sum_{k=1}^{N} U_{i+\frac{1}{2},j,k}, \quad \overline{V}_{i,j+\frac{1}{2}} = \sum_{k=1}^{N} V_{i,j+\frac{1}{2},k} \tag{23}$$

are vertically integrated (barotropic) volume fluxes; and we have used the identity

$$(\zeta_{i,j} + h_{i,j}) \cdot \Delta A_{i,j} \equiv \sum_{k=1}^{N} \Delta V_{i,j,k}, \tag{24}$$

where $h_{i,j}$ is independent of time. Setting $k = N$ in (23), consistency with (22)–(24) results in

$$W_{i,j,N+\frac{1}{2}} = 0, \tag{25}$$

as required by the kinematic boundary condition at the free surface.

Thus far we have assumed that the time step and time-stepping algorithm for the tracer (16) and for ζ (22) are the same. This would be the case if the barotropic and baroclinic components were advanced using the same small time step dictated by the stability criterion for the barotropic mode. In a split-explicit, free-surface model, the equation for free-surface (22) and the vertically integrated (2D) momenta are advanced using a much smaller time step than the tracer equations. Each baroclinic time step starts with computation of the right hand side of the 3D momentum equations. The right-hand-side components are integrated vertically to provide forcing terms for the barotropic mode. During the barotropic time stepping, the free surface and the barotropic velocity components are advanced with the short time step, then averaged over the sequence of the barotropic steps to prevent aliasing. Then the 3D momenta are advanced to the baroclinic time step $n + 1$, and vertical integrals of the new fields are subtracted from the similar values from the barotropic sub-model. The resultant differences are then uniformly distributed throughout the vertical column to make sure that the corrected 3D velocity components have the same vertical integrals as do the barotropic components. At the same time, the free surface ζ at the new baroclinic step is assigned its new state from the barotropic sub-model.

Because of the replacement of ζ at $n + 1$ with its fast-time-averaged value, it is no longer possible to reconstruct the vertical velocity via (21) in such a way that the top kinematic boundary condition (25) is respected. Alternatively, one might distribute the mismatch in (25) throughout the water column, so that the top boundary condition holds, but at the expense of discrepancy in (21), [4]. In either case, a conservative update of the tracer fields (16) loses its constancy preservation property.

3.4. Mode-splitting error

In the split-explicit method, the shallow water equations (SWE) are advanced in time, using the smaller CFL-limited time step, to obtain the depth-integrated mode, *e.g.*,

$$\frac{\partial \overline{U}}{\partial t} + \cdots = -gD\nabla_x \zeta + \{gD\nabla_x \zeta + f\}. \tag{26}$$

Here, g is the acceleration of gravity; $D = h + \zeta$ is total depth; $\overline{U} \equiv D\bar{u}$ is depth-integrated velocity (barotropic mass flux); $\nabla_x \zeta$ is a shorthand for $\partial\zeta/\partial x$; and

$$f = -\frac{1}{\rho_0}\int_{-h}^{\zeta}\frac{\partial P}{\partial x}\,\mathrm{d}z \tag{27}$$

is the vertically integrated pressure gradient. The latter is a functional of the topography, free-surface gradient, and the free surface itself, as well as the vertical distribution of density and its gradient,

$$f = f[\nabla_x\zeta, \zeta, \nabla_x\rho(x), \rho(z)]. \tag{28}$$

The term in curly brackets in (26) is interpreted as barotropic–baroclinic mode coupling. It is kept fixed during the barotropic time stepping while the first term on the right hand side evolves. The disadvantage of this approach is that after the barotropic time stepping is complete and the new free-surface field is substituted into the full baroclinic pressure gradient, its vertical integral will no longer be equal to the sum of the SWE-like pressure gradient (computed using the new free surface, ζ' say) and the original coupling term (based on the old free surface, ζ). This type of discrepancy is usually known as a *mode-splitting error*.

Unfortunately, this error plays the role of a disturbance, causing the vertically integrated pressure gradient to not be in equilibrium with the barotropic mass flux. The barotropic time stepping drives the barotropic part toward an equilibrium, but it is disturbed again due to the redefinition of the vertically integrated baroclinic pressure gradient Ref. [35] analyzed the stability of a coupled linearized system in an isopycnic vertical coordinate and showed that, if non-dissipative time stepping algorithms are used for both modes, the resultant model is unavoidably unstable. An alternative definition of the barotropic mode in an isopycnic model that reduces the mode splitting error is obtained by replacing both $gD\nabla_x\zeta$ terms in (26) with

$$\frac{\partial f}{\partial(\nabla_x\zeta)}\nabla_x\zeta + \frac{\partial f}{\partial\zeta}\zeta. \tag{29}$$

This replacement may be shown to remove the dominant mode splitting error [7].

3.5. Improved mode-splitting

The algorithms in ROMS take into account the effects of the non-uniform density field, resulting in a more accurate mode splitting method that is suitable for use in a terrain-following model. Consider a fluid element bounded horizontally by two vertical lines corresponding to the locations of ζ_i and ζ_{i+1} and vertically by the free surface and bottom. The horizontal component of the pressure-gradient force acting on this element is calculated by the integration of the pressure along the contour surrounding the fluid element:

$$F_{i+\frac{1}{2}} = \int_{-h_i}^{\zeta_i} P(x_i, z)\,\mathrm{d}z - \int_{-h_{i+1}}^{\zeta_{i+1}} P(x_{i+1}, z)\,\mathrm{d}z - \int_{x_i}^{x_{i+1}} P(x, -h(x))\left[-\frac{\partial h(x)}{\partial x}\right]\mathrm{d}x = J_i - J_{i+1} - J_{i+\frac{1}{2}}. \tag{30}$$

In (30) $P(x,z)$ is the hydrostatic pressure,

$$P(x,z) = g\int_{z'}^{\zeta_i}\rho(x,z')\,\mathrm{d}z'. \tag{31}$$

Assuming a finite-volume approach to approximate (31) and eventually (30) at the discrete level, the barotropic pressure-gradient force at the velocity point $i+1/2$ is a function of the density in the vertical columns i and $i+1$, as well as the free-surface elevations ζ_i, ζ_{i+1}. Hence,

$$F_{i+\frac{1}{2}} = f(\zeta_{i+1}, \bar{\rho}_{i+1,1}, \ldots, \bar{\rho}_{i+1,N}, \zeta_i, \bar{\rho}_{i,1}, \ldots, \bar{\rho}_{i,N}), \tag{32}$$

where the structure of the functional f depends upon the discretization details of the baroclinic pressure gradient, typically involving nonlinear interaction of the ζ_i and $\rho_{i,k}$ fields.

In the ROMS mode-splitting technique, it is assumed that ζ is changing during the barotropic time stepping while the density values $\{\bar{\rho}_{i,k}\}$ remain frozen and change only during the baroclinic time step. However, the nonlinear relation (32) holds in barotropic time. Of course, it would be prohibitively inefficient to recompute f in (32) at every barotropic step by vertical integration of the whole 3D pressure gradient. Instead, in each vertical column, once at every baroclinic time step before the barotropic mode calculation begins, ROMS computes a vertically averaged density,

$$\bar{\rho}(x) = \frac{1}{D}\int_{-h(x)}^{\zeta(x)} \rho(x,z)\,\mathrm{d}z \tag{33}$$

and a vertically averaged dynamical density,

$$\rho^*(x) = \frac{1}{\frac{1}{2}D^2}\int_{-h(x)}^{\zeta(x)} \left\{\int_z^{\zeta(x)} \rho(x,z')\,\mathrm{d}z'\right\}\mathrm{d}z, \tag{34}$$

where $D \equiv D(x) = \zeta(x) + h(x)$ is the total thickness of the water column. Changing the integration variable to $\sigma = (z-\zeta)/D$ yields

$$\bar{\rho}(x) = \int_{-1}^{0} \rho(x,\sigma)\,\mathrm{d}\sigma, \quad \rho^*(x) = 2\int_{-1}^{0}\left\{\int_{\sigma}^{0}\rho(x,\sigma')\,\mathrm{d}\sigma'\right\}\mathrm{d}\sigma,$$

which implies that $\bar{\rho}$ and ρ^* are actually independent of ζ as long as the density profile $\rho = \rho(\sigma)$ stays the same. Expressed in terms of $\bar{\rho}$ and ρ^*, (30) becomes

$$F_{i+\frac{1}{2}} = g\left\{\frac{\rho_i^* D_i^2}{2} - \frac{\rho_{i+1}^* D_{i+1}^2}{2} + \int_{x_i}^{x_{i+1}} \bar{\rho} D \frac{\partial h}{\partial x}\,\mathrm{d}x\right\}. \tag{35}$$

This is a finite-volume discretization of the pressure-gradient term in the vertically integrated momentum equation,

$$\frac{\partial}{\partial t}(D\overline{U}) + \cdots = -\frac{1}{\rho_0} g\left\{\frac{\partial}{\partial x}\left(\frac{\rho^* D^2}{2}\right) - \bar{\rho} D\frac{\partial h}{\partial x}\right\} = -\frac{1}{\rho_0} gD\left\{\rho^*\frac{\partial \zeta}{\partial x} + \frac{D}{2}\frac{\partial \rho^*}{\partial x} + (\rho^* - \bar{\rho})\frac{\partial h}{\partial x}\right\}. \tag{36}$$

If $\rho^* \equiv \bar{\rho} \equiv \rho_0$, the right side of (36) reverts back to the familiar SWE pressure-gradient term of (26), but in the general case non-uniformity of $\bar{\rho}$ and ρ^* leads to the appearance of two extra terms that are baroclinic in nature. The problem therefore reduces to the search for a suitable method of calculation of $\bar{\rho}$ and ρ^* from the 3D density field $\{\bar{\rho}_{i,k}\}$ and an appropriate discretization of (35) and (36).

To address the first issue, consider, *e.g.*, a piecewise-parabolic reconstruction of the vertical density profile from a set of discrete values $\{\bar{\rho}_k | k = 1, 2, \ldots, N\}$ that is interpreted as a set of grid-box averages within each vertical grid box H_k (Fig. 2),

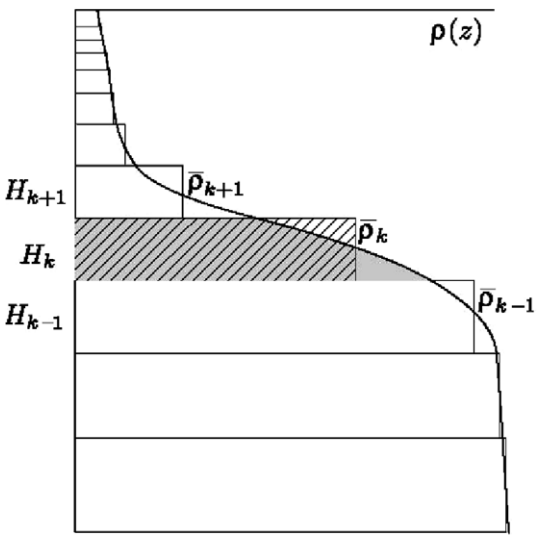

Fig. 2. Reconstruction of the vertical density profile by parabolic segments: for each $k = 1, 2, \ldots, N$, $\bar{\rho}_k$ are density averaged over grid boxes H_k of a vertically non-uniform grid. The shaded area is the same as the area of the diagonally hatched rectangle [7].

$$\rho(z') = \bar{\rho}_k + \frac{\rho_{k+\frac{1}{2}} + \rho_{k-\frac{1}{2}}}{H_k} z' + 6\left(\frac{\rho_{k+\frac{1}{2}} + \rho_{k-\frac{1}{2}}}{2} - \bar{\rho}_k\right)\left[\frac{z'^2}{H_k^2} - \frac{1}{12}\right]. \tag{37}$$

