Method

The sequential analysis begins with a state forecast, $\boldsymbol{\omega}_k^f$ at time t_k produced by a forecast model represented by the operator Ω. All observations (of temperature, salinity, sea-level, etc.) available at time t_k are collected into an observation vector $\boldsymbol{\omega}_k^o$. The sequential filter combines the forecast with the observations into an analysis, $\boldsymbol{\omega}_k^a$, which then provides the estimate of the true state. The analysis is then used as the initial condition for a subsequent forecast $\boldsymbol{\omega}_{k+1}^f = \Omega \boldsymbol{\omega}_k^a$.

The two-stage algorithm begins with a bias forecast from some earlier analysis $\beta_{k+1}^f = \mathbf{B}\beta_k^a$ where the bias model $\mathbf{B}$ has been constructed from an analysis of $\omega^f - \omega^o$ statistics. The results of *Chepurin et al. (2003)* indicate that half of the variance of the temperature $\omega^f - \omega^o$ in the upper tropical oceans may be reduced with a bias model that accounts for time-mean, annual cycle, and ENSO-related variability. Then at time t_{k+1} the bias and state analyses are computed in two stages:

$$\beta^a = \beta^f - \mathbf{L}\left[\omega^o - \mathbf{H}(\omega^f - \beta^f)\right], \tag{1a}$$

$$\omega^a = \tilde{\omega}^f + \mathbf{K}\left[\omega^o - \mathbf{H}\tilde{\omega}^f\right] \text{where } \tilde{\omega}^f = \omega^f - \beta^a \tag{1b}$$

where the gain matrices $\mathbf{K}$ and $\mathbf{L}$**,** which determine the impact of the observations, depend on terms such as the forecast error covariance (see *Chepurin et al. 2003* for the specification of K and L).

Specifying the (bias-corrected) forecast error covariances in a reasonable way is a key issue in data assimilation and is also a complex one. The presence of a near-surface mixed layer, for example, introduces discontinuities into the forecast error covariances. Within the mixed layer forecast errors are highly correlated in the vertical, while their vertical correlation may decrease substantially below the base of the mixed layer. The forecast error covariance of the depth of the mixed layer itself is correlated with forecast mixed layer temperature error (turbulence both cools and deepens the mixed layer). Geographic features can introduce discontinuities. Across the West Indies island chain, for example, forecast errors may be correlated horizontally at levels above the depth of the channels, and not below. Finally, the evolving flow field influences inhomogeneities. Forecast errors are more highly correlated along the axis of the Gulf Stream than across the Gulf Stream axis.

To account for these complex relationships we intend to begin with empirical models based on analysis of bias-corrected forecast-minus-observation statistics, some of which is described by *Carton et al. (2000a,b).*

Ocean model

The ocean model component of SODA-POP is the recently released Parallel Ocean Program POP2.0[1] with a 0.25x0.4 degree displaced pole grid (**Fig. 1**) and 50 vertical levels. This recent release of POP has several advantages including partial filled bottom cells allowing better vertical resolution and providing useful NetCDF output tools.

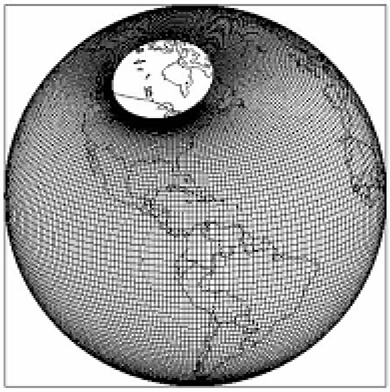

Fig. 1 Displaced pole grid. The horizontal resolution on the equator is 28km x 44km reducing to an approximately uniform 25km x 25 km in the western North Atlantic.

The use of a displaced pole opens up the option of resolving Arctic processes (SODA currently has a northern boundary at 62N). Bottom topography has been obtained from the 1/30 degree analysis of *Smith and Sandwell (1997)* (GTOPO30) with modifications for certain passages provided by Julie McLean (*McLean, personal communication, 2002*).

Vertical diffusion of momentum, heat, and salt are carried out using KPP mixing with modifications to address issues such as diurnal heating, while lateral subgrid-scale processes are modeled using bi-harmonic mixing. The new POP2.0 formulation does allow anisotropic horizontal viscosity, which has certain advantages as well (*Large et al., 2001*). The vertical levels we have chosen has 5m resolution near the surface and half of the vertical gridpoints (25) are in the upper 200m. In contrast, the grid-spacing in the deep ocean expands to 400m. Thus, the grid is weighted to resolve upper ocean processes (0-1000m), in keeping with our focus on seasonal to decadal variability. One efficient alternative to representing deep ocean processes is the model formulation of *Danabasoglu and McWilliams (2000)* where the temperature and salinity are constrained by their climatological values in the mid-depth and deep ocean.

