SODA: A Reanalysis of Ocean Climate

James A. Carton[1] and Benjamin S. Giese[2]

July 5, 2005

submitted to *Journal of Geophysical Research-Oceans*

[1]Department of Atmospheric and Oceanic Science
University of Maryland
College Park, MD 02742
carton@atmos.umd.edu

[2]Department of Oceanography
Texas A&M University
College Station, TX
b-giese@tamu.edu

Abstract

This paper describes the Simple Ocean Data Assimilation reanalysis of ocean climate variability. The reanalysis spans the 44-year period from 1958-2001, which complements the European Center for Medium Range Weather Forecasts ERA-40 atmospheric reanalysis. A second reanalysis experiment spans the period of QuikSCAT scatterometry from 2000-2004. The observation set is as complete as possible and includes the historical archive of hydrographic profiles supplemented by ship intake measurements, moored hydrographic observations, and remotely sensed SST and sea level. The forecast model utilizes Parallel Ocean Program physics with an average 0.25^{o}x0.4^{o}x40-level resolution. A sequential assimilation algorithm is used with a 10-day updating cycle. Because the introduction of altimeter sea level observations represents a major change in the observation base it is excluded from the basic reanalysis, while a parallel reanalysis experiment has been carried out including this data. The reanalysis and associated experiments are available online in monthly-averaged form, mapped onto a uniform 0.5^{o}x0.5^{o}x40-level grid. The reanalysis variables consist of three types, those that are well constrained by observations, those partly constrained by dynamical relationships to variables that are frequently observed, and those such as such as horizontal velocity divergence that are poorly constrained and may contain significant errors.

Evaluation of the reanalysis is carried out by comparisons with independent observations as well as by examination of observation-minus-forecast differences. Independent comparisons include comparisons with tide gauge sea level records and satellite altimetry, and comparisons with equatorial velocity in the Pacific. Sea level anomalies from their climatological monthly cycle during the nine-year period of overlap with the TOPEX/Poseidon altimeter record have correlations with a gridded form of the altimeter record

exceeding 0.7 throughout the tropics, decreasing to 0.4-0.5 at high latitudes. We examine the long-term bias in the reanalysis by comparing decadal average heat content to a recent, strictly observation-based, analysis. Both the reanalysis and the observation-based analysis show a gradual warming of the upper 700m of the North Atlantic with some cooling of the subpolar gyre. The tropical Pacific is characterized by an eastward shift of heat between the 1960s and 1990s. The reanalysis shows warming in the Southern Hemisphere that is not well resolved by the observations alone.

1. Introduction

In this paper we report on a multi-year Simple Ocean Data Assimilation (SODA) effort to reconstruct the changing climate of the global ocean, continuing the work of *Carton et al. (2000a,b)*. The results presented here cover the 44-year period 1958-2001 spanned by the European Center for Medium Range Forecasts ERA-40 atmospheric reanalysis and the 5 year period 2000-2004 spanned by the QuikSCAT scatterometer. The approach we take is to combine virtually all information available about the evolving fields of temperature, salinity, and sea level with a forward numerical simulation model driven by observed winds. The resulting ocean reanalysis provides an estimate of the evolving physical state of the ocean that represents an improvement over estimates based solely on either the sparse set of observations or numerical simulation.

Since becoming available in the mid-1990s atmospheric reanalyses have proved revolutionary in providing uniformly-mapped and regularly available samples of not only variables that are directly observed, but also indirect variables such as vertical velocity. However the atmospheric observing system was set up to support weather forecasting, not to produce multi-decadal reanalyses. As a result, the reanalyses are subject to systematic errors due to changes in the observing system. An example of these changes is the introduction of satellite observations of the atmosphere around 1979 (*Mo et al., 1995; Bengtsson et al., 2004*). Other changes have occurred in the conventional observations, such as in the average height of ship-board anemometers. These changes may well explain some of the trends evident in the reanalysis winds. For example, examination of ERA-40 surface stress in the tropics shows a weakening of the zonal stress in both the equatorial Pacific and Atlantic by 20% between the early 1960s and the early 1990s. The

errors in the winds project onto the circulation produced by ocean general circulation models driven by these winds. In addition to errors in winds, ocean circulation models are subject to internal errors due to unrepresentative ocean physics, inadequate numerics, internal variability, as well as errors in initial conditions. *Stockdale et al. (1998)* document many of these biases in their study of tropical circulation.

Our hope is that by using direct observations to correct these errors we will improve the ocean simulations. The ocean observing system, like its atmospheric counterpart, was not originally set up to produce ocean reanalyses. The main subsurface data set consists of temperature and salinity profile data whose distribution is concentrated along shipping routes in the Northern Hemisphere and in regions of commercial or military interest. Beginning in 1981 continuous global sensing of SST began with the deployment of infrared AVHRR (Advanced Very High Resolution Radiometer) sensors aboard NOAA polar orbiting satellites. Then in 1991 continuous observations of sea level began with the launch of a succession of satellite altimeters. While providing helpful information, the expanding observing system does introduce systematic errors into the ocean reanalysis.

In recent years there has been a proliferation of data assimilation algorithms that have been applied to problems in physical oceanography. These algorithms may be broadly divided into variational adjoint methods based on control theory and sequential estimation based on stochastic estimation theory, which for linear models leads to the Kalman Filter/Smoother. Comprehensive descriptions of alternatives in formulating algorithms and their relative advantages are provided by *Wunsch (1996)*, *Bennett (2002)*, and *Kalnay (2002).* The more sophisticated approaches, such as the four dimensional variational approach and full Kalman Filters suffer from high computational expense, increased sensitivity to the quality of the statistical assumptions, and in the latter case the potential for stability problems. To make more efficient

use of observations time-symmetric smoothing algorithms, which allow future observations to influence past events, are being explored but at the cost of more algorithmic complexity and computational expense. To address the high computational expense of these methods some researchers are exploring the use of ensembles to estimate forecast errors (e.g. *Evenson, 1994; Fukumori, 2002; Hamill, 2003; Ott et al., 2004*).