Here the local vertical coordinate z' spans the grid box H_k, so that $-\frac{H_k}{2} \leqslant z' \leqslant +\frac{H_k}{2}$, and $\rho_{k\pm\frac{1}{2}} \equiv \rho(\pm\frac{H_k}{2})$ are the density values at the upper and lower grid box interfaces $z = \pm\frac{H_k}{2}$ computed via an appropriate reconstruction algorithm. Regardless of the details of computing $\rho_{k\pm\frac{1}{2}}$, Eq. (37) guarantees that

$$\frac{1}{H_k}\int_{-H_k/2}^{+H_k/2} \rho(z')dz' \equiv \bar{\rho}_k \tag{38}$$

and leads to the discretization of vertically averaged density,

$$\bar{\rho}_i = \sum_{k=1}^{N} \bar{\rho}_{i,k} H_{i,k} \Big/ \sum_{k=1}^{N} H_{i,k}. \tag{39}$$

To compute ρ^* we note from 37,38 that the hydrostatic pressure in (30) can be expressed as a continuous function within each grid box H_k,

$$\begin{aligned} P(z') &= P_{k+\frac{1}{2}} + g\int_{z'}^{H_k/2} \rho(z'')\,\mathrm{d}z'' \\ &= P_{k+\frac{1}{2}} + gH_k\left\{\bar{\rho}_k\left[\frac{1}{2} - \frac{z'}{H_k}\right] + \frac{\rho_{k+\frac{1}{2}} - \rho_{k-\frac{1}{2}}}{2}\left[\frac{1}{4} - \frac{z'^2}{H_k^2}\right] + 2\left(\frac{\rho_{k+\frac{1}{2}} - \rho_{k-\frac{1}{2}}}{2} - \bar{\rho}_k\right)\left[\frac{z'}{4H_k} - \frac{z'^3}{H_k^3}\right]\right\}, \end{aligned} \tag{40}$$

where $P_{k+\frac{1}{2}}$ is the pressure at a depth corresponding to the interface between H_k and H_{k+1}

$$P_{N+\frac{1}{2}} = 0 \quad \text{and} \quad P_{k-\frac{1}{2}} = g\sum_{k'=k}^{N} \bar{\rho}_{k'} H_{k'}, \quad k = 1, \ldots, N. \tag{41}$$

It can be verified from (40) that $P(-H_k/2) \equiv P_{k-\frac{1}{2}}$ and that the pressure distribution and its first derivative are continuous across the grid box interfaces. Subsequent integration of (34) and (35) leads to

$$J_i = \int_{-h_i}^{\zeta_i} P_i(z)\,\mathrm{d}z = \sum_{k=1}^{N}\int_{-H_{i,k}/2}^{+H_{i,k}/2} P_i(z')\,\mathrm{d}z' = \sum_{k=1}^{N} H_{i,k}\bar{P}_{i,k}, \tag{42}$$

where

$$\overline{P}_{i,k} = P_{i,k+\frac{1}{2}} + \frac{1}{2}gH_{i,k}\left(\bar{\rho}_{i,k} + \frac{\rho_{i,k+\frac{1}{2}} - \rho_{i,k-\frac{1}{2}}}{6}\right) = \frac{P_{i,k+\frac{1}{2}} - P_{i,k-\frac{1}{2}}}{2} + gH_{i,k}\frac{\rho_{i,k+\frac{1}{2}} - \rho_{i,k-\frac{1}{2}}}{12} \tag{43}$$

is the pressure averaged over $H_{i,k}$. This further leads to the definition of the vertically averaged dynamical density as

$$\rho_i^* = \frac{1}{\frac{1}{2}\left(\sum_{k=1}^{N} H_{i,k}\right)^2} \cdot \sum_{k=1}^{N} H_{i,k}\left[\left(\sum_{k'=k+1}^{N} \bar{\rho}_{i,k'} H_{i,k'}\right) + \frac{1}{2}H_{i,k}\left(\bar{\rho}_{i,k} + \frac{\rho_{i,k+\frac{1}{2}} - \rho_{i,k-\frac{1}{2}}}{6}\right)\right]. \tag{44}$$

Using the identity,

$$\sum_{k=1}^{N} H_{i,k}\left[\left(\sum_{k'=k+1}^{N} H_{i,k'}\right) + \frac{1}{2}H_{i,k}\right] \equiv \frac{1}{2}\left(\sum_{k=1}^{N} H_{i,k}\right)^2$$

one can interpret (44) as just a weighted average. Furthermore, since

$$\sum_{k=1}^{N} H_{i,k} \equiv h_i + \zeta_i = D_i. \tag{45}$$

Eq. (42) may be expressed as

$$J_i = \frac{1}{2} g \rho_i^* D_i^2. \tag{46}$$

To approximate $J_{i+\frac{1}{2}}$, we assume that D, $\bar{\rho}$, and h are linear functions of the horizontal coordinate between points x_i and x_{i+1}:

$$\begin{aligned} J_{i+\frac{1}{2}} &= g \int_{x_i}^{x_{i+1}} \left(\bar{\rho}_i \frac{x_{i+1} - x'}{\Delta x} + \bar{\rho}_{i+1} \frac{x' - x_i}{\Delta x} \right) \left(D_i \frac{x_{i+1} - x'}{\Delta x} + D_{i+1} \frac{x' - x_i}{\Delta x} \right) \frac{h_{i+1} - h_i}{\Delta x} \\ &= g \frac{(\bar{\rho}_i + \bar{\rho}_{i+1})(D_i + D_{i+1}) + \bar{\rho}_i D_i + \bar{\rho}_{i+1} D_{i+1}}{6} (h_{i+1} - h_i). \end{aligned} \tag{47}$$

After some algebra, (46) and (47) yield

$$\begin{aligned} F_{i+\frac{1}{2}} = {} & g \frac{D_i + D_{i+1}}{2} \cdot \frac{\rho_i^* + \rho_{i+1}^*}{2} (\zeta_i - \zeta_{i+1}) + g \frac{D_i^2 + D_{i+1}^2}{4} (\rho_i^* - \rho_{i+1}^*) \\ & + g \frac{D_i + D_{i+1}}{2} \cdot \frac{(\rho_i^* - \bar{\rho}_i) + (\rho_{i+1}^* - \bar{\rho}_{i+1})}{2} (h_i - h_{i+1}) + \frac{(\bar{\rho}_{i+1} - \bar{\rho}_i)(D_{i+1} - D_i)(h_{i+1} - h_i)}{12}. \end{aligned} \tag{48}$$

The first three terms are obviously similar to the first, second, and third terms on the second line in (36), respectively. The fourth term in (48) is on the order of $O((\Delta x)^3)$ while all three preceding terms are $O(\Delta x)$, so the former is negligible relative to the others as $\Delta x \to 0$.

In the case of $\zeta_i = \zeta_{i+1} = 0$, $D_i = h_i$ and $D_{i+1} = h_{i+1}$, hence (48) becomes

$$F_{i+\frac{1}{2}}^{(0)} = g(\rho_i^* - \bar{\rho}_i) \frac{h_i^2}{2} - g(\rho_{i+1}^* - \bar{\rho}_{i+1}) \frac{h_{i+1}^2}{2} + g(\bar{\rho}_i - \bar{\rho}_{i+1}) \frac{h_i^2 + h_i h_{i+1} + h_{i+1}^2}{6}. \tag{49}$$

Unlike the SWE pressure gradient, this does not vanish unless there is a special balance between the densities ρ_i^*, ρ_{i+1}^*, $\bar{\rho}_i$, $\bar{\rho}_{i+1}$, and the unperturbed thicknesses, h_i and h_{i+1}. For example, if density is a linear function of depth, $\rho = \rho(z) = -\alpha z$ resulting in

$$\bar{\rho}_i = \frac{1}{h_i} \int_{-h_i}^{0} (-\alpha z) \, \mathrm{d}z = \frac{\alpha h_i}{2}, \tag{50}$$

$$\rho_i^* = \frac{2}{h_i^2} \int_{-h_i}^{0} \int_{z}^{0} (-\alpha z') \, \mathrm{d}z' \, \mathrm{d}z = \frac{\alpha h_i}{3}. \tag{51}$$

Then $F_{i+\frac{1}{2}}^{(0)}$ vanishes, as verified by direct substitution of these expressions into (49).

We therefore split (48) into

$$F_{i+\frac{1}{2}} = F_{i+\frac{1}{2}}^{(0)} + F'_{i+\frac{1}{2}}, \tag{52}$$

where

$$\begin{aligned} F'_{i+\frac{1}{2}} = {} & -\frac{1}{2} g \Big\{ (h_i + h_{i+1})(\rho_{i+1}^* \zeta_{i+1} - \rho_i^* \zeta_i) + \rho_{i+1}^* \zeta_{i+1}^2 - \rho_i^* \zeta^2 i \\ & + (h_{i+1} + h_i) \left[(\rho_{i+1}^* - \bar{\rho}_{i+1}) \zeta_{i+1} + (\rho_i^* - \bar{\rho}_i) \zeta_i + \frac{1}{3} (\bar{\rho}_{i+1} - \bar{\rho}_i)(\zeta_{i+1} - \zeta_i) \right] \Big\} \end{aligned} \tag{53}$$

contains all the terms of (48) with ζ. The transition from (48) to (49)–(53) has no approximations.

If the density field is a function only of depth, the baroclinic pressure gradient should vanish. However, in order to make $f_{i+\frac{1}{2}}^{(0)} = 0$ in (49), there must be cancellation between its terms that can be achieved only by having a special relation between ρ_i^* and $\bar{\rho}_i$. Except for a few special choices of the density profile (constant, linear, or quadratic in z), this cancellation is not exact, but rather relies on the numerical accuracy of the integration method. This is often referred to as *hydrostatic inconsistency*. The use of a high-order integration method does not eliminate this inconsistency; it just reduces the error. For example, dropping the term,

$$g H_\kappa \frac{\rho_{k+\frac{1}{2}} - \rho_{k-\frac{1}{2}}}{12}, \tag{54}$$

in (43) is equivalent to switching from a parabolic to a trapezoidal rule in integration of the hydrostatic equation. Examples of errors arising from hydrostatic inconsistency are shown in, *e.g.* [1].

3.6. Time stepping the coupled baroclinic–barotropic system

We now summarize the time-stepping algorithm in ROMS, focusing on the discrete-time interactions between the modes (Fig. 3).

Stage 1: Compute the right side for the 3D momentum equations at time step n. Apply this right side to advance the 3D momenta using a Leapfrog (LF) step combined with a half-step, backward interpolation with Adams–Moulton (AM3-like) coefficients (the result is time-centered at $n+1/2$). Because no meaningful barotropic mass fluxes time-centered at $n+1/2$ are available yet, set the vertical averages for the newly computed fluxes back to $\langle \overline{\mathbf{U}} \rangle^n$.

Stage 2: Advance the tracer variables in a similar manner with a LF step combined with an AM3 interpolation, placing the resultant values at $n+1/2$. This predictor algorithm is constancy preserving, though not conservative. (This is acceptable for now. Constancy preservation and conservation are simultaneously imposed in the corrector step below.)

Stage 3: Compute the right side for the 3D momentum equations from the mass fluxes and tracers (via density) at $n+1/2$. Vertically integrate everything and also compute and store vertically averaged densities, $\bar{\rho}_i$ ρ_i^*, using (33)–(38) time-centered at $n+1/2$. Apply the right side to the 3D momentum variables, but do not finalize the time step since H^{n+1} and $\langle \overline{\mathbf{U}} \rangle^{\mathbf{n+1}}$ are not available yet.

Stage 4: Compute the right side terms for the barotropic mode from barotropic variables using (48) for the pressure gradient and subtract it from the corresponding vertical integrals of the 3D right side computed in Stage 3 (*i.e.*, convert them into baroclinic-to-barotropic forcing terms). Advance the barotropic variables (slightly beyond the baroclinic time step, $n+1$, depending on the shape of the fast-time filter), performing a 2-way, fast-time averaging of barotropic variables on the way. The baroclinic forcing terms are kept constant during this procedure, but the barotropic pressure-gradient terms are recomputed by (48) with participation of $\bar{\rho}_i$ and ρ_i^* at every barotropic step. Once this is complete, update the vertical coordinate system $\{z_{i,j,k}, z_{i,j,k+\frac{1}{2}}, \boldsymbol{H}_{i,j,k}\}^{n+1}$ to be consistent with $\langle \zeta \rangle^{n+1}$.

