The semi-prognostic method

Richard J. Greatbatcha,*, Jinyu Shenga, Carsten Edenb, Liqun Tanga, Xiaoming Zhaia, Jun Zhaoa

aDepartment of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1
bIFM-GEOMAR Kiel, Düsternbrooker Weg 20, D-24105 Kiel, Germany

Available online 4 October 2004

Abstract

An overview is given of the semi-prognostic method, a new and novel technique that can be used for adjusting models to correct for systematic error. Applications of the method to a regional model of the northwest Atlantic Ocean, and to an eddy-permitting model of the entire North Atlantic, show improvement in the handling of the Gulf Stream/North Atlantic Current systems, especially in the “northwest corner” region southeast of Newfoundland where prognostic models show systematic errors of as much as 10 °C in the temperature field. Use of the semi-prognostic method also leads to improvement in the modelled flow over the eastern Canadian shelf. An advantage of the semi-prognostic method is that it is adiabatic; in particular, in spite of the improvement seen in the modelled hydrography, the potential temperature and salinity equations carried by the model are unchanged by the method. Rather, the method introduces a correction term to the horizontal momentum equations carried by the model. Adiabaticity ensures that the method does not compromise the requirement for the flow in the ocean interior to be primarily in the neutral tangent plane, and also ensures that the method is well-suited for tracer studies. The method is also easy to implement, requiring only an adjustment in the hydrostatic equation carried by the model. We also describe the use of the method as a diagnostic tool, for probing the important dynamic processes governing a phenomenon, and finally as a technique for transferring information between the different subcomponents of a nested modelling system.

Keywords: Continental shelf; Modelling; Ocean basin; Ocean circulation; Atlantic Ocean; Scotian Shelf

1. Introduction

Models are often prone to systematic error. Examples are the tendency for the Gulf Stream to separate too far to the north, and for a poor representation of the region to the southeast of Newfoundland known as the “northwest corner” (Lazier, 1994), where the North Atlantic Current turns first northward and then abruptly eastward towards Europe (see e.g. Willebrand et al., 2001). Associated with the poor representation of the Gulf Stream system, the model sea surface
temperature (SST) can differ by as much as 10°C over large areas of the northwest Atlantic Ocean. Such large errors in SST can be expected to impact negatively on both the North Atlantic storm track in a coupled atmosphere/ocean modelling system (Hoskins and Valdes, 1990), and on the uptake of tracers such as carbon, both issues of importance for climate modelling. In addition, poor representation of ocean current systems in models can lead to erroneous pathways for tracers in models, and impact negatively on the overall tracer budget (see Zhao et al., 2004, for an example).

In this paper, we describe a new and novel technique that can be used to correct models for systematic error. The method, called “the semi-prognostic method”, is a technique for transferring data into a model or between models and was originally introduced by Sheng et al. (2001). Sheng et al. used a regional model for the northwestern part of the North Atlantic Ocean, including the eastern Canadian shelf, to show that application of the semi-prognostic method leads to a significant improvement in the representation of the Gulf Stream and the North Atlantic Current, including the circulation along the shelf break. Subsequently, Eden et al. (2004) showed the success of modified versions of the method for improving the performance of a 1/3° × 1/3° eddy-permitting model of the whole North Atlantic Ocean. The biggest improvements are again in the pathways of the Gulf Stream and the North Atlantic Current system, but with the added bonus that the poleward heat transport by the model is brought into better accord with observational estimates. The modifications introduced by Eden et al. (2004) avoid spurious damping of mesoscale eddy variability and distortion of the model physics that are a byproduct of the method as it appeared in Sheng et al. (2001). Eden and Greatbatch (2003) used the method in a different context, this time exploiting the distortion of the model physics to determine the essential dynamics governing the behaviour of a damped, decadal oscillation in a model of the North Atlantic. However, the method has application beyond it’s use as an adjustment or diagnostic technique. In particular, it can be used in a nested modelling system as a means of transferring information between the different submodels (Zhai et al., 2004). This is a new and exciting application of the method that is discussed in Section 4.

The principal advantages of the semi-prognostic method are (i) its simplicity and (ii) the fact it is adiabatic. The latter property arises because the adjustments to the model are made in the horizontal momentum equations, leaving the model tracer equations (in particular, the conservation equations for potential temperature and salinity) unchanged. As such, the method is ideal for tracer studies (see Zhao et al., 2004) since, like the conservation equations for an active tracer, the conservation equations for a passive tracer are unchanged by the method. The semi-prognostic method can be contrasted with the robust diagnostic method of Sarmiento and Bryan (1982). In the latter, the model tracer equations are adjusted by the addition of Newtonian relaxation terms. These relaxation terms hold the model temperature and salinity close to climatology, but are also associated with strong sources and sinks in the model potential temperature and salinity equations, with the result that the method is highly diabatic. The advantages of using an adiabatic adjustment technique are discussed in Section 2.3. In Section 2.2, we note that the robust diagnostic method is an example of “nudging” (the diabatic relaxation terms “nudge” the model potential temperature and salinity towards climatology). In the semi-prognostic method, on the other hand, the correction applied to the horizontal momentum equations does not have the form of a relaxation term. It follows that while the correction term does “nudge”, the direction of the nudging is not specified.