Here we adopt a sequential approach in which a numerical model provides a first guess of the ocean state at the update time and a set of linear Kalman equations is used to correct the first guess. This correction is based on estimates of the errors contained in the model forecast (the difference between the forecast value and true value of a variable such as temperature at a particular location and time) and in the observations. This work is a continuation of the reanalysis effort described by *Carton et al. (2000a,b)*, which first brought together all the components necessary for a global ocean reanalysis. That original effort used a forecast model based on the Geophysical Fluid Dynamics MOM2 numerics with modest horizontal and vertical resolution, resulting in considerable model bias. Indeed, the observation-minus-forecast differences were more than twice as large as in the current model. The original model grid excluded the Arctic Ocean, while the surface winds were smoothed with a monthly average. Additionally, the profile data set was nearly 30% smaller than is currently available, and SST data was included only in the form of an already gridded product. Since then, increases in computational power, the availability of the ERA-40 reanalysis, improvements to the data assimilation, the expansion of the ocean observing system and historical data sets, the need for a more complete domain and treatment of the hydrologic cycle, as well as interest in estimates of biogeochemical quantities, have all provided motivation for us to continue development of this ocean reanalysis.

2. Overview of the reanalysis system

The SODA analysis begins with a state forecast produced by a general circulation ocean model based on the Parallel Ocean Program numerics (*Smith et al., 1992*), with an average 0.25^{o}x0.4^{o} horizontal resolution and 40 vertical levels with 10m spacing near the surface[1]. The use of a displaced pole opens up the option of resolving Arctic processes (**Fig. 1**). Bottom topography has been obtained from the $1/30^{o}$ analysis of *Smith and Sandwell (1997)* with modifications for certain passages provided by Julie McClean (*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. Sea level is calculated diagnostically using a linearized continuity equation, valid for small ratios of sea level to fluid depth *(Dukowicz and Smith 1994).*

Daily surface winds are provided by the European Center for Medium Range Weather Forecasts ERA-40 reanalysis (*Simmons and Gibson, 2002*) for the 44-year period from January 1, 1958 to December 31, 2001. Surface freshwater flux for the period 1979-present is provided by the Global Precipitation Climatology Project monthly satellite-gauge merged product (*Adler et al., 2003*) combined with evaporation obtained from bulk formula. 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 temperature.

The assimilation cycle is carried out every 10 days. Following the Incremental Analysis Update methodology of *Bloom et al. (1996),* an analysis at time t is followed by a five-day simulation. On day t+5 the assimilation package is called to produce estimates of the temperature and salinity updates. Then the

[1] 5m, 15, 25, 35, 46, 57, 70, 82, 96, 112, 129, 148, 171, 197, 229, 268, 317, 381, 465, 579, 729, 918, 1139, 1625, 1875, 2125, 2375, 2624, 2874, 3124, 3374, 3624, 3874, 4124, 4374, 4624, 4874, 5124, and 5374m.

simulation is repeated from time t with temperature and salinity corrections added incrementally at every timestep. This digital filter has the advantage of maintaining a nearly geostrophic relationship between the pressure and velocity fields with a minimum excitation of spurious gravity waves. The procedure also reduces bias in the forecast model by 50%.

Averages of model output variables (temperature, salinity, and velocity) are saved at five-day intervals. These average fields are remapped onto a uniform global 0.5^{o}x0.5^{o} horizontal grid (a total of 720x330x40 gridpoints) using the *horizontal grid Spherical Coordinate Remapping and Interpolation Package* with second order conservative remapping (*Jones, 1999*) and saved in netcdf format (each 5-dy file is 157Mb, so a 44-year data set at this reduced resolution is 497Gb, the monthly data set is a more manageable 83Gb). The mapping shifts the locations of the temperature and horizontal velocity grids, which are offset in the model, to the same set of remapped gridpoint locations. In the Northern Hemisphere remapping is unable to retain exact three dimensional integrals required to calculate basin-averaged quantities such as meridional heat flux. Thus, such basin-average fluxes should be calculated on the original grid. In addition to the basic data set, a number of derived quantities such as heat content (see **Table 1)** are computed on the original grid and then remapped onto the 0.5^{o}x0.5^{o} grid.

In order to take account of changes in the surface forcing data sets as well as to help identify the impact of the observations, we examine four experiments (**Table 2**). Experiment #1 is a simulation forced by ERA-40 winds in which no ocean observations are used for updating. Experiment #2 is a data assimilation analysis forced by ERA-40 winds in which all temperature and salinity observations are used in the updating procedure. This is our control experiment and we will refer to it as our '*reanalysis*' in the later sections. Experiment 3 is a data assimilation analysis forced by the QuikSCAT satellite wind product for

the five year period 2000-2004. The differences between this Experiment and the reanalysis is briefly discussed in Section 5. Finally, Experiment 4 is a data assimilation analysis in which the updating data additionally includes all available altimeter data beginning in January, 1992.

3. Preparation of data for reanalysis

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; Stephens et al., 2001*) with online updates through December, 2004. This data set has been 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. This data set represents an increase of some 1.7 million profiles relative to the WOD1998 data used in many recent studies. The data coverage, shown as a function of depth and time in **Fig. 2**, has several characteristics that may affect the temporal variability of the results. Before the late-1960s the primary temperature profiling instrument was the Mechanical Bathythermograph. This device was limited to depths shallower than 285m and thus did not sample the main thermocline in many parts of the ocean. With the exception of the International Geophysical Year (1957-8) the total number of observations reaching a depth of 300m was less than 5,000/yr.

In the late-1960s a new Expendable Bathythermograph began to be used, extending the depth of temperature coverage to 450m and below. It was later discovered that these instruments suffered from biases in droprates and much effort since then has been devoted to correcting for these biases. We use the droprate correction described by *Stephens et al. (2001)*. In the early 1990s the tropical ocean moored

thermistor data increased dramatically with the expansion of the Tropical Atmosphere/Ocean Triangle Trans Ocean Buoy Network, while a decrease in observations later in that decade reflects the end of the World Ocean Circulation Experiment and reductions in the Volunteer Observing Ship program.