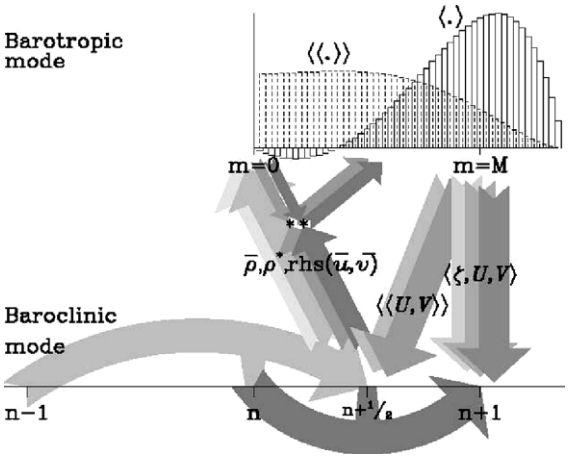

Fig. 3. Barotropic–baroclinic mode data exchange in ROMS: Curved horizontal arrows symbolize the predictor LF step combined with AM3 half-step-back interpolation of the result (light shading) and corrector sub-steps (dark shading). The four ascending arrows denote the 2-way, vertically averaged densities, $\bar{\rho}$ and ρ^*, and the vertically integrated right side for 3D momentum equations [the last two meet with the two small arrows symbolizing computation of the barotropic mode right hand side from barotropic variables; so that asterisks (**) note the computation of baroclinic-to-barotropic forcing terms, two small arrows ascending to the right]. The five large descending arrows symbolize 2-way, fast-time-averaged barotropic variables for backward coupling. Each arrow originates at the time when the data is logically available, regardless of the temporal placement of the corresponding variable [7].

Stage 5: Finalize the computation of the 3D mass fluxes begun in Stage 3 using the now available H^{n+1}, and set the vertical average to $\langle\langle \overline{\mathbf{U}} \rangle\rangle^{\mathbf{n+\frac{1}{2}}}$. from the barotropic mode.

Stage 6: Interpolate the 3D velocity components back in time to $n+1/2$ using a combination of the new time-step values (from Stage 5), values from the predictor step (Stage 4), and old-time step values. Set the vertical average of the resultant fields to $\langle \overline{\mathbf{U}} \rangle^{\mathbf{n+1}}$. Use the resultant velocity field and tracers at $n+1/2$ to compute the tracer fluxes and advance the tracers to $n+1$. This step is both conservative and constancy preserving.

3.7. Upstream-biased advection

The upstream-biased horizontal tracer advection in ROMS is constructed via a flux-form finite volume method with the upwinding/upstreaming being achieved through a simple min/max flux directional discriminator [33]. For a tracer quantity q, which would typically be the temperature or the salinity, the advection terms can be compactly written as

$$\Delta t \cdot [(F_{i+1/2,j,k} - F_{i-1/2,j,k}) + (G_{i,j+1/2,k} - G_{i,j-1/2,k})]/(\Delta x_{i,j} \Delta y_{i,j}), \tag{55}$$

where a zonal flux component has the form

$$\begin{aligned} F_{i-1/2,j,k} = {} & \Delta z_{i,j,k} \Delta y_{i,j} U_{i-1/2,j,k} \cdot (q_{i-1,j,k} + q_{i,j,k})/2 - \max(\Delta z_{i,j,k} \Delta y_{i,j} U_{i-1/2,j,k}, 0) \cdot (q_{i,j,k} - 2q_{i-1,j,k} + q_{i-2,j,k})/6 \\ & - \min(\Delta z_{i,j,k} \Delta y_{i,j} U_{i-1/2,j,k}, 0) \cdot (q_{i+1,j,k} - 2q_{i,j,k} + q_{i-1,j,k})/6 \end{aligned} \tag{56}$$

and a meridional flux component is defined as

$$\begin{aligned} G_{i,j-1/2,k} = {} & \Delta z_{i,j,k} \Delta x_{i,j} V_{i,j-1/2,k} \cdot (q_{i,j-1,k} + q_{i,j,k})/2 - \max(\Delta z_{i,j,k} \Delta x_{i,j} V_{i,j-1/2,k}, 0) \cdot (q_{i,j,k} - 2q_{i,j-1,k} + q_{i,j-2,k})/6 \\ & - \min(\Delta z_{i,j,k} \Delta x_{i,j} V_{i,j-1/2,k}, 0) \cdot (q_{i,j+1,k} - 2q_{i,j,k} + q_{i,j-1,k})/6. \end{aligned} \tag{57}$$

The accuracy of this formulation is easily verified under the simplification where the velocities are assumed to be globally constant and the grid metrics are ignored. Then, for example, the zonal (x-direction) advective flux (with a constant velocity $U > 0$ and retaining only the zonal grid indices) is

$$F_{i+1/2,j,k} - F_{i-1/2,j,k} = U \cdot (q_{i+1} - q_{i-1})/2 - U \cdot (q_{i+1} - 3q_i + 3q_{i-1} - q_{i-2})/6, \tag{58}$$

where the first flux term is a lower-order, spatially second-order accurate flux and the second term is a higher-order, spatially third-order accurate flux. The respective numerical truncation errors for these two flux terms on a regular grid with spacing h are:

$$\begin{aligned} & (q_{i+1} - q_{i-1})/2 = hq_x + (h^3/3!) \cdot q_{xxx} + \mathrm{O}(h^5), \\ & (q_{i+1} - 3q_i + 3q_{i-1} - q_{i-2})/6 = (h^3/3!) q_{xxx} - (h^4/12) \cdot q_{xxxx} + \mathrm{O}(h^5) \end{aligned} \tag{59}$$

from which we find

$$[F_{i+1/2,j,k} - F_{i-1/2,j,k}]/h = U \cdot [q_x + (h^3/12) \cdot q_{xxxx}] \tag{60}$$

and therefore the advection scheme is third-order accurate in space. The min/max switch facilitates the prescription of the higher-order flux contribution from the upstream direction. Note that the leading-order error term in this upstream-biased formulation is equivalent to a hyper-viscosity term.

3.8. Conservative parabolic spline based vertical tracer advection

An ideal choice for vertical advection of tracers in ROMS is based on a discretization employing the conservative parabolic splines given by Eq. (36); (Fig. 2). The necessary property that these splines be conservative is stated in Eq. (38).

Splines are piecewise lower-order polynomials whose values and first derivatives are required to be continuous from one grid element to another. The spline represented by Eq. (43) can be written as follows for two consecutive vertical grid elements:

$$\begin{aligned} S_k(\zeta) &= q_k + [(F_{k+1/2} + F_{k-1/2})/\Delta z_k] \cdot z + 6[(F_{k+1/2} + F_{k-1/2})/2 - q_k] \cdot [z^2/(\Delta z_k)^2 - 1/12], \\ S_{k+1}(\zeta) &= q_k + [(F_{k+3/2} + F_{k+1/2})/\Delta z_{k+1}] \cdot z + 6[(F_{k+3/2} + F_{k+1/2})/2 - q_{k+1}] \cdot [z^2/(\Delta z_{k+1})^2 - 1/12], \end{aligned} \tag{61}$$

where now q is the tracer under consideration, F is the inter-facial flux, Δz is the grid element thickness and ζ is the dimensionless, local grid coordinate variable. For S_k, this coordinate is such that $-\Delta z_k/2 \leqslant z \leqslant +\Delta z_k/2$ and for S_{k+1}, it is $-\Delta z_{k+1}/2 \leqslant z \leqslant +\Delta z_{k+1}/2$.

As S is a second-order polynomial, the condition of derivative continuity requires that the first derivatives of S_k and S_{k+1} should match at the common elemental boundary such that

$$S'_k(+\Delta z_k/2) = S'_{k+1}(-\Delta z_{k+1}/2), \tag{62}$$

with S' being the derivative of S with respect to z. This derivative condition gives rise to a tri-diagonal system of equations for the inter-facial flux, F which after multiplying through by $\Delta z_k \cdot \Delta z_{k+1}$ is

$$2\Delta z_{k+1} F_{k-1/2} + (2\Delta z_{k+1} + \Delta z_k) F_{k+1/2+} \Delta z_k F_{k+3/2} = 3 \cdot (\Delta z_{k+1} q_k + \Delta z_k q_{k+1}). \tag{63}$$

This system is solved in ROMS via the traditional Gaussian elimination method with a forward sweep to generate the matrix coefficients and a back substitution to achieve the solution.

Upon obtaining the inter-facial flux values, they are thereafter transformed into the true tracer flux by multiplying them with the corresponding vertical ω-velocity (the velocity following the vertical σ-levels) which reside at the inter-facial locations due to the staggering of the ROMS vertical grid. As $\omega = 0$ at both the bottom and top of the model grid, this condition translates to a non-flux condition at both ends of the vertical grid. Finally, the tracers are updated for the vertical advection contribution as

$$Q_k \rightarrow Q_k + \Delta t \cdot (F_{k+1/2} - F_{k-1/2})/\Delta z_k. \tag{64}$$

4. Skill assessment

4.1. Skill assessment in a tidally driven estuary

Estuarine salinity structure is a result of the interplay between the buoyancy flux from riverine inflow, advection by tides and the estuarine circulation, and mixing. Accurate numerical predictions of, say, the time-dependent salinity field thus depend critically on the model representation of tidal and sub-tidal motions as well as the sub-gridscale turbulence closure parameterizations for mixing of momentum and salt.

We applied the ROMS model to simulate the tidal dynamics and salt transport in the Hudson River Estuary [36] over a 50-day period with large variations in tidal forcing and river discharge. The Hudson River Estuary is located along the northeast coast of the US (Fig. 4) and provides New York State a major water resource for water quality control, navigation, and recreation. The model domain of the Hudson River Estuary extends from the Battery to the Federal Dam. The average width in the lower reach is about 2 km, broadens to nearly 5 km north of Hastings, and then narrows to nearly 500 m near the dam. This domain is horizontally discretized with 200 along-channel and 20 cross-channel cells. Grid resolution in the region of salt intrusion (first 40 km) has along-channel cell spacing on the order of 300 m and lateral spacing approximately 100 m. From km 40 to 250 the grid spacing telescopes linearly to reduce resolution in the upstream fresh water zone and approaches 4 km in length at the northern boundary. In the vertical, the bathymetry varies from 2 to 25 m with a mean depth of 8 m. The vertical dimension is discretized with 20 terrain-following sigma levels and a vertical stretching parameter allows increased resolution near the surface and bottom boundaries.