Forcing: SODA1.0 is forced with daily averaged winds provided by the NCEP/NCAR reanalysis 1948-present *(Kistler, et al., 2001)*. The advantage of this choice over the monthly COADS analysis used in SODA includes improved temporal resolution and greater consistency with atmospheric reanalysis products. The disadvantage of this wind product is the presence of bias, particularly the weakness of the equatorial winds (see e.g. *Josey et al., 2002; Goswami and Sengupta, 2003*). We address bias in the stress by adding a steady spatially-dependent term to each stress component to correct for time-mean bias and multiplying by a steady spatially-dependent term to properly scale the

---

[1] (http://www.acl.lanl.gov/climate/models/pop/current_release/UsersGuide.pdf)

stress variance. Both of these terms are determined by comparison with Quickscat scatterometer wind stress during the years of overlap. Extending the winds back to the early 1940s will require merging reanalysis and COADS winds for a few years.

In addition to a time-mean bias, studies have shown that changes in the historical atmospheric measurement program have caused changes in the surface analysis (*Mo et al., 1995*). This includes a gradual increase in Northern Hemisphere upper air observations from 1948-1957. In 1973 there was another significant improvement associated with changes in quality control, and another in the mid- to late 1970s due to the introduction of satellite-based temperature retrievals.

SODA1.2 is forced by ERA40 daily-averaged winds available from ECMWF[2], which will cover the period from mid-1957 to 2001.

Surface freshwater flux provides a key forcing field for this reanalysis. For the period 1979-present we intend to use precipitation provided by the Global Precipitation Climatology Project[3] monthly merged product combined with evaporation obtained from bulk formula and the UNESCO river discharge transports (*Vörösmarty, et al. 1998*). Under the Arctic ice surface salinity will be relaxed to the monthly Polar science center Hydrographic Climatology 2.1[4] (*Steele et al., 2001*) in order to account for seasonal melting/freezing. Monthly polar sea ice coverage is based on satellite estimates of ice concentrations (*Parkinson, 2001*) for the period 1979-pres edited by John Weatherly of ERDC-CRREL. We do not plan to attempt a complete mass budget for sea ice (taking account of the relationship between changes in ice coverage and mixed layer salinity) except to the extent that the mass budget is reflected in the sea ice and salinity climatologies.

Surface heat flux boundary conditions are provided by a bulk formula for heat flux (which is reduced under polar ice). However, this boundary condition is relatively unimportant because of the use of near-surface temperature observations to update mixed layer temperatures (see below). Our preliminary tests suggest the analysis system requires about 15,000 GAUs to complete 50 years in the cheapest batch on the NCAR IBM-SP Bluesky.

Data The basic subsurface temperature and salinity data sets consist of approximately 7×10^6 profiles, of which two-thirds have been obtained from the World Ocean Database 2001 (*Boyer et al., 2002*) and are extended by operational temperature profile observations from the National Oceanographic Data Center\NOAA temperature archive, including observations from the TAO/Triton mooring thermistor array and ARGO drifters. The profile data is concentrated along commercial shipping lanes. Mixed layer temperature observations are available from the COADS surface marine observation set (*Diaz et al., 2002*). Satellite altimeter sea level from GEOSAT, ERS/1-2, TOPEX/POSEIDON and JASON is used beginning in the mid-1980s. Data checking for

---

[2] (http://www.ecmwf.int/research/era/)

[3] http://precip.gsfc.nasa.gov/

[4] http://psc.apl.washington.edu/Climatology.html

this analysis includes checks for duplicate reports and errors in the recorded position and time of observations, for static stability, for deviation from climatology, and checks on the relationship between temperature and salinity. Substantial quality control is already in the WOD2001. Our additional quality control (including buddy-checking, examination of forecast-minus-observation differences, and vertical stability) eliminates roughly 5% of the profiles.

Chlorofluorocarbons have been introduced into the global environment primarily since the 1960s as the result of various anthropogenic activities. Because they are extremely nonreactive, because different CFCs have been introduced at different rates, and because of the significant data archive that has accumulated, they provide an extremely useful tracer of the ventilation of the ocean and the movements of water masses (*Bullister, 1989; Dutay et al., 2002*). The CFCs are just one among a series of tracers (tritium/helium, transient tracers). As a first step we intend to introduce the chlorofluorocarbons (CFC-11, CFC-12) as passive tracers following the Ocean Carbon Modeling Intercomparison Project OCMIP[5] protocol (*Orr et al., 1999*), relying partly on CCSM for guidance. Although the results will be global our own analysis will focus on the North and tropical Atlantic where a number of previous data sets and observational studies are available for comparison (e.g. *Doney et al., 1997; Jenkins 1998*).

---

[5] http://www.ipsl.jussieu.fr/OCMIP/phase2/simulations/CFC/HOWTO-CFC.html