The data counts for salinity (**Fig. 2b**) are lower than those for temperature by a factor of 3-4 because this variable is not measured by the bathythermographs, but by several less common deeper profiling instruments. The reduction in salinity observations after 1995 is even more dramatic because of the delay in getting salinity observations into the data archives. Recently a new observing system called ARGO has been deployed, which makes profiles to 1000-2000m available in near-realtime, causing a profound increase in temperature sampling at 1000m and in salinity sampling at all depths.

All profile data are subject to nonrandom errors. A series of quality control filters has already been applied to temperature and salinity profiles prior to entry into the World Ocean Database 2001. These include, for example, location checks and checks on local stability. We apply some additional quality control to our data set including buddy-checking and examination of forecast-minus-observation differences, which eliminates an additional 5% of the profiles. Finally, examination of the results from previous reanalysis experiments allows us to eliminate a handful of additional outliers that have not been identified in previous objective quality control procedures.

In addition to the temperature profile data a large number of near-surface temperature observations are available both in the form of *in situ* observations (bucket and ship-intake temperatures from the COADS surface marine observation set of *Diaz et al., 2002*) and from satellite remote sensing. The *in situ* observations are made at a nominal depth of 1m and are treated as an estimate of mixed layer temperature.

Here we use the ungridded nighttime NOAA/NASA AVHRR Oceans Pathfinder Sea Surface Temperature Data, 'best' data, at 0.5^o resolution (*Vazquez et al, 1995*). These observations are available beginning January, 1981 and average 25,000 observations per week. Use of nighttime retrievals reduces the problem of correcting for skin temperature effects due to daytime heating (*Ewing and McAlister, 1960*). Although a smaller data set than surface temperature, mixed layer salinity observations average more than 10,000/yr since 1960 (*Bingham et al. 2002*).

Satellite altimeter sea level information is available from a number of satellites beginning with GEOSAT in the mid-1980s and continuing with a succession of satellites (ERS/1-2, TOPEX/POSEIDON and JASON) beginning in 1991. A number of corrections are applied to the altimeter data to account for geophysical effects such as delay due to tropospheric water, and for unresolved processes such as tides. We rely on the corrections applied to the CNES/AVISO (http://las.aviso.oceanobs.com/las/servlets/dataset) and use the 1^o along-track averaged data. Using the alongtrack averaging reduces observation error due to noise and unresolved processes and reduces the data set to a manageable size.

4. Data assimilation module

The two-stage sequential algorithm that we use is described by *Chepurin et al. (2005)*. The update step begins with a bias forecast from an earlier analysis $\beta_{k+1}^f = \mathbf{B}\beta_k^a$. The state, ω^a, and bias β^a analyses at time t_{k+1} are computed using linear filters based on the difference between observations ω^o and bias-corrected forecasts mapped onto the observation locations $\mathbf{H}(\omega^f - \beta^f)$:

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

$$\omega^a = \tilde{\omega}^f + \mathbf{K}[\omega^o - \mathbf{H}(\omega^f - \beta^a)] \quad (1b)$$

where the gain matrices **K** and **L,** which determine the impact of the observations, depend on terms such as the model error covariance. The state forecasts are provided by the ocean general circulation model. The bias forecasts are provided by an empirical model developed following *Chepurin et al. (2005)* based on a principal component analysis of the observation-minus-forecast differences and includes a steady term and seasonal terms. For the experiments listed in **Table 1** we set $\mathbf{L} = [0]$ (no bias correction).

Specifying the (bias-corrected) model 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 model error covariances. Within the mixed layer model errors are highly correlated in the vertical, while their vertical correlation may decrease substantially below the base of the mixed layer. Following *Carton et al. (2000a,b)*, the model error covariances of temperature and of salinity are assumed to behave as a Gaussian function of horizontal distance. The decay scales allow for zonally oriented error covariances in the tropics, becoming more isotropic towards midlatitude. The major and minor axes of the error covariance ellipses are rotated to allow for flow-dependence following the method of *Riishojgaard (1998)*, in order to reduce the error covariance perpendicular to coasts or strong fronts such as the Gulf Stream.

Model errors are allowed to be correlated in the vertical as well. Vertical correlation of model errors becomes important in allowing profile observations to influence the analysis below the base of the cast.

However, the vertical influence scale depends on the relative position of the base of the cast and the depth of the pycnocline. A lookup table allows us to reduce the vertical correlation of model error across depths of strong stratification.

Because there are fewer salinity observations than temperature observations, covariances between temperature and salinity errors play a critical role in updating the salinity fields. If the covariance is too strong water masses cannot evolve in time as observed, if the covariance is too weak new water masses will be created erroneously. We use results from a previous reanalysis to determine the temperature salinity error covariances. Comparisons in the vicinity of Station S (Bermuda) and Aloha (Hawaii) show that the relationship between temperature and salinity (**Fig. 3**) is similar to that in the climatological atlas (*Boyer et al., 2001*; *Stephens et al., 2001*). However, the upper and intermediate water masses at these locations do evolve in time, and this evolution is reflected in the scatter in the temperature/salinity relationship. At both locations the scatter is somewhat broader than the monthly climatology.

Because there are many more sea level observations than either temperature or salinity observations beginning in the 1990s our results are quite sensitive to the covariances between sea level errors and either temperature or salinity errors (the baroclinic component of sea level). These covariances are largest in the tropics where motion below the thermocline remains weak despite strong time-dependent flows in the upper ocean. This feature of the tropics implies a strong correlation between the height of sea level and the depth of isopycnal surfaces (a 1cm increase in sea level corresponds to a 1-2m change in the isopycnal depth. We use a truncated Taylor Series expansion to derive the model error covariances between temperature and salinity and sea level following *Carton et al, 1996*). An analysis of observation–minus–forecast differences from a previous reanalysis is then used to construct a lookup table.