Boundary conditions are zero wind stress at the free surface and a bottom roughness length $z_0 = 0.002$ m. At the northern boundary, a depth-integrated flow was imposed to provide transport into the domain that was consistent with observed flows for the time period. Salinity was set to zero. At the southern end, the observed time series of water level was used to force the depth-integrated momentum through a reduced physics boundary

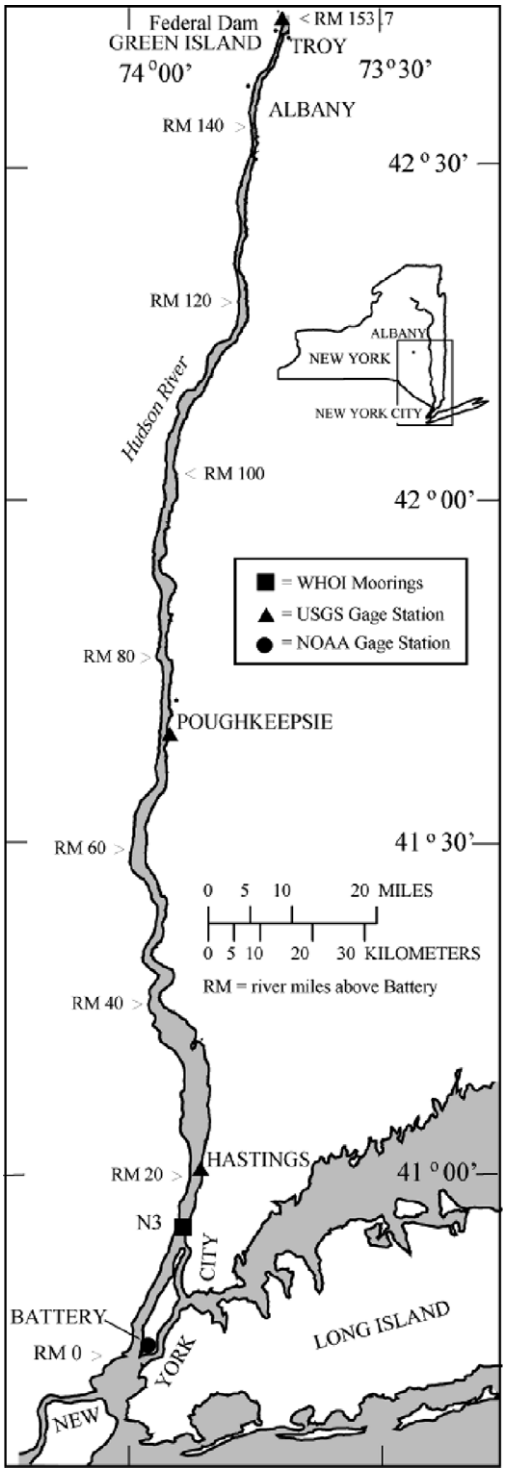

Fig. 4. Hudson River, observation site locations, and model domain [36].

condition. The salinity at the southern end was imposed using a gradient boundary condition. The salinity gradient depends on the magnitude of the salinity, with near-zero gradients as the salinity approaches zero and oceanic values, and a maximum gradient as the salinity reaches mid-estuarine values.

The 50-day simulation period includes three spring tides and two neap tides with spring tides centered on days 117, 132, and 145. The magnitude of semidiurnal tidal sea level amplitude approaches 0.8 m during spring tides and decreases to nearly 0.4 m during neap tides. River discharge varied from 400 $m^3 s^{-1}$ to a freshet that reached 2400 $m^3 s^{-1}$ from days 132 to 140. The increased fresh water transport was coincident with the weakest of the spring tides.

Validation is conducted with respect to a measurement program consisting of a 43-day deployment of oceanographic equipment from April 23 to June 5, 2002 (year days 113–156; [37]). Sensors to measure pressure, temperature, conductivity, and vertical profiles of velocity were positioned on bottom-mounted tripods and taught wire mooring arrays at six locations at cross section N3, near river km 23. In addition, nine along-channel hydrographic surveys were carried out to measure the vertical structure of salinity along the main axis of the estuary.

Predictive skill is based on quantitative agreement between model and observations. Using a method presented in [38] we define:

$$\text{Skill} = 1 - \frac{\sum |X_{\text{model}} - X_{\text{obs}}|^2}{\sum (|X_{\text{model}} - \overline{X_{\text{obs}}}| + |X_{\text{obs}} - \overline{X_{\text{obs}}}|)^2}, \tag{65}$$

where X is the variable being compared with a time mean. Perfect agreement between model results and observations will yield a skill of one and complete disagreement yields a skill of zero. Model skill was evaluated for all prognostic quantities.

Skill for the different model results are listed in Table 1. The model has skill greater than 0.85 in predicting barotropic quantities of sea level, depth-averaged velocity and barotropic transport (flux of water). This skill is highest in the lower reach of the estuary where the study was focused. For example, the model has significant skill at N3 in simulating the observed time series of surface and bottom salinity (Fig. 5) as well as the vertical structure of salinity and velocity (Fig. 6). The simulation has least skill (0.68) in reproducing the near-surface circulation, a possible consequence of the omission of explicit surface forcing (*e.g.*, the winds). In this application, alternative vertical mixing schemes, drawn from the GLS formulation, show little systematic impact on skill (Fig. 6).

4.2. Nitrogen cycling in the coastal ocean

Continental shelves play a key role in the global cycling of biologically essential elements such as nitrogen and carbon. The shelves are known to be highly productive, the majority of the oceanic burial of organic carbon occurs on continental shelves and the adjacent slope, and shelf sediments are important sites for denitrification. However, the exact role of continental shelf processes in global biogeochemical cycling remains poorly understood.

Table 1
Predictive skill of model parameters [36]

Parameter	Skill (Eq. (65))
Sea level	0.85–0.95
Depth-average velocity	0.92
Salinity	0.85
Vertical velocity profiles	0.89
Vertical salinity profiles	0.77
Salt intrusion length	0.87
Longitudinal salinity gradient	0.82
Estuarine circulation:	
Near bottom	0.78
Near surface	0.68
Barotropic transport	0.91
Salt flux	0.91

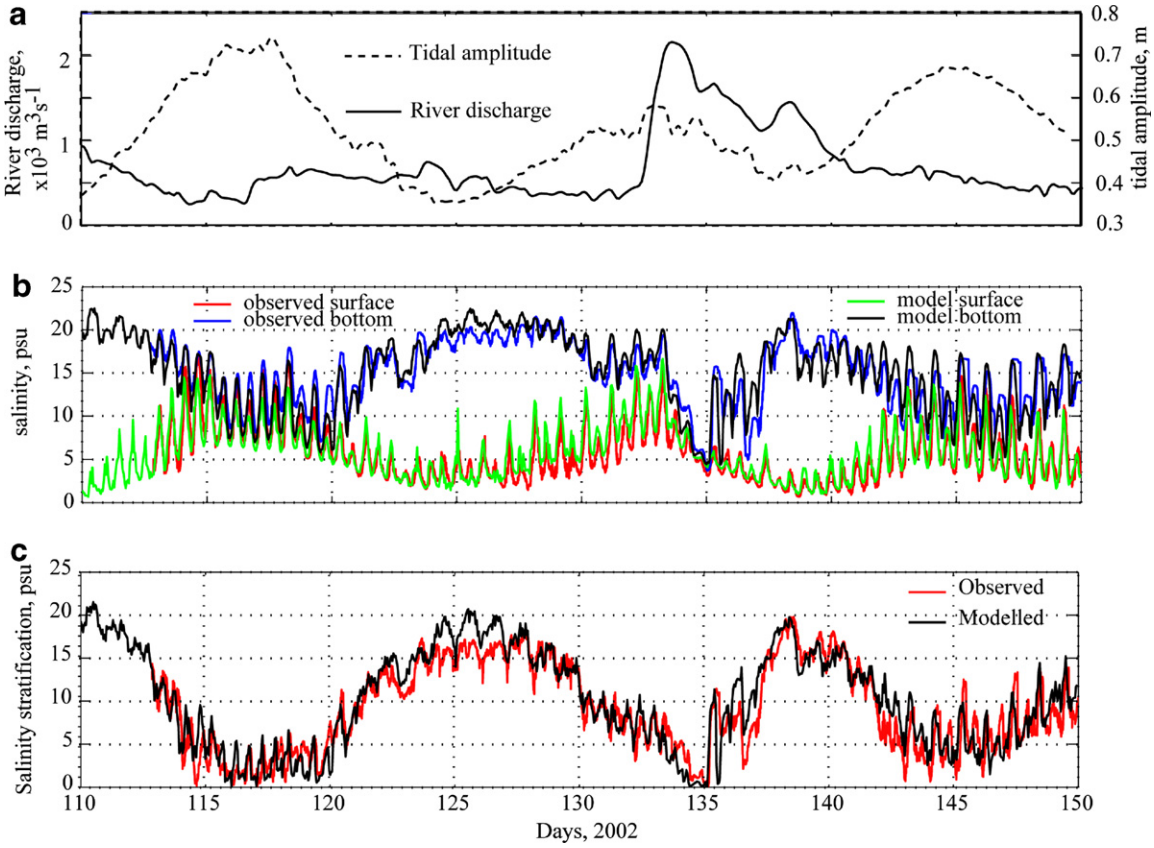

Fig. 5. (a) Environmental conditions of river discharge at Green Island dam and amplitude of semidiurnal sea level. Comparison of observational and model results at site N3 for (b) surface and bottom salinity and (c) vertical salinity stratification [36].

Here ROMS is implemented for the NorthEast North American (NENA) continental shelf [24]. The hydrography of the region is to a large extent influenced by remote processes, including the relatively cool and fresh shelf water entering from the north, as well as the entrainment of warm and salty waters originating from the Gulf Stream. The influences of these remote influences are incorporated by nesting of the NENA domain within a ROMS North Atlantic (NA) model. The horizontal resolution of the NA model is 10 km, and 30 sigma levels are used in the vertical direction. Prior experience with ROMS in the NA has shown that a horizontal resolution near 10 km is sufficient to obtain a satisfactory Gulf Stream separation and nominal eddy variability [6].

The NENA model has been implemented with similar resolution to the NA configuration (10 km in the horizontal direction, 30 vertical levels) and is forced with the same 3-day average surface fluxes. A one-way nesting procedure is employed whereby 3-day average temperature, salinity and baroclinic velocity from the basin-wide simulation are imposed at the oceanic perimeter of the shelf model. The depth-average velocity and sea surface height boundary conditions [39,40] are used to allow the outward radiation of gravity waves generated within the domain without upsetting the perimeter tracer fluxes. Coastal freshwater inputs are applied using observed river flow data from USGS and Canadian Rivers. Tidal currents and the associated mixing are relatively small in most of our model domain, except for the Gulf of Maine/Bay of Fundy system; tidal forcing is therefore omitted.

A coastal biogeochemical model of Fasham type (Fig. 7; Section 2.4) was initialized with the state of the NA model at 1 January 1993, spun up for 1 year, and integrated for the subsequent 2 years (1994 and 1995). This simulation is evaluated with a particular focus on the Middle Atlantic Bight (MAB). The MAB is the region of the eastern continental shelf of the United States that extends from Nantucket Shoals in

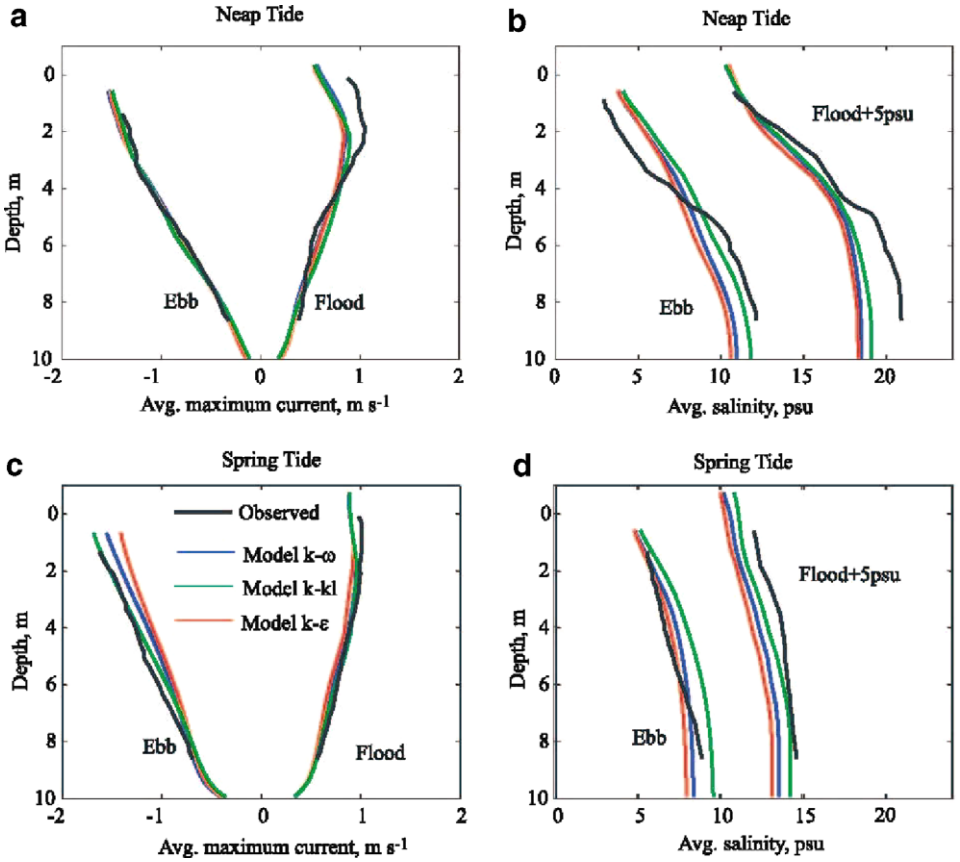

Fig. 6. Vertical structure of modeled and observed tidal currents and salinity at site N3. Average maximum flood and maximum ebb tidal current profiles during (a) neap tides and (b) spring tides. Average salinity profiles at maximum flood and maximum ebb during (c) neap tides and (d) spring tides. Flood tide salinity profiles are offset 5 psu [36].

the north to Cape Hatteras in the south. Available data sets from the MAB were used to create a merged biogeochemical dataset, including chlorophyll, nitrate and ammonium concentrations. A monthly climatology of these variables for the inner and outer shelf (delineated by the 50-m isobath) was generated for model/data comparison.