In this overview, we begin in Section 2.1 by describing the method as it appeared in Sheng et al. (2001). The interpretation of the method is discussed in Section 2.2, the advantages of an adiabatic approach in Section 2.3, and the distortion of the model physics and the use of the method as a diagnostic technique (Eden and Greatbatch, 2003) in Section 2.4. This naturally leads to a discussion of modified versions of the method in Section 2.5, following Eden et al. (2004). Some applications of the semi-prognostic method are shown in Section 3, illustrating the improvements in model performance. In Section 4,
we briefly describe the use of the semi-prognostic method as a technique for nesting models, and finally, in Section 5, we provide a summary and conclusions.

2. Formulation and interpretation

2.1. The standard method

The semi-prognostic method is applicable to hydrostatic models. For simplicity, we begin by considering models that use height coordinates in the vertical, that is the common z-coordinate, where \( z \) measures geopotential height from a reference level, or sigma-coordinate in which the model vertical coordinate varies between 0 at the ocean surface to \(-1\) at the ocean bottom (see Greatbatch and Mellor (1999) for a review of the different vertical coordinates used in models). Application to models using other vertical coordinate systems are discussed briefly in Section 2.7. As applied to height coordinate models, the semi-prognostic method involves replacing the instantaneous density variable in the model hydrostatic equation by a linear combination of the model density, \( \rho_m \), and the input density \( \rho_c \). Usually the latter is computed from climatological hydrographic data, but it might also be density from the same model (as when the method is being used as a diagnostic tool; Eden and Greatbatch, 2003) or a different model (as in a nested modelling system, e.g. Zhai et al. (2004)). For simplicity, in this and the next section, we assume that \( \rho_c \) is the climatological density unless stated otherwise. It follows that the model hydrostatic equation is changed from

\[ \frac{\partial \rho}{\partial z} = -g\rho_m \]  

(1)

to

\[ \frac{\partial \rho}{\partial z} = -g[z\rho_m + (1 - z)\rho_c], \]  

(2)

where \( p \) is the pressure variable carried by the model and \( z \) is a parameter with \( 0 \leq z \leq 1 \). If \( \rho = \rho(\theta, S, p_{\text{ref}}) \) is the model equation of state (here \( \theta \) is potential temperature, \( S \) is salinity), then

\[ \rho_m = \rho(\theta_m, S_m, p_{\text{ref}}) \]  

(3)

and

\[ \rho_c = \rho(\theta_c, S_c, p_{\text{ref}}), \]  

(4)

where \( \theta_m, S_m \) are the instantaneous, prognostic potential temperature and salinity variables carried by the model and \( \theta_c, S_c \) are the input climatological potential temperature and salinity data. It should be noted that (2) is applied at each time step, so that \( \theta_m, S_m \) are evolving with time, and \( \theta_c, S_c \) are usually seasonally varying. Here, we have assumed a simplified equation of state in which the pressure dependence is replaced by an approximation, \( p_{\text{ref}} \), to the actual pressure, with \( p_{\text{ref}} \) depending only on height, \( z \) (Dewar et al., 1998). However, there is no restriction that prevents using the full equation of state in which \( p_{\text{ref}} \) is replaced by the physical pressure, \( p^* \). (As noted below, \( p^* \) is not the same as the pressure variable, \( p \), carried by the model.)

We next note that if \( z = 1 \), the model is a pure prognostic model, and there is no influence from the input density, \( \rho_c \). On the other hand, putting \( z = 0 \) turns the model into a diagnostic model (see Greatbatch and Mellor (1999) for a discussion of “diagnostic” versus “prognostic” models). In diagnostic models, the model density, \( \rho_m \), plays no role in the model dynamics; in fact, in this limit, the potential temperature and salinity fields carried by the model act as passive tracers. In all the applications to be discussed, \( z = 0.5 \) (Sheng et al. (2001) provide some justification for this choice). The issue of choosing \( z \) is one we return to in Section 2.6.

In order to understand the effect of using Eq. (2), rather than the conventional Eq. (1), we first write Eq. (2) in the form

\[ \frac{\partial \rho}{\partial z} = -g\rho_m + g(1 - z)(\rho_m - \rho_c). \]  

(5)

We now divide the pressure variable, \( p \), into two parts \( p = p^* + \tilde{p} \), where \( p^* \) is the physical pressure and \( \tilde{p} \) is a new variable associated with the

---

1In the applications to be described, we have used gridded data. Use of the “smoothed” and “tapered” methods described in Section 2.5 helps to reduce the effects of mismatches between the input data and the bottom topography.
correction term, \( g(1 - z)(\rho_m - \rho_o) \). Since \( p^* \) is the physical pressure, it satisfies the standard hydrostatic equation

\[
\frac{\partial p^*}{\partial z} = -g \rho_m \tag{6}
\]

with

\[
p^* = g \rho_o \eta \tag{7}
\]

at \( z = 0 \). Here \( \eta \) is the upwards displacement of the sea surface from mean sea level, \( z = 0 \), and \( \rho_o \) is a representative density for sea water. It is important to appreciate that although the method is implemented in the model by adjustment of the density variable that is seen in the model hydrostatic equation, the method does not disturb the hydrostatic balance, Eq. (6), associated with the physical pressure, \( p^* \). In view of Eqs. (6) and (7), and since the model pressure variable, \( p \), satisfies \( p = g \rho_o \eta \) at \( z = 0 \), it follows that \( \hat{p} \) satisfies

\[
\frac{\partial \hat{p}}{\partial z} = g(1 - z)(\rho_m - \rho_o) \tag{8}
\]

with

\[
\hat{p} = 0 \tag{9}
\]

at \( z = 0 \). Substituting for the model pressure variable in the model’s horizontal momentum equations then gives

\[
\frac{\partial \mathbf{v}}{\partial t} + \mathbf{f} \times \mathbf{v} = -\frac{1}{\rho_o} \nabla p^* - \frac{1}{\rho_o} \nabla \hat{p} + \cdots \tag{10}
\]

where \( \mathbf{v} \) is the horizontal velocity vector, \( \mathbf{f} \) is a vector that points in the upwards vertical direction and has magnitude equal to the local value of the Coriolis parameter, and \( \nabla \) is the horizontal gradient operator. It follows from Eq. (10) that the semi-prognostic method adjusts the model by adding the term \(-1/\rho_o \nabla \hat{p} \) to the horizontal momentum equation carried by the model.