We evaluate the impact of the observations on the resulting reanalysis by comparing the reanalysis to a simulation using the same model and surface forcing, but with the assimilation code switched off. The assimilation alters the mean stratification of the upper ocean, causing the thermocline to be sharper, and amplifies the seasonal cycle of mixed layer temperature and salinity somewhat, as these are too weak in the simulation. When the seasonal cycle is removed the assimilation is shown to increase variability throughout the ocean by 50%, and nearly doubling the variability in western boundary current regions. Much of this increased variability represents a strengthening of the oceanic mesoscale currents. However, low frequency variability is enhanced as well. The amplitude of sea level variations correlated with the NINO3 index, a measure of the strength of the El Nino signal, increases by 20% and its spatial structure is altered as well.

5. Reliability of analysis and preliminary results

In this section we present some comparisons of the reanalysis to mainly independent observation sets to provide the reader with a sense of how reasonable the results are. The comparisons are organized by timescale, beginning with mean transports at key locations, progressing to consideration of decadal changes in global temperature, monthly sea level, and finally to consideration of 5-dy average velocity.

Mean volume transports for several key passages and sections are listed in **Table 3**. Transports in the Gulf Stream and Kuroshio portions of the northern subtropical gyres are similar to observations. Much of the observed volume transport into the Gulf Stream system enters through the Florida Straits where an extensive cable monitoring program has been in place for decades (*Behringer and Larsen, 2001*). The

reanalysis volume transport through the Straits is 8Sv too low and is made up for water is passing to the east of the Bahamas before joining the Gulf Stream. This problem seems to be independent of details in the wind field, as an earlier reanalysis using the NCEP/NCAR reanalysis winds had similarly low transports at Florida Straits (and is a common feature of eddy-permitting simulations).

Transport in the Antarctic Circumpolar Current at Drake Passage is high (157Sv) in comparison with observations (134Sv, *Cunningham et al., 2003*). The processes controlling this transport are still a subject of debate. According to *Gent et al. (2001)* the 20Sv excess in modeled zonal transport with respect to uncertain observations could result from a moderate (3Sv) excess in meridional overturning circulation off the Antarctic shelf. In contrast to this excess, the warm water exchanges between the Pacific and Indian Oceans through the Indonesian Archipelago and between the Indian and Atlantic Oceans by way of the Agulhas Current are both similar to the observed estimates.

Decadal variations in the storage of ocean heat is a key reanalysis target because of its societal importance and because this variable could easily reflect bias in the ocean model. We choose for comparison the results of the observational study of *Levitus et al. (2005)*, who rely on the same temperature data set (WOA01). Comparison of heat content anomalies (**Fig. 4)** shows a gradual warming in both data sets, with the most pronounced warming occurring in the Atlantic and eastern Pacific. There is some cooling in the southwestern tropical Pacific (this result is influenced by the intense 1997 Nino) and in the subpolar gyre of the North Atlantic. There are also some differences between the two estimates. The reanalysis shows more warming in the Southern Hemisphere of the Atlantic and Indian Oceans as well as a band of cooling along the equator in the Indian Ocean.

We next examine monthly averaged reanalysis sea level, which in the tropics is closely related to monthly averaged heat content. Two independent sea level data sets are available, multi-decadal tide gauge sea level (which is compared in *Carton et al., 2005*) and satellite altimetry. Here we compare reanalysis sea level anomalies from their climatological monthly values with a gridded sea level product from AVISO for the 9 years 1993-2001 based on TOPEX/Poseidon altimetry.

In this comparison we remove the seasonal cycle in order to focus on the more challenging issue of reproducing nonseasonal changes. When the seasonal cycle is removed, the monthly reanalysis sea level and TOPEX/Poseidon altimetry are correlated in excess of r = 0.7 throughout the tropical Pacific and Indian Oceans, around 0.6 in midlatitude North and South Pacific , dropping to 0.4 – 0.5 in the tropical Atlantic and in other midlatitude regions. The high correlations in the tropical Pacific and Indian Oceans reflects the ability of the reanalysis to reproduce ENSO-related variability. The high correlations in the North Pacific at least partly reflect the high data coverage. Lower correlations in the tropical Atlantic have been a longstanding difficulty in reanalysis products (*Chepurin and Carton, 1999*) whose cause is still not clear. Lower correlations near western boundaries in midlatitudes reflects the presence there of intense mesoscale variability. To explore the similarity in distribution of this intense variability we next examine sea level in one such region: the western North Atlantic.

The magnitude of variability in reanalysis and TOPEX/Poseidon altimeter sea level is shown in **Fig. 5**. The mean position of the Gulf Stream in the reanalysis (indicated by the position of maximum variability) is quite similar to its mean position in the altimeter observations. The maximum reanalysis sea level variability is about 30% higher than the maximum TOPEX/Poseidon sea level variability. This apparent excess variability may reflect limitations in spatial sampling in the altimeter observations, as the altimeter

variability increases by 50% when computed using multiple satellites (*Le Traon et al., 2001*). A second variability maximum occurs in the eastern Gulf of Mexico due to the presence of meanders and rings in the Loop Current. The variability at the location of this second maximum is lower than observed, reflecting the somewhat lower frequency of these rings in the reanalysis than observed.

Finally, we examine the reanalysis currents at the equator where geostrophy, and thus observations of temperature do not play such a strong constraint. The location of the observations is 140^{o}W in the central Pacific, the longitude of maximum mean trade winds and maximum eastward Equatorial Undercurrent velocity (peaking at 100m depth). It is also the location of the western edge of the zone of 20-30 day Tropical Instability Waves. Surface currents show the westward South Equatorial Current intensifies in boreal fall in response to a strengthening of the trade winds. Thus a comparison at this location reflects a variety of different physical processes. The instrument used to record currents is an upward looking Acoustic Doppler Current Profiler (ADCP). Because of a limitation of the ADCP our shallowest comparison is at 25m, while the deepest comparison is at 150m (**Fig. 6**).