An example of spatial patterns in surface chlorophyll is given in Fig. 8, where the simulated mean surface chlorophyll for July 1994 is shown in comparison to the SeaWiFS chlorophyll for July 2003. Chlorophyll concentrations are highest near the coast, decrease on the outer shelf and again in the slope waters with local chlorophyll maxima in the Gulf Stream front; lowest concentrations are found in the Sargasso Sea. The model underestimates surface chlorophyll on Georges Bank when compared to SeaWiFS chlorophyll, owing to the neglect of tides and tidal mixing.

A more quantitative assessment of model/data agreement is shown in a Taylor diagram [41] in which the ratio of the standard deviations of the simulated and observed fields, their correlation, and their centered RMS difference can be displayed by one point in a single 2D diagram (Fig. 9). The ensemble of SeaWiFS points in Fig. 9 illustrates the discrepancies to be expected when comparing a single year to the climatology. Note that model results are expected to deviate somewhat from the climatology owing to unresolved internal variability, imperfect initial and forcing conditions, and most importantly in our case owing to interannual variability. Hence apparent differences between the test and the reference fields may be statistically or practically insignificant.

The expected level of difference in pattern correspondence due to interannual variability can be gauged by the comparison of the individual SeaWiFS years (1998–2004) with the climatology. If the whole domain is

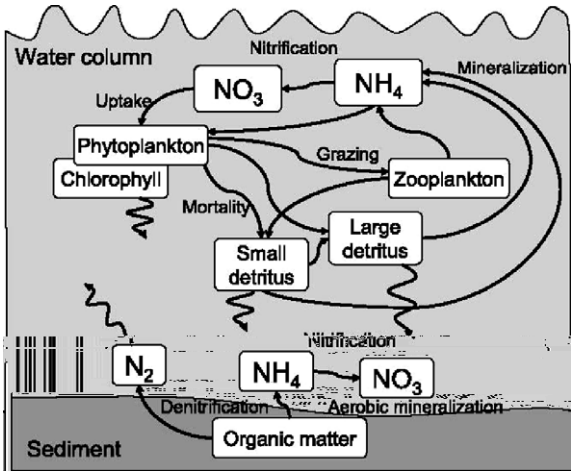

Fig. 7. Schematic of the biological model applied to the NorthEast North Atlantic model [24].

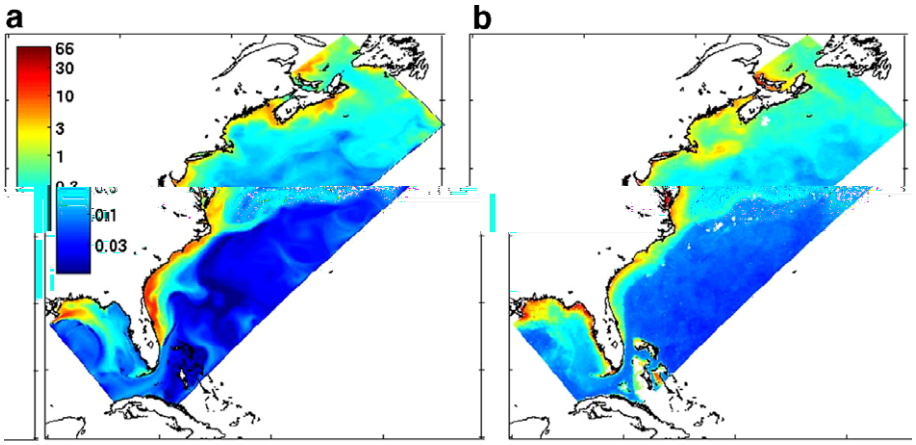

Fig. 8. (a) Model-simulated mean surface chlorophyll for July of 1994 and (b) SeaWiFS mean chlorophyll for July of 2003 (same colorscale) [24].

considered, the match between an individual SeaWiFS year and the climatology is within 10% of the amplitudes of variations ($0.9 \leqslant \sigma_n \geqslant 1.1$), with a correlation $R > 0.9$, and an RMS difference $E < 0.4$. With the exception of 1998, a similar match is obtained when the analysis is restricted to the MAB. 1998 is anomalous with amplitudes of variation significantly larger than the climatology and the other SeaWiFS years. The patterns of modeled surface chlorophyll and the SeaWiFS climatology are within 40% of the amplitudes of variation ($0.6 \leqslant \sigma_n \geqslant 1.4$) and have a lower correlation (0.4–0.8). The model's spatial resolution is thus sufficient to capture the variability of surface chlorophyll that is represented by the 9-km SeaWiFS data. When the statistics are restricted to the MAB, the simulated patterns match the SeaWiFS climatology more closely in terms of centered RMS difference and correlation than SeaWiFS 1998.

4.3. Basin-scale response to large-scale air-sea coupling

Regional ecosystem variability is an integrated response to local/short- and basin/climate-scale changes. Ecosystem dynamics have been linked to local topographic and coastal features and meso-scale activity, as well as to long-term and basin-scale ocean states and inter-annual changes caused by phenomena such as

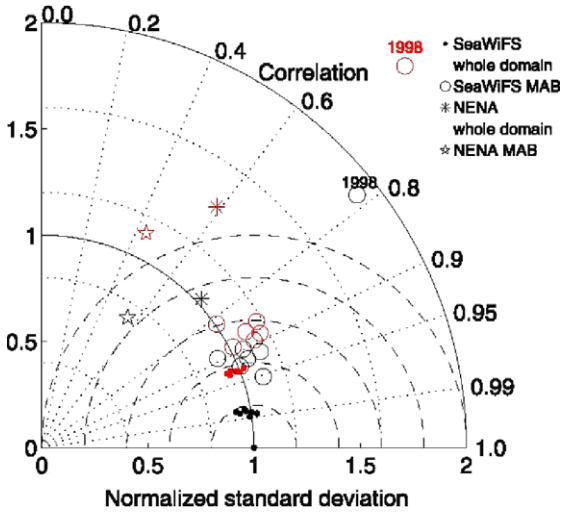

Fig. 9. Taylor diagram comparing monthly mean fields of log-transformed chlorophyll estimates from the model and SeaWiFS (the "test" fields) with a log-transformed SeaWiFS climatology (the "reference" field indicated by the black dot on the x-axis) for the whole model domain and for the Middle Atlantic Bight only. The radial distances from the origin are proportional to the ratio standard deviations; the azimuthal positions indicate the correlation coefficient; and the distance between the "test" points and the "reference" point indicates the centered RMS difference. The definitions of the total time–space (red) and spatial–annual (black) statistics [24].

the El Niño/Southern Oscillation. One of the main challenges is to understand the cumulative effects of the large- and regional-scale physics on ecosystem dynamics and variability.

A retrospective hindcast of the circulation in the North Pacific basin has been conducted to assess basin-scale variability on inter-annual and episodic time scales [42]. The North Pacific basin model domain used for this work ranges from 30 °S to 65 °N and from 100 °E to 290 °E in latitude and longitude, respectively. The average horizontal grid spacing is approximately 0.4° and 30 vertical layers are used. The resulting grid has $476 \times 238 \times 30$ points. The stretched generalized terrain-following coordinates (Section 3.1) result in a vertical distribution of grid points that is enhanced in the surface and benthic boundary layers. The baroclinic and barotropic timesteps are 1600 and 36 s, respectively. Vertical mixing is determined by the KPP scheme. As a consequence of implicit scale-selective smoothing in the third-order upwind advection scheme (Section 3.7), no explicit horizontal dissipation or diffusion is needed. A non-dimensional quadratic bottom drag coefficient of 3×10^{-3} is used. The model bathymetry is interpolated from ETOPO5 [43] with a minimum depth set to 40 m. An attempt was made to keep isobaths from intersecting side walls and to keep shallow areas intact. Nevertheless, at the current resolution, the shelf areas are poorly resolved and topographic smoothing creates significant discrepancies. Tides and riverine inputs are omitted in these computations.

In order to evaluate the basin-wide skill of the hindcast simulations, we compare the model output with the sea surface height fields from the SSALTO/DUACS 14 delayed time product distributed by AVISO at [44]. This product combines Topex/Poseidon, ERS-1, 2 and Jason altimetry data into weekly global sea surface height anomaly maps with the spatial resolution of 1°. Following [45], the comparison is performed using 4° longitude × 1° latitude resolution. Fig. 10 shows the root-mean-square (RMS) and correlation coefficient for the model and data sea surface height time series over the entire 1996–2002 period. This comparison assesses the model ability to reproduce large-scale features of the ocean circulation in the North Pacific. The analysis shows positive correlations in most regions; in the tropical and Eastern Pacific positive correlations can be higher than 0.8. The RMS difference is below 4 cm in many areas with the errors generally smallest in the Eastern Pacific.

The poorest correlation and largest RMS differences occur in the Western Pacific where the model is unable to reproduce the observed monthly variability. We note that in this region the dominant mode of monthly variability is related to the western boundary current and its associated dynamics (*e.g.*, meanders, eddy formation, etc.) which at the current spatial resolution of 0.4° the model is unable to fully reproduce. Another

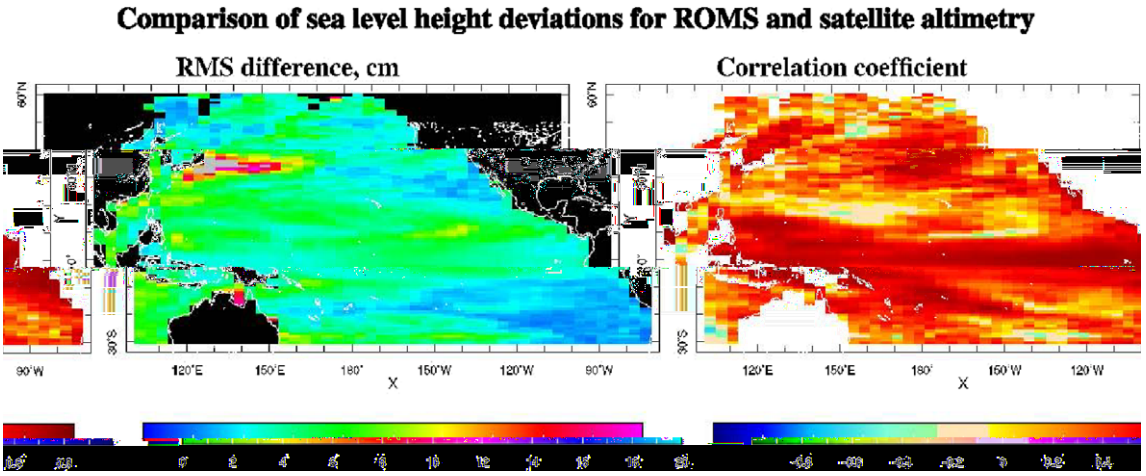

Fig. 10. Comparison for $4^\circ \times 1^\circ$ monthly means of the ROMS North Pacific simulation for 1996–2002 forced by daily fluxes from the NCEP–NCAR Reanalysis [52] with sea level heights from SSALTO/DUACS time-delayed merged altimetry maps [42].

region where the correlation deteriorates is in the eastern and northern Gulf of Alaska and along the southern side of the Aleutian Islands. These are regions where long-lived (several years) anti-cyclonic eddies, denoted as Haida, Sitka, and Yakutat eddies, have been observed [46,47], but are poorly represented by the model at the current resolution.