2.2. Interpreting the correction term

We first note that the correction term, \(-1/\rho_o \nabla \hat{p} \), is an “interactive” forcing, by which we mean that it depends on the model state, in particular the difference between the model density, \( \rho_m \), and the input density, \( \rho_o \). It is for this reason that use of the semi-prognostic method in its standard form distorts the model dynamics, an issue discussed further in Section 2.4. It should also be noted that the correction term is not equivalent to adding a relaxation term to the momentum equations (in this sense the semi-prognostic method differs from the standard form of “nudging”). In particular, the semi-prognostic method does not attempt to constrain the model horizontal velocity to remain close to an input horizontal velocity field, as, for example, in the suggestion of Holloway (1992). Indeed, there is no direct constraint placed on the model velocity field. Rather, the method relies on the impact of the correction term on the “balanced” flow; that is, that part of the flow that evolves on time scales long compared to \( 1/f \). To see this, consider the linearised horizontal momentum equation with the local time derivative term neglected; that is

\[
\mathbf{f} \times \mathbf{v} = -\frac{1}{\rho_o} \nabla p^* - \frac{1}{\rho_o} \nabla \hat{p} + \cdots \tag{11}
\]

We can then divide the horizontal velocity into two parts \( \mathbf{v} = \mathbf{v}_u + \mathbf{v}_c \) where

\[
\mathbf{f} \times \mathbf{v}_c = -\frac{1}{\rho_o} \nabla \hat{p}; \quad \mathbf{f} \times \mathbf{v}_u = -\frac{1}{\rho_o} \nabla p^* + \cdots \tag{12}
\]

Here, \( \mathbf{v}_c \) is the instantaneous correction to the velocity field associated with the semi-prognostic correction term, and is analogous to an Ekman contribution to the velocity (as discussed in Chapter 9 of Gill (1982)). This is quite different from the “nudging” approach discussed by Woodgate and Killworth (1997), or the assimilation technique of Oschlies and Willebrand (1996).

In view of Eqs. (8) and (9), it is clear that the correction term is zero at the surface, \( z = 0 \), but can be non-zero below. It is of interest to note that in writing Eqs. (7) and (9), it was assumed that the sea surface height variable, \( \eta \), carried by the model is the physical sea surface height. A different choice, such as putting \( \eta = \eta^* + \hat{\eta} \) by analogy with the decomposition applied to the model pressure variable, has no effect on the model solution. There is, therefore, an ambiguity as to exactly how the sea surface height variable carried by the model should be interpreted and, in the absence of
any compelling reason to do otherwise, we take $\eta$ to be the physical sea surface height.

It should also be noted that the vertical average of the correction term, $-1/\rho_o \nabla \hat{p}$, is, in general, nonzero. It follows that when there is variable bottom topography (as in the real ocean), there will be a contribution from this term to the barotropic mode (that is the forcing for the vertically averaged flow in the model) that has the same form as the JEBAR term (see Mertz and Wright (1992) and Greatbatch et al. (1991) for a discussion of JEBAR). This aspect of the correction term can play an important role in the model adjustment (although there is also an important baroclinic component). It is also of interest that because the correction term appears as a horizontal gradient, it does not appear explicitly in the vorticity equation (although it does contribute to the vorticity balance through the JEBAR-like term and through nonlinear and frictional coupling).

On the other hand, the divergence of the correction term appears in the equation for the horizontal divergence.

The correction term, $-1/\rho_o \nabla \hat{p}$, should be interpreted as a simple way to take account of processes that are missing from the model physics. Some authors have advocated implementing parameterisations for mesoscale eddy processes in the horizontal momentum equations of a model (e.g. Holloway, 1992; Greatbatch and Lamb, 1990; Greatbatch, 1998; Greatbatch and McDougall, 2003). However, because the correction term appears as a horizontal gradient, and has no curl, it does not correspond to a flux of potential vorticity, as in Greatbatch (1998) (in fact, the semi-prognostic method does not interfere directly with potential vorticity conservation). Rather, it has the same form as the term arising from the gradient of the eddy kinetic energy (EKE) that appears in the averaged horizontal momentum equations (see Section 5 in Greatbatch (1998)). If the correction term can be interpreted as including a contribution from the unresolved EKE, then an argument can be made for splitting the sea surface height variable, $\eta$, as $\eta = \eta^e + \eta^o$, with $\eta^o$ corresponding to the EKE of the unresolved flow at the surface. However, as noted above, making this interpretation does not in any way affect the model solution.