At 25m the observed record is limited to the years 1990-1995. The reanalysis shows a seasonally varying South Equatorial Current that is some 10cm/s stronger than observed, with a larger difference at the beginning of the record (in year 1990). The strength of the Tropical Instability Waves at this depth may also be somewhat stronger than observed, while the seasonally varying component of the currents is reasonably accurate. The core of the Equatorial Undercurrent is close to 100m depth. There the observed and reanalysis zonal currents both show strong, well correlated (r= 0.79, increasing to 0.86 when the time series are filtered with a two-month average) seasonal and interannual variations. The disappearance of the Undercurrent in response to the arrival of the 1997 El Nino is well-represented, although the recovery of

the Undercurrent appears to be delayed by about a month in the reanalysis. Finally, below the core of the Undercurrent at 150m the variability is much reduced.

Comparison of the reanalysis to Experiment #3 in which the winds are replaced with daily winds from the QuikSCAT scatterometer (**Fig. 6**) shows little change in the zonal velocities at 0^{o}N, 140^{o}W. Indeed, comparison of sea level from the reanalysis and from Experiment #3 during the two year period of overlap also shows a close correspondence. These results indicate that the change of wind products does not overwhelm the impact of the *in situ* observations and thus for some applications it might be possible to join the two reanalyses to obtain a longer record. This is perhaps not surprising as the ERA40 reanalysis actually includes observations from the ERS series of scatterometers beginning in 1991.

6. Summary

SODA is an ongoing program to develop reanalyses of the upper ocean for the benefit of climate studies as a complement to atmospheric reanalyses. Here we report on a set of reanalysis experiments spanning the 44-year period 1958-2001 with an extended reanalysis from 2000-2004. The data used in these reanalyses includes virtually all available hydrographic profile data as well as ocean station data, moored temperature and salinity time series, and surface temperature and salinity observations of various types. Two key satellite data sets are also included, SST from nighttime infrared observations and sea level from a sequence of altimeters. The data assimilation package is built around a general circulation ocean model using Parallel Ocean Program numerics with eddy-permitting resolution. The data assimilation algorithm is based on sequential analysis with a 10 day update cycle. Model error covariances are based on the results of *Carton et al. (2000a)* modified to allow flow-dependence. Flow-dependence in the error covariances is

important in that it reduces the influence of observations on the analysis when the two are separated by strong fronts such as the north wall of the Gulf Stream.

The results are made available in monthly average format on a uniform 0.5^{o}x0.5^{o}x40-level grid. In addition to the state variables (temperature, salinity, velocity, and sea level), we produce a set of two-dimensional derived quantities such as material properties on constant density surfaces. These reanalysis variables can be classified into three types, with type **A** consisting of those variables that are directly constrained by observations such as upper ocean temperature, type **B** consisting of those variables such as upper ocean geostrophic velocity that are strongly constrained by dynamical relationships to type **A** variables and with some observations directly affecting the variable. Type **C** variables are both poorly constrained by our set of observations and by strong dynamical relationships. Type **C** variables include deep ocean variables such as temperature, salinity, and velocity, or upper ocean variables such as horizontal divergence or momentum flux. Use of type **C** variables must be viewed with considerable caution.

In order to provide the user with some indications of the reliability of the reanalysis we provide several comparisons to independent data sets. *Carton et al. (2005)* compare the reanalysis sea level to a select set of 20 tide gauge records with an average length of 37 years, while here we present a comparison to the nine-year overlap with satellite altimetry. Correlations of altimeter and reanalysis sea level when the strong seasonal cycle is removed are highest in the tropics, exceeding $r = 0.7$, and generally decrease with increasing latitude. The root-mean-square variability of reanalysis sea level is somewhat larger than that obtained from a single altimeter, but in line with results from multiple altimeters.

The Pacific equatorial currents present an intriguing target for comparison because they are not as strongly constrained by geostrophy as currents elsewhere in the ocean, and because they play a central role in interannual climate variability. The comparison of reanalysis and observed zonal current at 0^{o}N, 140^{o}W in the central basin is reassuringly successful.

A primary concern in developing this reanalysis is the identification, impact, and reduction of forecast bias. One way we approach the identification of bias is through examination of the observation-minus-forecast difference data that are produced as part of the assimilation updating cycle. Our examination reveals, for example, that in the absence of bias correction (equation 1a) the observations act systematically to intensify the zonal slope of the thermocline in the equatorial Pacific relative to the model forecast. The observations also act to warm the forecast mixed layer in the Summer Hemisphere and cool the forecast mixed layer in the winter hemisphere, thus strengthening the seasonal cycle of mixed layer temperature. To determine the impact of bias on the ability of the reanalysis to represent decadal climate we compare decadal changes in 0/700m heat content from the reanalysis with corresponding data-only estimates of *Levitus et al. (2005)*. The comparison of the two sets of estimates shows that the spatial structure of the decadal heat content anomalies is reassuringly similar throughout the northern oceans where the data coverage is more extensive. In the southern oceans where the data coverage is sparse the two sets of estimates diverge. There the reanalysis shows enhanced heat storage while the data-only estimate shows little change between decades.

As has the ocean observing system, the meteorological observing system has changed substantially in recent decades. In order to evaluate the impact of changes in the meteorological observing system on the reanalysis we carry out a reanalysis experiment for the five year period 2000-2004 using the same model

and observation set as in the reanalysis, but with a different wind data set, in this case produced by the QuikSCAT scatterometer. The two-year period of overlap between the reanalysis, which ends in year 2001 and this reanalysis experiment shows some, but not an overwhelming, impact of the change in wind field on the circulation in the tropics, and diminished impact at higher latitudes.

The effort described here is ongoing. We anticipate devoting continued attention to the problem of model forecast bias and to mitigating the effects of spatial inhomogeneities and temporal changes in the atmospheric and oceanic observing systems. Aspects of the physical model that need attention include the role played by ice in the freshwater balance at high latitude (currently largely ignored). The inclusion of passive geochemical tracers such as the Chlorofluorocarbons and tritium/^{3}He will provide important additional information regarding processes such as water mass formation, entrainment, and mixing.