While many features of the North Pacific circulation and its variability are reproduced by the model, the comparisons shown in Fig. 10 also indicate deficiencies in the numerical solution. We make the observation that the patterns of error seen in Fig. 10 reflect features of the oceanic circulation (most obviously western intensification). We speculate that inaccurate representation of sub-gridscale processes (below 0.4° in this case) is the main contribution to the errors in the large-scale mean circulation. However, we cannot discount at this stage that errors in the forcing fields are also a prominent contributor to overall error levels.

The dominant mode of observed inter-annual variability in the Equatorial and Eastern Pacific is related to the El Niño/La Niña cycle. Cross correlation coefficients of sea surface height anomalies between a station near the equator (station 1 at 9.3 °N, off the coast of Panama) and various stations along the coast of North America are presented in Table 2. The correlation coefficient for the stations within the region of maximum El Niño influence (south of 28 °N) are 0.77–0.83 which correspond to the values (0.75–0.84) derived from the altimetry analysis of [48]. Our correlation value between the Equator and Alaska of 0.67 is considerably higher than the 0.41 of the satellite derived product. This is likely due to the inadequate representation in the model of the steep and narrow shelf off the coast of North America which limits the effectiveness of shelf-wave scattering.

4.4. Sea ice

Regional sea ice distribution and dynamics may be expected to respond rapidly and sensitively to variability in the Earth's climate system. The potential regional and/or global consequences of sea ice changes include sea

Table 2
Correlation coefficients of sea surface height anomalies between station 1 (latitude 9.3 °N) and stations along the coast of North America [42]

Station #	Location	Latitude (North) of station	Correlation coefficient (r)
3	Southern Mexico	21.1	0.83
4	Baja California	27.9	0.77
6	California	46.7	0.69
10	Alaska	58.9	0.69

level rise, increased freshwater fluxes to low-latitude coastal regions (*e.g.*, the MAB), and modifications or damage to regional ecosystems.

The dynamic–thermodynamic sea ice model of Section 2.5 has been coupled to a three-dimensional ocean general circulation model for the purpose of conducting ocean climate dynamical downscaling experiments for the Barents Sea region [28]. The regional model domain is shown in Fig. 11. The horizontal grid size varies between 7.8 and 10.5 km, with an average of 9.3 km. In the vertical, 32 coordinate levels are used, with enhanced resolution near the surface and bottom. The generic length scale (GLS) scheme [13] was used for sub-gridscale mixing of mass and momentum, with the two-equation k–kl model parameters. The time step used in the simulation was 450 s for both the ocean internal mode and the ice thermodynamics; the ratio of ocean internal to external mode time step is 45. The ratio of ice thermodynamic to dynamic time step is 60. Radiation open boundary conditions ([39,40]) were prescribed for barotropic normal velocity components and the free surface, respectively. Flow relaxation scheme [49] open boundary conditions were employed for three-dimensional velocity components and tracers.

The regional model was forced at the boundaries with interpolated 5-day mean fields from a larger-area model (LAM) and with tidal velocities and free-surface heights from eight constituents of the Arctic Ocean Tidal Inverse Model by [50]. The large-area model is used to supply boundary and initial conditions to the regional Barents model. A stretched spherical coordinate grid [51] is used in the horizontal, with the North Pole situated in central Asia and the South Pole situated in the Pacific Ocean west of North America. In the Barents Sea region, the horizontal resolution of the LAM is approximately 50 km. There were 30 generalized coordinate levels, stretched to increase vertical resolution near the surface and bottom. A time step of 1800 s was used for both the ocean internal mode and ice thermodynamic time step. A ratio of 40 was used between the ocean internal and external mode time steps. A ratio of 60 was used between ice thermodynamic and dynamic time steps. No tides were included in the LAM simulation. The vertical mixing scheme employed was the LMD [8] parameterization. The atmospheric forcing was obtained from the NCEP/NCAR reanalysis data [52]. Daily mean wind stress, and latent, sensible, downward shortwave radiative and net longwave radiative heat fluxes were applied as surface forcing after correcting for differences in model and NCEP surface conditions, such as in surface temperature and ice concentration. The flux corrections applied were developed

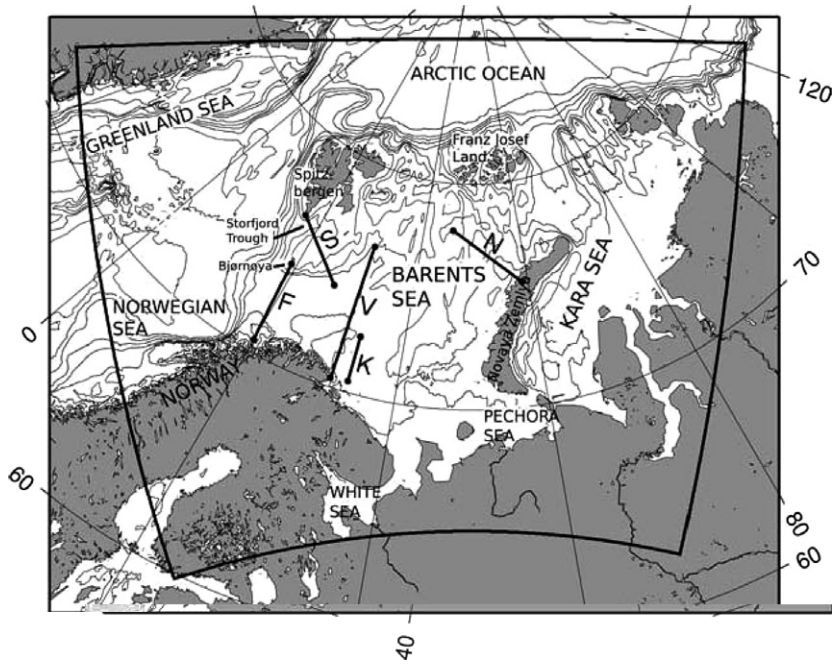

Fig. 11. Barents Sea regional model domain showing the location of sections: F (Fugløya-Bjørnøya), K (Kola), V (Vardø North), S (Spitzbergen Bank) and N (West Novaya Zemlya) [28].

by [53] and provide a feedback between the model surface temperature and applied heat fluxes, thus minimizing problems with drift in model surface temperatures.

Surface forcing for the regional model was the same as that applied in the large-area model simulation, but with the exception that the NCEP/NCAR Reanalysis cloud cover fraction was modified to provide the same monthly mean cloud cover climatology over the period 1983–2002 as the International Satellite Cloud Climatology Project (ISCCP) cloud cover data [54]. This necessitated the modification of the downward shortwave radiation and net longwave radiation fluxes to be consistent with the new cloud data. In the first simulations of Barents Sea ice cover, we found excessive melting of sea ice in summer by shortwave radiation due to the NCEP/NCAR Reanalysis cloud cover fraction being too low by a factor of approximately 0.75. This problem was largely, but not entirely, remedied by the use of ISCCP cloud cover data. Initial conditions for the Barents regional model were obtained from the archived 5-day mean large-area model fields interpolated to January 1, 1990.

The Barents simulation was conducted for the period 1990–2002. Temperatures have been averaged across portions of three sections in the Barents Sea: Fugløya-Bjørnøya, Vardø North and Kola, sections F, V and K, respectively, in Fig. 12. The Institute of Marine Research in Bergen, Norway, has been monitoring the Fugløya- Bjørnøya section six times per year and the Vardø North section for four times per year for 28 years. Section-mean temperatures have been computed over the depth range of 50–200 m and between 71.5° and 73.5 °N from the Fugløya-Bjørnøya section and between 72.25° and 74.25 °N from the Vardø North section. The overall error statistics are provided in Table 3. There is negligible bias in the modeled Fugløya-Bjørnøya results, but the bias in the Vardø North and Kola results are appreciable. The higher bias at the Kola section than at Vardø North is attributable to the 0–200 m depth range at Kola relative to the 50–200 m range at Vardø North used in constructing the averages. While the root-mean-square of the error after removal of the bias (denoted RMSEu) is comparable in the three sections, the bias error at the Vardø North and Kola

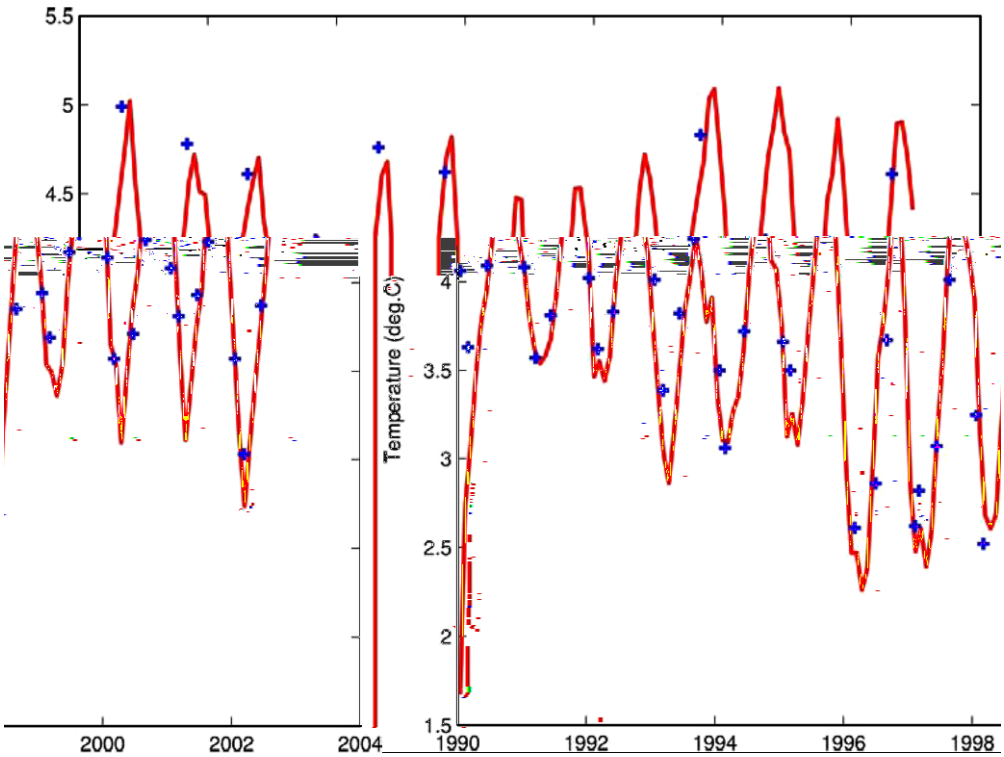

Fig. 12. Modeled monthly mean (red) and observed (blue) time series of the mean Vardø N section temperatures. Averages are computed over a depth range of 50–200 m, and from 72.25° to 74.25 °N. Note that the model values shown have been reduced by a constant 0.5 °C.[28].

sections is over twice the RMSEu and accounts for over 80% of the total mean-square error. As will be noted below, the cause of this positive bias error is likely the excessive inflow of Atlantic Water in the model results.