### 2.3. The advantages of an adiabatic approach

It was noted earlier that the robust diagnostic method has the disadvantage that it is strongly diabatic because of the source terms $-\gamma(\theta_m - \theta_C)$ and $-\gamma(S_m - S_C)$ that are added to the prognostic equations for potential temperature, $\theta$, and salinity, $S$, where here $\gamma$ is a Newtonian relaxation coefficient (Sarmiento and Bryan, 1982). These terms interact nonlinearly with the mixing processes that are resolved by the model physics (e.g. convective overturning) and make the method difficult to use in studies of passive tracers. More seriously, since diapycnal mixing is known to be weak in the ocean interior (e.g. Gregg, 1989; Ledwell et al., 1998), the circulation pathways should be constrained, to a good first approximation, to be in the neutral tangent plane (McDougall, 1987), whereas the relaxation terms allow strong diapycnal flow. An extreme case is the use of a sponge layer along the open boundaries of an ocean model, where model temperature and salinity are constrained to be close to observed values and strong water mass conversion takes place to mimic the inflow and outflow of water through what is often a closed boundary in the model code. Even in the coastal ocean, where explicit diapycnal mixing might be expected to be larger, observations suggest that diapycnal mixing can sometimes be surprisingly weak (Sundermeyer and Ledwell, 2001).

### 2.4. Distortion of the model physics

To see how the standard method distorts the model physics, it is helpful to cast the semi-prognostic method in terms of the shallow water equations. For this purpose, we consider a two density layer system in which the lower layer is infinitely deep and at rest (the so-called 1_2-layer model). The linearised equations for the active upper layer of mean depth $H$, including the semi-prognostic correction, are then

$$\frac{\partial u}{\partial t} - f v = -g \rho \frac{\partial h}{\partial x} - (1 - z) g \frac{\partial h}{\partial x} - Ku + f x,$$

(13)
\[
\frac{\partial v}{\partial t} + fu = -z g' \frac{\partial h}{\partial y} - (1 - z) g' \frac{\partial h_c}{\partial y} - Kv + F_y,
\]

(14)

\[
\frac{\partial h}{\partial t} + H \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = -\gamma h,
\]

(15)

where \( g' \) is reduced gravity, \( K \) and \( \gamma \) are Rayleigh friction and Newtonian damping coefficients (included to mimic damping effects), and \( F_x, F_y \) represent the model forcing (e.g. surface wind stress). Here, \( h \) is the downward displacement of the interface between the two layers, \( h_c \) corresponds to the input hydrographic data (that is, \( \rho_c \) in Eq. (2)), and it has been assumed that correction factor \( \alpha \) is spatially uniform.

It is clear from Eqs. (13)–(15) that the effect of using the standard method is to change the gravity wave speed, \( c \), from \( \sqrt{gH} \) to \( \sqrt{zgH} \). In other words, the gravity wave speed is changed by a factor of \( \sqrt{z} \). Likewise, the propagation speed for non-dispersive, long, baroclinic Rossby waves, which depends on \( c^2 \), is changed by a factor of \( z \).

It is also apparent from the modification to the horizontal pressure gradient term in Eqs. (13)–(15) that the mesoscale eddy field is damped by the method (although not necessarily eliminated, as shown in Fig. 2). The modifications to the method introduced by Eden et al. (2004) are designed to overcome these drawbacks, as discussed in the Section 2.5.

Another interesting aspect of Eqs. (13)–(15) is that the semi-prognostic “forcing” term, \((-\partial h_c/\partial x, -\partial h_c/\partial y)\) has the same form as atmospheric pressure forcing of the barotropic mode (see Gill, 1982, Chapter 9). In the limit of zero damping (i.e. \( K = \gamma = 0 \)), the equilibrium solution to this forcing corresponds to the inverse barometer solution with \( h = -(1 - z) h_c/x \) (implying sloping isopycnals) and \( u = v = 0 \). In reality, this equilibrium solution is never achieved, partly because of the damping that is present (combined with the long, decadal adjustment time associated with the baroclinic modes), but also because of the forcing that is applied to the barotropic mode through the equivalent of the JEBAR term (see above). A particularly interesting case is the limit \( z \to 0 \). In this limit the model becomes “diagnostic”; that is, the flow field depends on the input density field, \( \rho_c \), but not the model density field, \( \rho_m \). In this limit, the model velocities are non-zero, even when \( K = \gamma = 0 \), which appears to contradict the inverse barometer solution. It should be noted, however, that as \( z \to 0 \), so does the gravity wave speed \( \sqrt{zgH} \), which, in turn, implies that the time required to reach the inverse barometer solution becomes infinite, and the inverse barometer solution is never achieved.

Eden and Greatbatch (2003) show how the distortion of the model physics can be turned to advantage and used as a diagnostic tool. These authors describe perturbation experiments in which a model of the North Atlantic is perturbed by adding forcing corresponding to the positive and negative phases of the North Atlantic Oscillation (NAO) (see Greatbatch (2000) and Hurrell et al. (2003) for reviews of the NAO). Following the switch-on of the anomalous forcing, the meridional overturning circulation in the model adjusts on a decadal time scale to a new equilibrium. Eden and Greatbatch (2003) used the standard semi-prognostic method to determine whether the decadal adjustment is achieved by wave propagation (e.g. boundary waves and/or long, baroclinic Rossby waves) or by advective processes. To do this, they applied the standard method using the unperturbed model density field as the input density, \( \rho_c \), with \( z = 0.5 \). Repeating the perturbation experiment on the semi-prognostic model shows an adjustment with a very similar time scale to that found in the fully-prognostic model, showing that wave propagation cannot be important (since otherwise the adjustment time would be greatly increased), and establishing the importance of advective processes. By carrying out perturbation experiments with different signs for the perturbation forcing, the authors were also able to establish that anomalous advection plays a role in modulating the adjustment time scale (this is because anomalous geostrophic advection is reduced by a factor of \( \sqrt{z} \), as can be seen from Eqs. (13) and (14)).