Acknowledgements

This effort has relied on many people. We gratefully acknowledge the contributions of our colleagues G. Chepurin to the coding and data set preparation and X. Cao who carried out the simulation. Without their efforts this work would not be possible. J. McClain of Scripps provided critical assistance with the numerical model. We depend on a number of individuals and institutions who maintain basic data sets. S. Levitus and the National Oceanographic Data Center/NOAA provided the quality-controlled hydrographic data, R. Reynolds provided the in situ and satellite SST observations. Altimetry has been obtained from the Archiving, Validation, and Interpretation of Satellite Oceanographic data (AVISO) website. Surface winds have been obtained from the European Centre for Medium Range Weather Forecasts (ECMWF), while scatterometer winds are from the Physical Oceanography DAAC of the Jet Propulsion Laboratory, and rainfall has been provided by the Global Precipitation Climatology Project. Computer resources have been provided by the National Center for Atmospheric Research and by the Office of Naval Research under a High Performance Computing Initiative grant to J. McClain. Finally, we would like to express our gratitude to the National Oceanic and Atmospheric Administration, the National Aeronautics and Space Administration and particularly to the National Science Foundation for providing key support for this long-term effort.

References

Adler, R. F., G. J. Huffman, A. Chang, R. Ferraro, P.-P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, and E. Nelkin, 2003: The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation analysis (1979-present). *J. Hydrometeorol.*, **4**, 1147–1167.

Baringer, M., and J.C. Larsen, 2001: Sixteen years of Florida current transport at 27^{o}N, *Geophys. Res. Letts.*, **16**, 3179-3182.

Bengtsson, L. S. Hagermann, and K.I. Hodges, 2004: Can climate trends be calculated from reanalysis data?, *J. Geophys. Res.*, **109**, D11111, doi:10.1029/2004JD004536.

Bennett, A. F., 2002: Inverse modeling of the ocean and the atmosphere, Cambridge University Press, Cambridge, UK, 352pp.

Bingham, F. M., S. D. Howden, and C. J. Koblinsky, 2002: Sea surface salinity measurements in the historical database, *J. Geophys. Res.*, **107**, 8019, doi:10.1029/2000JC000767.

Bloom, S.C., L.L. Takacs , A.M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*,**124**, 1256-1271.

Boyer, T.P., C. Stephens, J.I. Antonov, M.E. Conkright, , L.A. Locarnini , T.D. O'Brien, and H.E. Garcia, 2002: World Ocean Atlas 2001, Volume 2: Salinity. S. Levitus, Ed., NOAA Atlas NESDIS 49, U.S. Government Printing Office, Wash. D.C., 165 pp..

Bryden, H.L., L.M. Beal, and L.M. Duncan, 2003: Structure and transport of the Agulhas Current and its temporal variability. submitted to the *J. Phys. Oceanogr*.

Carton, J.A., B.S. Giese, X. Cao, and L. Miller, 1996: Impact of TOPEX and thermistor data on retrospective analyses of the tropical Pacific Ocean, *J. Geophys. Res.* , **101**,14,147-14,159.

Carton, J.A., G.A. Chepurin, X. Cao, and B.S. Giese, 2000a: A Simple Ocean Data Assimilation analysis of the global upper ocean 1950-1995, Part 1: methodology, *J. Phys. Oceanogr*., **30,** 294-309.

Carton, J.A., G.A. Chepurin, and X. Cao, 2000b: A Simple Ocean Data Assimilation analysis of the global upper ocean 1950-1995 Part 2: results, *J. Phys. Oceanogr*., **30,** 311-326.

Carton, J.A., B.S. Giese, and S. A. Grodsky, 2005: Sea level rise and the warming of the oceans in the SODA ocean reanalysis, *J. Geophys. Res.*, in press.

Chepurin, G.A., and J.A. Carton, 1999: Comparison of retrospective analyses of the global ocean heat content, *Dynam. Atmos. Oceans*, **29**, 119-145.

Chepurin, G.A., J.A. Carton, and D. Dee. 2005: Forecast Model Bias Correction in Ocean Data Assimilation, *Monthly Weather Review*, **133**, 1328–1342.

Cunningham, S. A., S. G. Alderson, B. A. King, and M. A. Brandon, 2003: Transport and variability of the Antarctic Circumpolar Current in Drake Passage, *J. Geophys. Res.*, **108**, 8084, doi:10.1029/2001JC001147.

Diaz, H., C. Folland, T. Manabe, D. Parker, R. Reynolds, and S. Woodruff, 2002: Workshop on Advances in the Use of Historical Marine Climate Data. *WMO Bulletin*, **51**, 377-380.

Dukowicz, J., and R. D. Smith, 1994: Implicit free-surface method for the bryan-cox-semtner ocean model. *J. Geophys. Res.*, **99**, 7991–8014.

Evensen G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10,143-10,162.

Ewing G., and E. D. McAlister, 1960: On the thermal boundary layer of the ocean, *Science*, **131**, 1374.

Gent, P., W. Large, and F. Bryan, 2001: What sets the mean transport through Drake Passage? *J. Geophys. Res.*, **106**, 2693–2712.

Hamill, T.M., 2003: Ensemble-based atmospheric data assimilation: A tutorial. University of Colorado and NOAA-CIRES Climate Diagnostics Center, Boulder, CO 46pp.

Hogg, N.G., 1992: On the transport of the Gulf Stream between Cape Hatteras and the Grand Banks. *Deep-Sea Res.*, **39**, 1231-1246.

James, C., M. Wimbush, and H. Ichikawa, 1999: Kuroshio meanders in the East China Sea, *J. Phys. Oceanogr*, **29**, 259-272.

Jones, P.W., 1999: First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates. *Monthly Weather Review*, **127**, 2204–2210.

Kalnay, E., 2002: Atmospheric modeling, data assimilation, and predictability, Cambridge University Press, Cambridge, UK, 342pp.