Shown in Fig. 12 are monthly mean model temperatures and averaged temperatures from the Vardø North hydrographic sections. A constant value of 0.5 °C was subtracted from the model Vardø North temperatures to remove the bias of +0.48 °C and simplify the comparison with the observations. The results from both sections show the model is in generally good agreement with the observed seasonal and interannual fluctuations.

Sea ice concentration in the Barents Sea can exhibit considerable variation both seasonally and inter-annually. Typical ice concentration distributions from the maximum and minimum ice extent periods during the

Table 3
Mean error (bias) and root-mean-square error with the bias removed for the Fugløya-Bjørnøya, Vardø North and Kola sections mean temperatures based on the period 1991–2002 [28]

Section	Mean error (_C)	RMSEu (C)bias removed
Fugløya-Bjørnøya	0.00	0.26
Vardø North	0.48	0.23
Kola	0.63	0.26

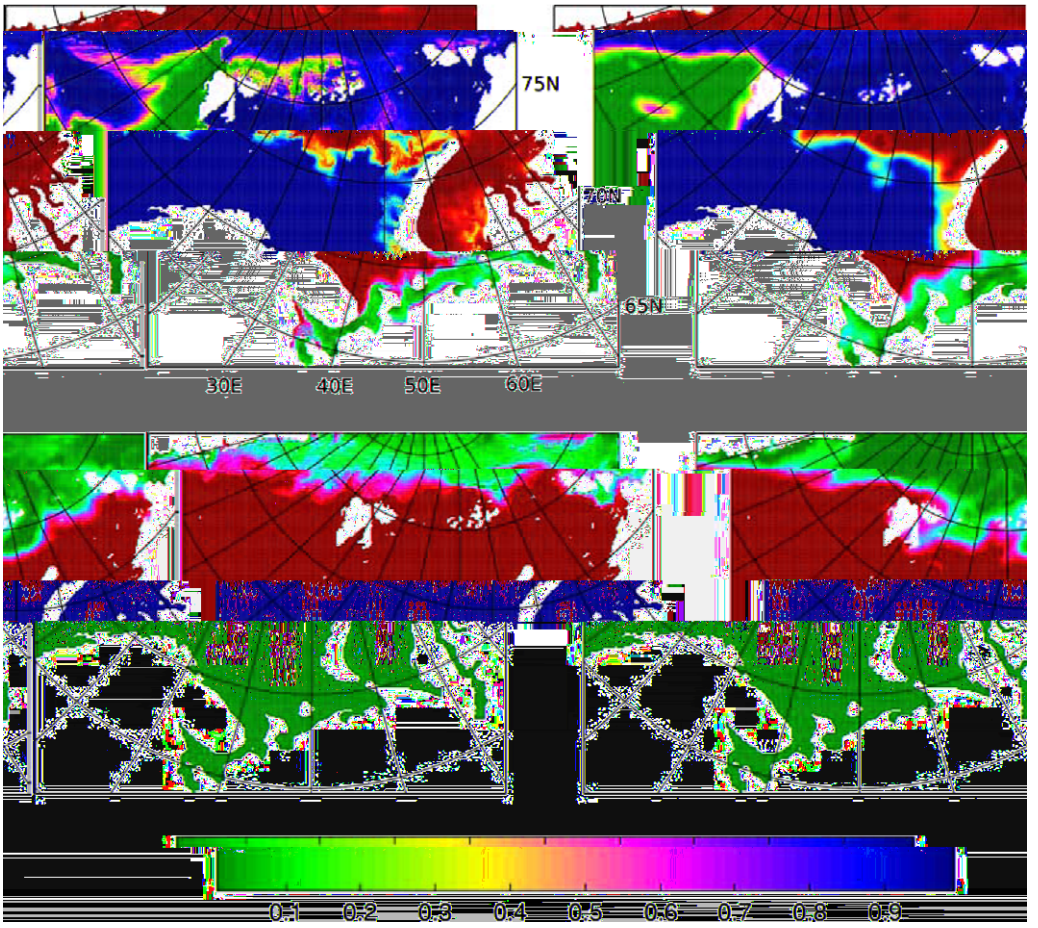

Fig. 13. Modeled versus Special Sensor Microwave/Imager (SSM/I) daily mean ice concentrations. The top row is from March 20, 1993, the lower row is from September 20, 1993. The left-hand column contains the model fields, the right-hand panel contains the SSM/I fields. The March modeled ice edge locations in the northern and southern Barents are in good agreement with observations, but show too much ice in the Greenland Sea and too little north and east of Spitzbergen. The September model results show an ice edge displaced to the north of the observed ice edge [28].

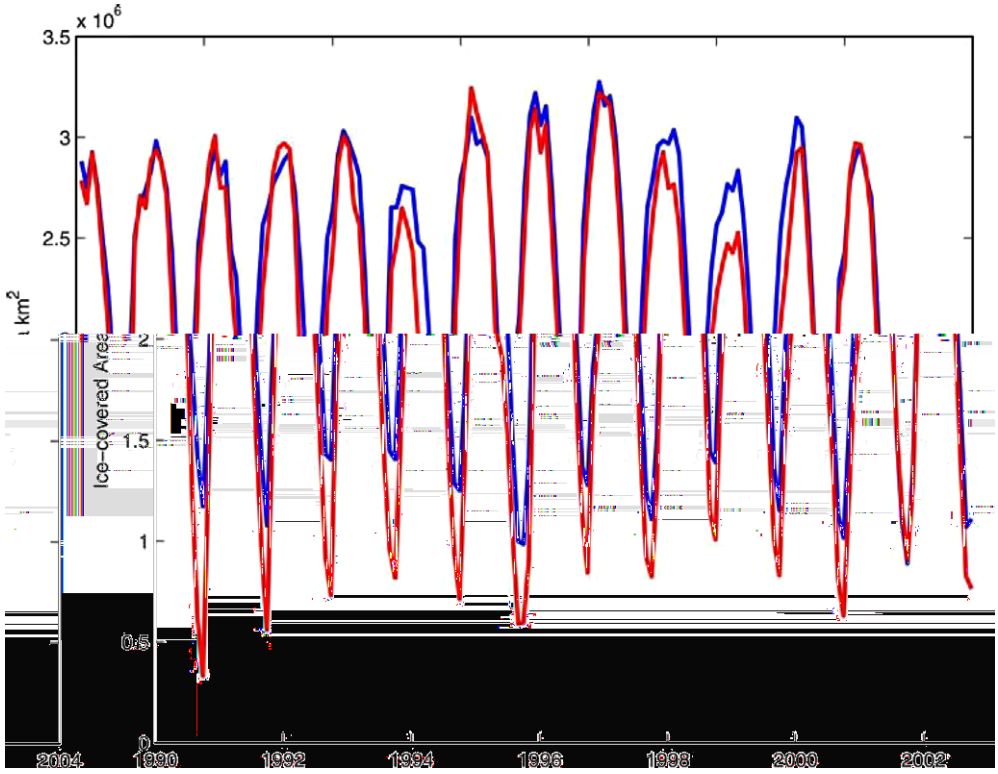

Fig. 14. Monthly mean modeled (red) and SSM/I (blue) time series of the ice-covered area in the model domain [28].

year are shown in Fig. 13. Given that the fields shown are daily mean values and, thus, effectively snapshots of ice distributions, the agreement between model fields and observations is remarkably good. The SSM/I (Special Sensor Microwave/Imager) data has a spatial resolution of 25 km, whereas the model resolution is 9 km. The satellite values are linearly interpolated onto the model grid and will appear somewhat smoother than the modeled values. The locations of the March modeled ice edges in the northern Barents (76 °N) and in the south-eastern Barents (69 °N) are in good agreement with observations. The model results show too much ice in the Greenland Sea and too little ice north of Spitzbergen along 82–85 °N, between 20° and 50 °E. The September model results are in good general agreement with the observations, but show too much ice-melt. The modeled ice edge is too far to the north.

To obtain an integral estimate of model performance, the monthly mean total areal ice cover has been tabulated from model results and from SSM/I satellite data. Fig. 14 shows that the model results successfully track the seasonal and inter-annual variability, but that there is a systematic under-estimation of summer ice cover, consistent with the September results shown in Fig. 13. The modeled under-estimate of ice concentration in summer is likely attributable to excessive shortwave radiation. It was argued previously that the positive bias in modeled temperatures 'downstream' of the Barents Opening is likely due to excessive inflow of Atlantic Water. If that is the case, one should expect to see a negative bias in the ice cover in the central Barents Sea. This condition is confirmed; the model results show a nearly constant offset of -1×10^5 km^2 from the observations in the 12-month running mean.

5. Future capabilities and the outlook for operational prediction systems

As demonstrated in these recent validation efforts, the Regional Ocean Modeling System has substantial skill in producing deterministic (forward-model) simulations across a range of space/time scales and oceanic system types. As suggested above, much of this skill can be credited to a robust computational engine that

combines novel methods, *e.g.*, higher-order spatial treatments, quasi-monotone advection, and conservative time-stepping. Nonetheless, limitations to the forward model algorithms still persist, most importantly having to do with the parameterization of sub-gridscale phenomena. The availability of, and access to, high-quality forcing data (*e.g.*, atmospheric fluxes) and validation fields (*e.g.*, in the oceanic interior) are also a perennial concern.

Through the efforts of its distributed network of development and applications specialists, the ROMS system continues to evolve at a rapid pace. Many additional capabilities, not reviewed in detail here, are in advanced stages of implementation and/or testing. Some of these include: an option for *non-hydrostatic physics*; dynamical enhancements to incorporate, *e.g.*, *near-shore processes* (wetting/drying, radiation stresses, surface waves); the capability to perform simulations on *block-structured grids*; *individual-based models* for higher trophic levels which explicitly incorporate behavior; and *advanced methods for data assimilation*. The latter are of particular importance for multi-purpose regional ocean prediction.

Forecasting systems based upon ROMS, which combine available observing networks and state-of-the-art methods for data assimilation, are under active development and initial application. For example, we are now implementing data assimilation methods for ROMS which allow a relatively inexpensive computation of the gradient of the fit cost function using the adjoint model [55]. We are also testing the 4D variational assimilation (4DVAR) approach as well as the Inverse Ocean Modeling system (IOM; [56]) representer method formalism for ROMS data assimilation. Proto-type demonstrations of ROMS in forecasting mode have been conducted in several oceanic regions including the US East Coast, the East Australia Current System, and the Southern California Bight.

One location, having advantages for testing of these new techniques, is the California Current System. The CCS is dynamically rich; it is forced both locally by atmospheric fluxes and remotely by larger-scale climate variability (*e.g.*, El Nino signals), and exhibits intense *in situ* mesoscale turbulence. In addition, the California Cooperative Oceanic Fisheries Investigations (CalCOFI) program has sampled the Southern California Bight

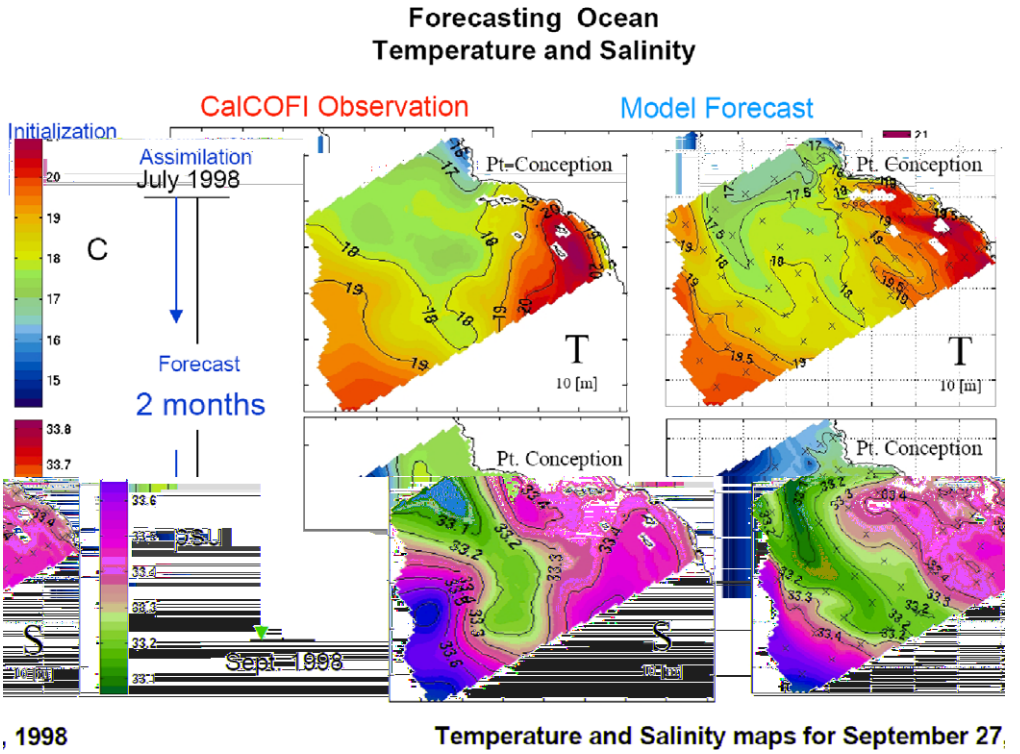

Fig. 15. Example of model forecast for the 1998 September CalCOFI cruise. The model was initialized with a simple data assimilation procedure in the month of July 1998 and then integrated forward with climatological winds and no further assimilation.