Finally, we note that there is no direct effect on barotropic adjustment, although coastal trapped waves that depend on the density stratification can be affected.
2.5. Modified methods

It was noted above that use of the standard semi-prognostic method distorts the model physics and damps the mesoscale eddy field. To overcome these disadvantages, Eden et al. (2004) introduced a number of modifications to the standard method. In what they call the “smoothed” method, the correction term, \( g(1 - z)(\rho_m - \rho_c) \), in Eq. (5) is smoothed so that it applies only on large spatial scales (typically several hundred kilometers). Use of the “smoothed” method avoids damping of the mesoscale eddy field and also transient waves with spatial scales smaller than the smoothing scale. The “smoothed” method also has the advantage that small scale features in the input hydrographic data, and which are often unreliable, are also smoothed out and do not feed into the model.

To avoid interaction with the model boundaries, Eden et al. introduced the “tapered” method in which the correction term is tapered to zero near to sloping bottom topography. Tapering avoids spurious interaction between the input density field, \( \rho_c \), and the model bottom topography, an issue that plagued the early diagnostic calculations and was one of the original motivations for the robust diagnostic model of Sarmiento and Bryan (1982). Since both the smoothed and tapered methods involve spatial smoothing of the correction term, they are often used in combination.

In what Eden et al. call the “mean” method, the correction term \( g(1 - z)(\rho_m - \rho_c) \) is computed using only annual means for \( \rho_m \) and \( \rho_c \). Strictly, these should be running annual means (i.e. the annual mean of \( \rho_m \) should be updated every time step), but in practice, this may not be necessary (see Eden et al. (2004) for discussion on this point). The “mean” method avoids damping or distorting any physical processes with time scales shorter than annual (including the mesoscale eddy field and coastal trapped waves) and leaves the model free to compute it’s own seasonal cycle. Eden et al. show results obtained using combinations of these methods, and also when the correction term is applied only in a limited depth range (e.g. between 800 and 200 m depth, as seems to be sufficient for the North Atlantic model they consider).

Finally Eden et al. discussed the “corrected-prognostic method” in which the time history of the correction term from a spin-up run is stored and then substituted for the correction term in subsequent model runs. In this approach, the correction is no longer flow-interactive, and all the damping and distortion effects associated with the standard method are eliminated.

2.6. Choosing \( z \)

Sheng et al. (2001) applied the standard method to a limited area model of the northwest Atlantic Ocean. The use of the semi-prognostic method led to a dramatic improvement in the model performance, especially in the representation of the Gulf Stream and the “northwest corner” to the southeast of Newfoundland. Particularly pleasing is the improvement in the unconstrained potential temperature, \( \theta \), and salinity, \( S \), fields. One by-product is that the magnitude of the semi-prognostic correction, which depends on the difference between the model \( \theta_m, S_m \)-fields and the input climatology, \( \theta_c, S_c \), is much less than would be supposed by directly comparing the \( \theta, S \)-fields from the prognostic model run with the climatology (see Fig. 1).

In their calculations, Sheng et al. put \( z = 0.5 \), although they also note that the model results are not sensitive to choosing \( z \) in the range 0.4 to 0.6. Sheng et al. argued that \( z = 0.5 \) is a reasonable choice based on the so-called best linear unbiased estimator (BLUE) method. BLUE can be used to optimally blend two independent model runs. In Sheng et al., the diagnostic and prognostic model runs are taken as being independent, and the model velocities on the eastern Canadian shelf compared with the inventory of current meter data held at the Bedford Institute of Oceanography. The error variance based in the difference between monthly mean model (in Year 2) and observed currents is found to be about the same for both the diagnostic and prognostic model runs. BLUE can then be used to argue that the two models can be blended with equal weights. However, a difficulty in trying to apply BLUE to a semi-prognostic model is that in Eq. (2), it is the instantaneous model density, \( \rho_m \), that is “blended” with the...
climatological density, $\rho_c$. These two fields are not independent since $\rho_m$ depends on $\rho_c$ from earlier times during the model run. Ideally one would want to estimate $z$ optimally by choosing the value that gives the best fit between the model results and the observed data. Such an approach has not been attempted, but remains a challenge for future work.

Finally, there is no requirement for $z$ to be spatially uniform; indeed, it may only be necessary for $z$ to be non-zero in certain localised areas. This is also a topic for future research.

2.7. Models using different vertical coordinate systems

In models that use a generalised vertical coordinate, the hydrostatic equation always takes the form

$$\frac{\partial A}{\partial s} = -R.$$  \hspace{1cm} (16)

For example, in a standard $z$-coordinate model, $A$ is pressure, $s$ is height $(z)$, and $R = g\rho$, as in Eq. (1). Likewise, in the atmospheric model of
Hoskins and Simmons (1975), $A$ is geopotential height, $s \ln(\sigma)$ and $R = T$, where $T$ is temperature and $\sigma$ is pressure normalised by the surface pressure. In a model that uses density as its vertical coordinate (e.g. MICOM—see Bleck and Boudra, 1981), $A$ is the Montgomery potential and $R$ is the geopotential. In all these cases, the semi-prognostic method can be applied exactly as described above; that is by replacing (16) by

$$\frac{\partial A}{\partial s} = \alpha R_m - (1 - \alpha)R_c.$$ (17)

When the input data is from a climatological data set, it is important that $R_c$ has been averaged in the same coordinate system as the model. So, for example, if the semi-prognostic method is applied to the atmospheric model of Hoskins and Simmons (1975), the input temperature climatology, $T_c$, should consist of temperature data that has been averaged in the normalised-pressure coordinate system used by the model. Likewise, for application to the MICOM model, the input data should be the geopotential height averaged on constant density surfaces.