Levitus S, J. Antonov J, and T. Boyer, 2005: Warming of the world ocean, 1955-2003, *Geophys. Res. Letts.*, **32**, Art. No. L02604.

Le Traon, P.Y., G. Dibarboure, and N. Ducet, 2001: Use of a High-Resolution Model to Analyze the Mapping Capabilities of Multiple-Altimeter Missions, *J. Atmos. and Oceanic Tech.*,**18**, 1277–1288.

Meyers, G., R.J. Bailey and A.P. Worby, 1995: Geostrophic transport of Indonesian throughflow. *Deep-Sea Res.*, **42**, 1163-1174.

Mo, K. C., X. L. Wang, R. Kistler, M. Kanamitsu, and E. Kalnay, 1995: Impact of satellite data on the CDAS-Reanalysis system, *Mon. Wea. Rev.*, **123**, 124-139.

Riishojgaard, L. P., 1998: A direct way of specifying flow-dependent background error, *Tellus*, **50A**, 42-57.

Simmons, A.J., and J.K. Gibson, 2000: The ERA-40 Project Plan, Series #1, ECMWF, Shinfield Park, Reading, UK, 63pp.

Smith, W. H. F., and D. T. Sandwell, 1997: Global seafloor topography from satellite altimetry and ship depth soundings, *Science*, **277**, 1957-1962.

Smith, R.D., J.K. Dukowicz, and R.C. Malone, 1992: Parallel ocean general circulation modeling, *Physica D*, **60**, 38–61.

Sprintall, J., S. Wijffels, A.L. Gordon, A. Ffield, R. Molcard, R. Dwi Susanto, I. Soesilo, J. Sopaheluwakan, Y. Surachman, and H.M. van Aken, 2004: INSTANT:A New International Array to Measure the Indonesian Throughflow, *Eos*, **85**, pp. 369 and 376.

Stephens, C., J. I. Antonov, T. P. Boyer, M. E. Conkright, R. A. Locarnini, T. D. O'Brien, H. E. Garcia, 2002: World Ocean Atlas 2001, Volume 1: Temperature. S. Levitus, Ed., NOAA Atlas NESDIS 49, U.S. Government Printing Office, Wash., D.C., 167 pp., CD-ROMs.

Stockdale, T.N., A.J. Busalacchi, D.E. Harrison, and R. Seager, 1998: Ocean modeling for ENSO. *J. Geophys. Res*., **103,** 14,325-14355.

Vazquez, J. and A. Tran, R. Sumagaysay, E. A. Smith and M. Hamilton, NOAA/NASA AVHRR Oceans Pathfinder Sea Surface Temperature Data Set User's Guide Version 1.2, JPL Tech. Rep., 53 pp., 1995.

Wunsch, C., 1996: The Ocean Circulation Inverse Problem. Cambridge University Press. Cambridge, UK, 442 p.

Table 1 Variables available in the SODA1.4.2 monthly netcdf data set. Unless otherwise specified, fields are two dimensional. Type (defined in text) refers to variables only in the upper 1000m.

Variable	Type	Description
Temp (40 levels)	A[1]	temperature
Salt (40 levels)	B[1]	salinity
U,V (40 levels)	B[1]	horizontal velocity
TAUX, TAUY		ERA-40 wind stress components mapped onto the analysis grid[2].
HC500	A	Average temperature in the upper 500m
HC125	A	Average temperature in the upper 125m
SSH	B	Sea surface height
DYNHT	B	Dynamic height 0/1000db
PV_SIGMA25	B	Potential Vorticity ($f(\partial\rho/\partial z)/\rho_o$) on the σ_{25} isosurface
T_SIGMA25	B	temperature on the σ_{25} isosurface
S_SIGMA25	B	salinity on the σ_{25} isosurface
D_SIGMA25	B	depth of the σ_{25} isosurface
T_SIGMA26.5	B	temperature on the $\sigma_{26.5}$ isosurface
S_SIGMA26.5	B	salinity on the $\sigma_{26.5}$ isosurface
D_SIGMA26.5	B	depth of the $\sigma_{26.5}$ isosurface

[1]for levels 1-23. These variables become Type C for levels 24-40.
[2]provided for the user's convenience.

Table 2 List of Experiments currently available.

#	Name	Period	Winds	Assimilation Data
1	Simulation1.4.0	1958-2001	ERA-40	No updating.
2	SODA1.4.2	1958-2001	ERA-40	Temperature and salinity observation set used.
3	SODA1.4.3	2000-2004	QuikSCAT	As in SODA 1.4.2
4	SODA1.4.4	1992-2001	ERA-40	As in SODA 1.4.2 with the addition of ERS1/2, TOPEX/Poseidon, and JASON altimetry.

Table 3 Time-mean volume transports (10^6 m^3/s) through some major passages for the ocean reanalysis (Experiment #2).

Passage	***Observations***	***Reanalysis***
ACC-Drake Passage	137 ± 8	157
(*Cunningham et al., 2003*)	134 ± 21[1]	
Kuroshio off Shikoku Island (34°N)	42	39
(*James et al., 1999*)		
Gulf Stream off Cape Hatteras (35°N)	45	45
(*Hogg, 1992*)		
Florida Straits[2] (26°N)	32	24
(*Behringer and Larsen, 2001*)		
Agulhas (33°S)	70 ± 4	68
(*Bryden et al. 2003*)		
Indonesian Throughflow	12[3]	15
(*Meyers et al., 1995*)		

[1]Second estimate derived from summarizing previous estimates. Uncertainty given here is the average of the range of uncertainties reported.

[2]Reanalysis transport is computed between Florida and Cuba, while the observations are made between Florida and the Bahamas.

[2]Observed transports are still quite uncertain according to *Sprintall et al. (2004*). The reanalysis transports do not include the transport through Sunda Straits.