(SCB) region of the CCS for over 50 years, providing a unique time series of physical, chemical, and biological data (*e.g.* [57,58]).

The CCS has been extensively studied using ROMS in seasonal cycle simulations, retrospective hindcast and short-term and long-term forecasts [59]. Assuming that unstable mesoscale eddies dominate the physical balances, one anticipates that model fitting and/or forecasting skill is achievable to at least the eddy turn-around time scale, which is roughly one to three months. Indeed, some preliminary results reveal remarkable comparability between flows forecast from July 1998 to those observed 2 months later in September 1998 (Fig. 15). If atmospheric forcing dominates the flow variability, however, then the forecast skill is limited only by the quality of the atmospheric forecast (roughly 2 weeks for deterministic patterns, roughly 1 year for El Nino teleconnection signatures, and many decades for shifts in the climate due to greenhouse gas forcing scenarios, such as used by [60]). The 1997–1998 El Nino and the 1999–2000 La Nina time periods were sampled particularly well in the SCB. These will give us a unique opportunity to test fitting and forecasting skill and to assess dynamic and ecosystem balances over inter-annual time periods.

In summary, the timely confluence of three factors – robust, multi-purpose marine modeling systems; long-duration observational datasets, both new and retrospective; and advanced assimilation methodologies – is poised to bring about a period of rapid evolution in our ability to forecast and to manage the ocean circulation and its resources. The Regional Ocean Modeling System, as well as other marine modeling systems under concurrent development, can be expected to contribute importantly to this progress. In an initial proof of concept, the ROMS system described above is being ported to operational use within the National Oceanic and Atmospheric Administration for the forecasting of water level and currents in several US estuaries.

Acknowledgments

The development and application of the ROMS model has been supported by grants from the Office of Naval Research (N00014-05-1-0363), the National Science Foundation (OCE00-02892, OCE01-13461, OCE04-35577, ITR OCE01-21332, CCE-LTER OCE-0417616), the National Oceanic and Atmospheric Administration (NA17RJ1231 through ECPC), the Department of Energy (DE-FG02-04ER63857), and the National Aeronautics and Space Administration (NNH04ZYS008N). Development and application of the ROMS model has also been supported by the US Geological Survey Community Sediment Modeling Project. The authors thank Mr. David Robertson and Ms. Madeline Gazzale for technical assistance.

References

[1] D.B. Haidvogel, A. Beckmann, Numerical Ocean Circulation Modeling, Imperial College Press, London, 1999, 318 pp.
[2] A.F. Blumberg, G.L. Mellor, A description of a three-dimensional coastal ocean circulation model, Three-Dimensional Coastal Ocean Models, vol. 208, N. Heap, New York, NY, 1987.
[3] D.B. Haidvogel, J.L. Wilkin, R. Young, A semi-spectral primitive equation ocean circulation model using vertical sigma and orthogonal curvilinear horizontal coordinates, Journal of Computational Physics 94 (1991) 151–185.
[4] Y. Song, D.B. Haidvogel, A semi-implicit ocean circulation model using a generalized topography-following coordinate system, Journal of Computational Physics 115 (1) (1994) 228–244.
[5] http://www.myroms.org/.
[6] D.B. Haidvogel, H.G. Arango, K.S. Hedström, A. Beckmann, P. Malanotte-Rizzoli, A.F. Shchepetkin, Model evaluation experiments in the North Atlantic Basin simulations in nonlinear terrain-following coordinates, Dynamics of Atmospheres and Oceans 32 (2000) 239–281.
[7] A.F. Shchepetkin, J.C. McWilliams, The regional oceanic modeling system (ROMS) a split-explicit, free-surface, topography-following-coordinate oceanic model, Ocean Modelling 9 (4) (2005) 347–404.
[8] W.G. Large, J.C. McWilliams, S.C. Doney, A review and model with a nonlocal boundary layer parameterization, Reviews of Geophysics 32 (1994) 363–403.
[9] S. Durski, S.M. Glenn, D. Haidvogel, Vertical mixing schemes in the coastal ocean: Comparison of the level 2.5 Mellor–Yamada scheme with an enhanced version of the K-profile parameterization, Journal of Geophysical Research 109 (C01015) (2004), doi:10.1029/2002JC001702.
[10] G.L. Mellor, T. Yamada, Development of a turbulence closure model for geophysical fluid problems, Reviews of Geophysics and Space Physics 20 (1982) 851–875.
[11] L. Umlauf, H. Burchard, A generic length-scale equation for geophysical turbulence models, Journal of Marine Research 61 (2003) 235–265.

[12] W. Rodi, Turbulence models and their applications in hydraulics – a state of the art review, Technical report, Int. Assoc. for Hydraul. Res., Delf, The Netherlands, 1984.
[13] J.C. Warner, C.R. Sherwood, H.G. Arango, R.P. Signell, B. Butman, Performance of four turbulence closure models implemented using a generic length scale method, Ocean Modelling 8 (2005) 81–113.
[14] R. Styles, S.M. Glenn, Modeling stratified wave and current bottom boundary layers on the continental shelf, Journal of Geophysical Research 105 (C10) (2000) 24,119–24,139.
[15] R. Styles, S.M. Glenn, Modeling bottom roughness in the presence of wave-generated ripples, Journal of Geophysical Research 107 (C8) (2002), doi:10.1029/2001JC000864.
[16] R.L. Soulsby, Bed shear-stresses due to combined waves and currents, Advances in Coastal Morphodynamics: An Overview of the G8-Coastal Morphodynamics Project, 1995, pp. 4.20–4.23.
[17] W.D. Grant, O.S. Madsen, Movable bed roughness in unsteady oscillatory flow, Journal of Geophysical Research 87 (C1) (1982) 469–481.
[18] P. Nielsen, Suspended sediment concentrations under waves, Coastal Engineering 10 (1986) 23–31.
[19] M.Z. Li, C.L. Amos, SEDTRANS96: the upgraded and better calibrated sediment-transport model for continental shelves, Computers & Geosciences 27 (2001) 619–645.
[20] O.S. Madsen, Spectral wave-current bottom boundary layer flows, in: Coastal Engineering 1994, Proceedings, 24th International Conference Coastal Engineering Research Council, 1994, pp. 384–398.
[21] P.L. Wiberg, C.K. Harris, Ripple geometry in wave-dominated environments, Journal of Geophysical Research 99 (C1) (1994) 775–789.
[22] T.M. Powell, C.V. Lewis, E.N. Curchitser, D.B. Haidvogel, A.J. Herman, E.L. Dobbins, Results from a three-dimensional, nested biological–physical model of the California Current System and comparisons with statistics from satellite imagery, Journal of Geophysical Research 111, in press,

[43] ETOPO5, Data announcement 88-MGG-02, Digital relief of the surface of the Earth, 1988.
[44] http://ftp.cls.fr/pub/oceano/enact/msla/merged.
[45] A. Kaplan, M.A. Cane, D. Chen, D.L. Witter, R. Cheney, Small-scale variability and model error in tropical Pacific sea level, Journal of Geophysical Research 109 (C02001) (2004), doi:10.1029/2002JC001743.
[46] W. Crawford, J.Y. Cherniawsky, M.G. Foreman, Multi-year meanders and eddies in Alaskan Stream as observed by TOPEX/Poseidon altimeter, Geophysical Research Letters 27 (7) (2000) 1025–1028.
[47] C. Ladd, N. Kachel, C. Mordy, P. Stabeno, Observations from a Yakutat eddy in the northern Gulf of Alaska, Journal of Geophysical Research 110 (C3) (2002), doi:10.1029/2004JC002710.
[48] P. Strub, C. James, Altimeter-derived surface circulation in the large-scale NE Pacific gyres. Part 2: 1997–1998 el niño, Progress in Oceanography 53 (2–4) (2002) 185–214.
[49] H. Engedahl, Use of the flow relaxation scheme in a three-dimensional baroclinic ocean model with realistic topography, Tellus 47A (1995) 365–382.
[50] L. Padman, S. Erofeeva, A barotropic inverse tidal model for the Arctic Ocean, Geophysical Research Letters (2004), doi:10.1029/2003GL019003.
[51] M. Bentsen, G. Evensen, A.D. Jenkins, Coordinate transformation on a sphere using conformal mapping, Monthly Weather Review 127 (12) (1999) 2733–2740.
[52] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, A. Leetma, R. Reynolds, R. Jenne, D. Joseph, The NCEP/NCAR 40-year reanalysis project, Bulletin of American Meteorological Society 77 (3) (1996) 437–471.
[53] M. Bentsen, H. Drange, Parameterizing surface fluxes in ocean models using the NCEP/NCAR reanalysis data, RegClim General Technical Report No 4, Norwegian Institute for Air Research, Kjeller, Norway, 2000, pp. 44–57.
[54] R.A. Schiffer, W.B. Rossow, ISCCP global radiance data set—a new resource for climate research, Bulletin of American Meteorological Society 66 (12) (1985) 1498–1505.
[55] A.M. Moore, H.G. Arango, E. Di Lorenzo, B.D. Cornuelle, A.J. Miller, D.J. Neilson, A comprehensive ocean prediction and analysis system based on the tangent linear and adjoint of a regional ocean model, Ocean Modelling 7 (2004) 227–258.
[56] E. Di Lorenzo, A. Moore, H. Arango, B.D. Cornuelle, A.J. Miller, B. Powell, B.S. Chua, A. Bennett, Weak and strong constraint data assimilation in the inverse Regional Ocean Modeling System (ROMS): development and application for a baroclinic coastal upwelling system, Ocean Modeling 16 (2007) 160–187.
[57] D. Roemmich, J. McGowan, Climatic warning and the decline of zooplankton in the California Current, Science 267 (1995) 1324–1326.
[58] B.M. Hickey, Coastal oceanography of western North America from the tip of Baja to Vancouver Island, in: A.R. Robinson, K.H. Brink (Eds.), The Sea, The Global Coastal Ocean: Regional Studies and Syntheses, vol. 11, New York, 1998, pp. 345–393.
[59] E. Di Lorenzo, A.J. Miller, N. Schneider, J.C. McWilliams, The warming of the California Current: Dynamics and ecosystem implications, Journal of Physical Oceanography 35 (2005) 336–362.
[60] G. Auad, A.J. Miller, E. Di Lorenzo, Long-term forecast of oceanic conditions off California and their biological implications, Journal of Geophysical Research 111 (2006) C09008, doi:10.1029/2005JC003219.