### 3. Applications of the method

Eden et al. (2004) applied the semi-prognostic method, and modifications of the method, to an eddy-permitting model of the North Atlantic Ocean driven by seasonally varying forcing. The model is part of the $z$-coordinate, FLAME group of models (Dengg et al., 1999) based on the GFDL MOM code (Pacanowski, 1995). The model spans the Atlantic Ocean between $20^\circ$S and $70^\circ$N, has a horizontal resolution of $1/3^\circ \cos \phi$ (where $\phi$ is latitude) and 45 levels in the vertical. Readers are referred to Eden et al. (2004) for the model details. In all cases, Eden et al. (2004) use $\alpha = 0.5$, and the input hydrographic data is a modified version of the climatology of Boyer and Levitus (1997). Fig. 1 compares the annual mean temperature at $50\text{m}$ depth east of Newfoundland in the climatology (Fig. 1a), and in the prognostic (Fig. 1b) and standard semi-prognostic (Fig. 1c) model versions. It is important to appreciate when looking at Fig. 1 that the potential temperature equation carried by the model is completely unconstrained, despite the improvement that can be seen in the modelled temperature field. Rather, it is the model momentum equations that have been modified, as in Eq. (10). In Fig. 1a, the “northwest corner” is the region of warmer water east of Newfoundland bounded on the west by the North Atlantic Current and the Grand Banks of Newfoundland, and on the north by the North Atlantic Current after it has turned abruptly eastward near $52^\circ$N. In the fully prognostic model, the “northwest corner” is completely missing. Instead of flowing northward to the east of the Grand Banks, the North Atlantic Current separates from the tail of the Grand Banks and heads directly towards Europe. Consequently, the water temperature in the “northwest corner” region is as much as $10^\circ$C cooler in the prognostic model run than in the climatology. Use of the standard semi-prognostic method (Fig. 1c) leads to a significant improvement, with much warmer water now being found in the “northwest corner” region. Fig. 1d shows the annual mean temperature from a $1/12^\circ \cos \phi$, eddy-resolving version of the FLAME model, from which it is clear that the semi-prognostic method captures the improvement in performance achieved by the increased resolution, but at a fraction of the computational cost. The semi-prognostic model also shows an improvement in the handling of the Gulf Stream separation in the model, and Eden et al. (2004) show that the improvement in the “northwest corner” region is also found using modified versions such as the “mean” and “smoothed/tapered” methods, and even “corrected-prognostic” versions. Sheng et al. (2001) find a similar improvement in the performance of their regional model of the North-west Atlantic (which also has $1/3^\circ$ resolution) using the standard version, and Zhang et al. (2004) show the advantage of the semi-prognostic method for carrying out multi-year simulations using a coupled ice/ocean model.

As noted in the introduction, the large error in temperature exhibited by the prognostic model in Fig. 1, and which translates into a similar error in SST, could negatively impact the representation of the atmospheric storm track in a coupled model (Hoskins and Valdes, 1990), and distort the uptake...
of tracers by the ocean model. As noted by Eden et al. (2004), the error in SST leads to a region of heat uptake by the ocean model in the “northwest corner” region. This region of heat uptake is not found in heat flux products generated by atmospheric models, but is a common feature in heat fluxes generated by ocean models due to the systematic error in the path of the North Atlantic Current.

Fig. 2 shows the EKE at 50 m depth in the prognostic model (Fig. 2a), and semi-prognostic models using the standard (Fig. 2b), the “mean” (Fig. 2c), and the “smoothed/tapered” combined (Fig. 2d) methods. The damping of the EKE by the standard method is clearly evident. Re-assuringly, the level of EKE is recovered (and even exceeded) in the “mean” and “smoothed/tapered” runs, including an enhanced penetration of eddy activity into the “northwest corner” region associated with the improved flow field there. The level of EKE in all these cases is lower than observed and is an unavoidable consequence of using insufficient horizontal resolution; the $1/12$ eddy-resolving model does much better in this respect.

![Fig. 2. EKE at 50 m depth.](image)

Fig. 2. EKE at 50 m depth. Contoured is the logarithm of EKE in m$^2$/s$^2$ at powers of 10 with interval 0.25: (a) the prognostic, FLAME model; (b) the standard, semi-prognostic model; (c) the “mean” semi-prognostic model; (d) the “smoothed and tapered” semi-prognostic model. From Eden et al. (2004).
On the other hand, the horizontal distribution of near surface EKE is in general agreement with observational estimates (Stammer et al., 1996).

Fig. 3 shows the poleward heat transport as a function of latitude in the standard semi-prognostic model, the prognostic version of the same model, the eddy resolving (1/12°) model, and the 1/3° model used for the DYNAMO model intercomparison study (Willebrand et al., 2001). Also shown for comparison are estimates based on observations. Disappointingly, the standard semi-prognostic model has the weakest and most unrealistic heat transport of all the cases shown. As noted by Eden et al., the standard semi-prognostic method being used here is causing a shortcut in the meridional overturning circulation associated with spurious upwelling at the latitude of the Gulf Stream separation. The problem results from a mismatch between the input density field, \( \rho_c \), and the model bottom topography, leading to spurious vertical velocities and an enhanced “Veronis effect” (Böning et al., 1995). The problem is cured by using the tapered method. This is illustrated in Fig. 4 which shows the poleward heat transport when modified versions of the semi-prognostic method are used (including “corrected-prognostic” model versions). This time only the “mean”, and the “mean, corrected-prognostic” versions show the reduced heat transport, these being the two cases shown in the figure that do not use tapering, while the other cases all give heat transports comparable to that from the 1/12° eddy-resolving model shown in Fig. 3.

Finally in this section, Fig. 5 compares the horizontal flow field in diagnostic, prognostic and standard semi-prognostic versions of the regional model of Sheng et al. (2001), showing the circulation over and around the Grand Banks of Newfoundland (the actual model domain is considerably bigger than the region shown in the figure). In the diagnostic and semi-prognostic models, the offshore branch of the Labrador Current follows the shelf break around the Banks and feeds southwestward to the Scotian shelf, much as observed (e.g. Loder et al., 1998). The prognostic model, on the other hand, is seriously in error, with flow in the opposite (incorrect)
direction along the southern edge of the Banks (Fig. 5c) in association with the poor representation of the Gulf Stream and North Atlantic Current systems. Such a large error, in turn, has a major effect on the water properties of the Scotian Shelf and Mid-Atlantic Bight to the south. The prognostic model also fails to capture the southward leakage of water through the Avalon Channel near the southeast coast of Newfoundland, a feature that is found in the diagnostic and semi-prognostic model versions, and in observations.

Fig. 5. The mean horizontal circulation at 61 m depth during February of Year 2 taken from semi-prognostic, diagnostic and prognostic versions of the regional model of Sheng et al. (2001). It should be noted that the actual model domain covers a much bigger area than that shown here.
4. Use of the semi-prognostic method as a nested-modelling technique

In this section we briefly describe the use of the semi-prognostic method as a means of transferring information between the different submodels comprising a nested modelling system. Often, detailed small-scale information is required within a limited area domain (e.g. in support of the oil and gas industry on the shelf), and it makes sense to nest a high-resolution, inner model inside a coarser-resolution, outer model. Sometimes, also, a model domain may contain choke points (e.g. the Gulf Stream separation region) where very high resolution would be an advantage, even though such high resolution is not necessary over the rest of the domain. The use of a two-way nested modelling system may then be desirable, with the inner, high-resolution model feeding back information to the coarser-resolution outer model.

A difficulty in developing a two-way nesting scheme is the compatibility problem (e.g., mass conservation) at the grid interface (Ginis et al., 1998). Furthermore, undesirable numerical noise may result from the change of the grid resolution at the grid interface and an additional damping is required (Fox and Maskell, 1995). Kurihara et al. (1979) proposed a nesting technique in which information is exchanged between the two models in a narrow zone (dynamic interface) near the grid interface. The Kurihara et al. (1979) scheme has been successfully applied to the ocean by Ginis et al. (1998). An alternative nesting technique developed by Oey and Chen (1992) is to embed a fine grid (inner model) inside a coarse grid (outer model) and use the inner model variables to update the outer model variables over the sub-region where the two grids overlap. Oey and Chen’s scheme has the advantage of allowing a two-way interaction not only at the grid interface but also directly inside the common subregion where the two grids overlap.

The use of the semi-prognostic method as a nesting technique is illustrated below (more detail can be found in Zhai et al., 2004). In the semi-prognostic approach, the focus is not on the exchange of information around the boundaries of the inner model, but rather on the exchange of information between the inner and outer models within the interior of the inner model domain. Indeed, in the applications shown in Fig. 6, there is nothing special about the treatment of the open boundaries of the inner model (the treatment is, in fact, the same as in Sheng et al. (2001), but using outer model output rather than climatological data as input). As illustrated in Fig. 6 (see below), a virtue of the semi-prognostic method is its ability to keep the inner model “on track” with the outer model (and vica versa). But perhaps the principal advantage of the semi-prognostic method over other nesting techniques is its simplicity and ease of implementation.

For the examples shown in Fig. 6, the outer model is the same as that described in Sheng et al. (2001) and has roughly 25 km resolution, while the inner model has the domain shown in the figure, and has a resolution of roughly 7 km. Climatological, seasonally varying forcing is used, as in Sheng et al. (2001). (The use of the same nested modelling system to simulate the passage of a severe storm over the shelf is described in Zhai (2004).) In the COW (Conventional One-Way nesting) case, the inner model is connected to the outer model only through the specification of the open boundary conditions for the inner model, these being taken from the outer model. In the OSP (Original Semi-Prognostic) case, in addition to applying open boundary conditions to the inner model as in the COW case, information is also transferred between the interiors of the models using the version of the semi-prognostic method described in Sheng et al. (2001); that is using Eq. (2),

$$\frac{\partial p}{\partial z} = -g[\alpha \rho_m + (1 - \alpha) \rho_c]. \tag{18}$$

where $\rho_m$ is the instantaneous model density variable and $\rho_c$ is now the density carried by the other submodel over their common domain (in this case, the domain of the high-resolution submodel). In two-way nesting (used to produce Fig. 6), Eq. (18) is used to transfer information from both the outer to the inner model, and back again, from the inner to the outer model. (The method can also be used for one-way nesting, in which case the feedback from the inner to the outer model is
suppressed). Writing Eq. (18) as

$$\frac{\partial p}{\partial z} = -g\rho_m + g(1-z)(\rho_m - \rho_c),$$  \hspace{1cm} (19)

the term $g(1-z)(\rho_m - \rho_c)$ is clearly the term responsible for the information transfer between the models. In the smoothed semi-prognostic (SSP) case, this term is smoothed spatially, as in the “smoothed” method described Section 2.5. For the case shown in Fig. 6, the correction term applied to the inner model is smoothed over 16 grid points (that is 112 km) so that inner model feels the outer model only on large spatial scales. (One consequence of this is that physical processes that are captured by the inner model on scales less than 112 km will not be adversely affected by the nesting procedure). On the other hand, no spatial filtering is applied to the correction term felt by the outer model.

The most striking feature of the instantaneous snapshots shown in Fig. 6 is the very different flow field in the COW case compared to both the outer model.
model and also the SSP and OSP cases. This difference is quite systematic and is also present in time-averaged fields (see Zhai et al., 2004). Indeed, in the COW case, the flow field within the inner model domain has drifted away from that in the outer model, even though the outer model is used to provide the boundary conditions for the inner model. The result is that the Gulf Stream is too far to the north and is banked up against the continental slope in an unrealistic fashion. In the OSP and SSP cases, on the other hand, where information is being exchanged between the inner and outer models in the interior of the inner model domain, the flow field is quite similar to that in the outer model, showing the usefulness of the semi-prognostic method for keeping the inner model on track with the outer model (and vice-versa when there is two-way nesting, as here). The greater detail apparent in the SSP case illustrates the advantage of using the smoothed method in order to gain the full advantage of the higher resolution used in the inner model. For an example of the nesting technique applied to the Caribbean Sea, readers are referred to Sheng and Tang (2004).

5. Summary and discussion

In the preceding sections, we have reviewed the semi-prognostic method, as it appeared in Sheng et al. (2001) and as subsequently modified by Eden et al. (2004). The method was introduced by Sheng et al. (2001) as a technique for adjusting hydrostatic ocean models to correct for systematic error. The results shown in Sheng et al. (2001), Eden et al. (2004) and in Section 3 of this paper confirm the success of the method, on both the regional (the eastern Canadian shelf, northwest Atlantic) and the basin (North Atlantic Ocean) scale. What is particularly pleasing is the improvement in the modelled temperature and salinity fields, despite the fact the tracer equations carried by the model are completely unmodified by the method. Rather, the correction is applied to the horizontal momentum equations, with the advantage that the method is adiabatic. The semi-prognostic method can be contrasted with the robust diagnostic method of Sarmiento and Bryan (1982), in which sources and sinks are added to model's potential temperature and salinity equations, making that method highly diabatic. Adiabaticity, on the other hand, ensures (i) that there is no compromise to the requirement that the flow be primarily in the neutral tangent plane in the ocean interior, and (ii) that the method is well-suited for use in tracer studies (e.g. Zhao et al., 2004). The semi-prognostic method is also easy to implement in model code, since all that is required is to adjust the density that is used in the model's hydrostatic equation (see Section 2.7 for models that do not use height coordinates in the vertical).

We noted in Section 2, that the standard method used by Sheng et al. (2001) distorts the model dynamics by reducing wave propagation speeds and damping the mesoscale eddy field. The modified methods introduced by Eden et al. (2004) overcome these drawbacks, while maintaining the benefits of the standard method (see Sections 2.4 and 2.5 for the details). We also noted that Eden and Greatbatch (2003) turn the distortion to advantage by using it to diagnose the important dynamical processes in a model of the North Atlantic Ocean.

The semi-prognostic method is a technique for transferring information into and between models. As such, it is really a data assimilation technique. The adiabaticity of the semi-prognostic method points to a certain kinship with the assimilation scheme of Cooper and Haines (1996). These authors describe a method for assimilating satellite altimeter data into a model that involves applying uniform vertical displacements to the model's subsurface isopycnals. As such, the Cooper and Haines method preserves water mass properties as well as the (planetary geostrophic) potential vorticity field on isopycnal surfaces, two properties it shares with the semi-prognostic method. It is interesting to speculate on the use of the semi-prognostic method as tool for ocean data assimilation. For example, the displaced isopycnal field of Cooper and Haines could be taken as the input hydrography, \( \rho_c \) in Eq. (2). Clearly, the use of the semi-prognostic method as an assimilation technique is a topic for future research. Further work is also required to perfect the use of the semi-prognostic method as a nested modelling
technique, as described in Section 4 (see Zhai et al., 2004), and on the modified methods described in Section 2.5 in a search for ways to further mitigate the distortion to the model physics inherent in the standard method (see Section 2.4). Another area where future research is required is in the choice of the parameter $a$ (see Eq. (2)), a topic discussed in Section 2.6. In particular, it may not be necessary for $a$ to be spatially uniform. Indeed, there are probably parts of the model domain where we can put $a = 1$, so that the model is purely prognostic in those regions. The spatial distribution of the semi-prognostic correction term (see Eqs. (5) and (8)) may also provide insight into model deficiencies, and may therefore provide useful information on how to improve the model physics.

Acknowledgements

This work has been supported by funding from NSERC, CFCAS and CICS in support of the Canadian CLIVAR Research Network and by NSERC, MARTEC (a Halifax based company) and the Meteorological Service of Canada (MSC) in support of the NSERC/MARTEC/MSC Industrial Research Chair. Conversations with Claus Böning, Klaus Fraedrich, Chris Hughes, Youyu Lu, Frank Lunkeit, Andreas Oschlies, Drew Peterson, Jürgen Willebrand and Dan Wright helped clarify our thinking about the semi-prognostic method and are gratefully acknowledged. Comments from two anonymous referees were helpful when revising the manuscript. We are also grateful to the MSC for the use of supercomputer facilities at the Canadian Meteorological Centre in Dorval, Quebec.

References