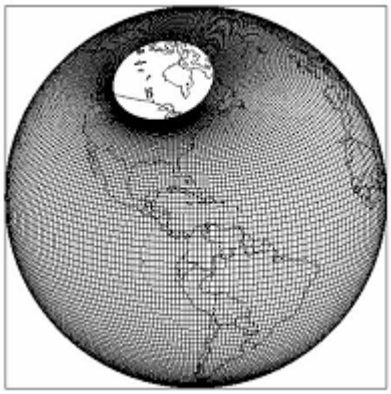

Fig. 1 Model grid showing displaced-pole geometry.

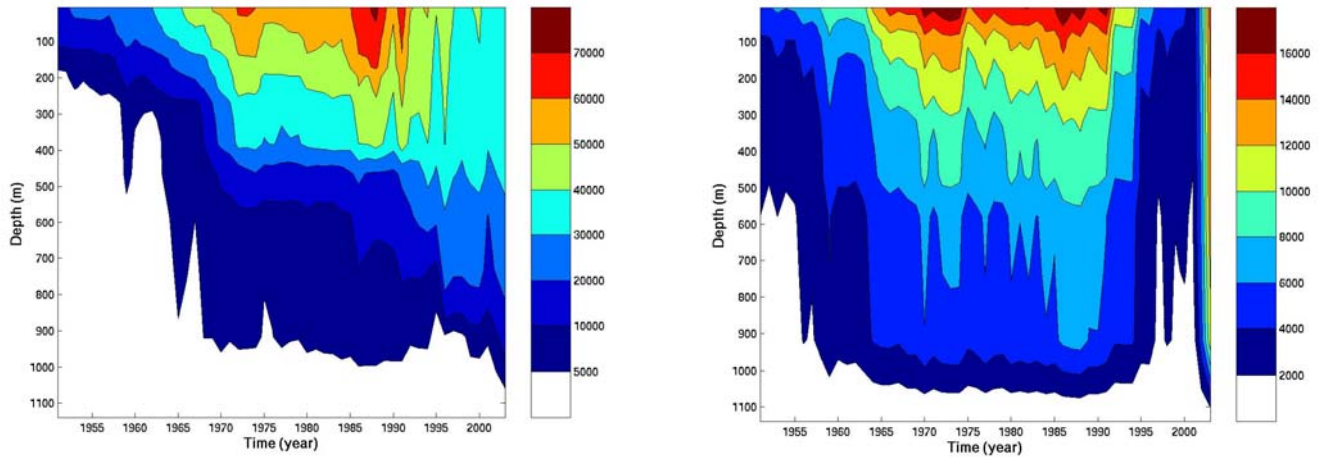

Fig. 2 Contours of the number of annual observations of (a) temperature and (b) salinity versus depth based on the World Ocean Database 2001 augmented with observations since 1995 from the Global Temperature Salinity Profile Program (see text for discussion). Data is limited to the upper 1000m

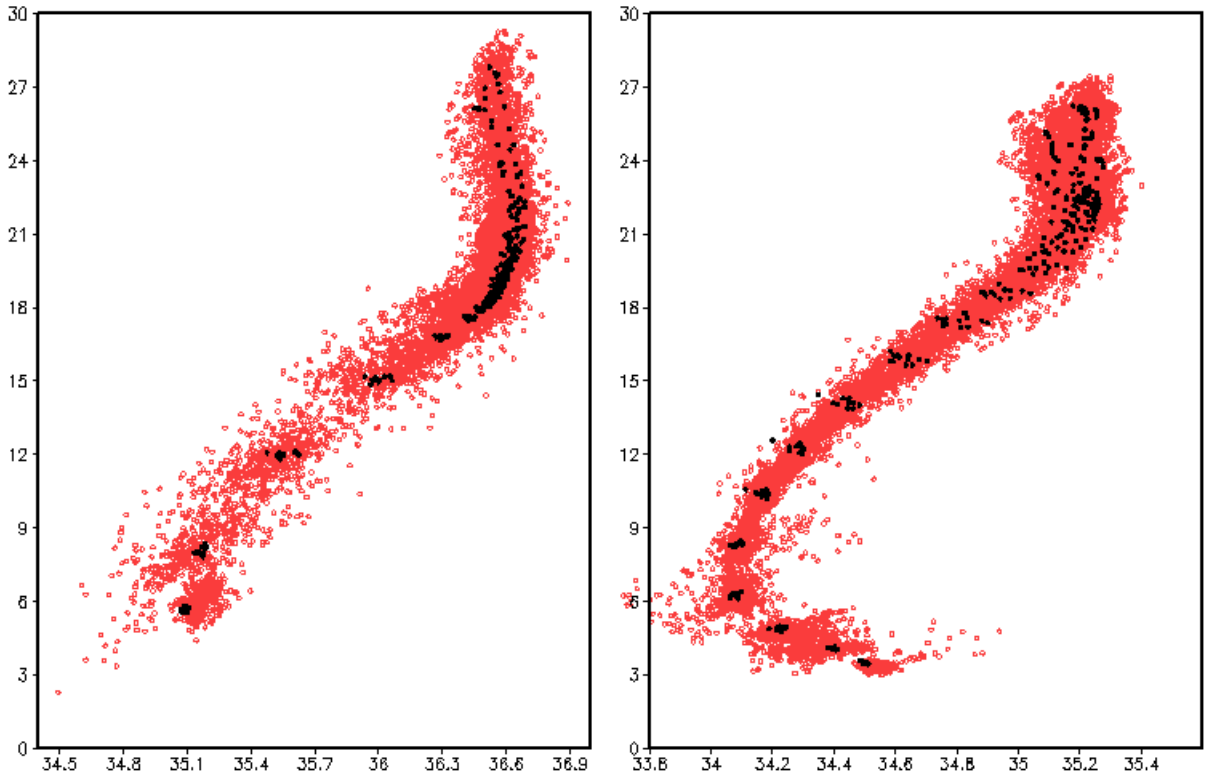

Fig. 3 Scatter diagrams of reanalysis potential temperature versus salinity to 1000m depth at two stations, a) Bermuda (30^{o}N, 64^{o}W) and b) Hawaii (23^{o}N, 158^{o}W). Climatological monthly temperature/salinity from WDB01 evaluated at the model levels is superimposed.

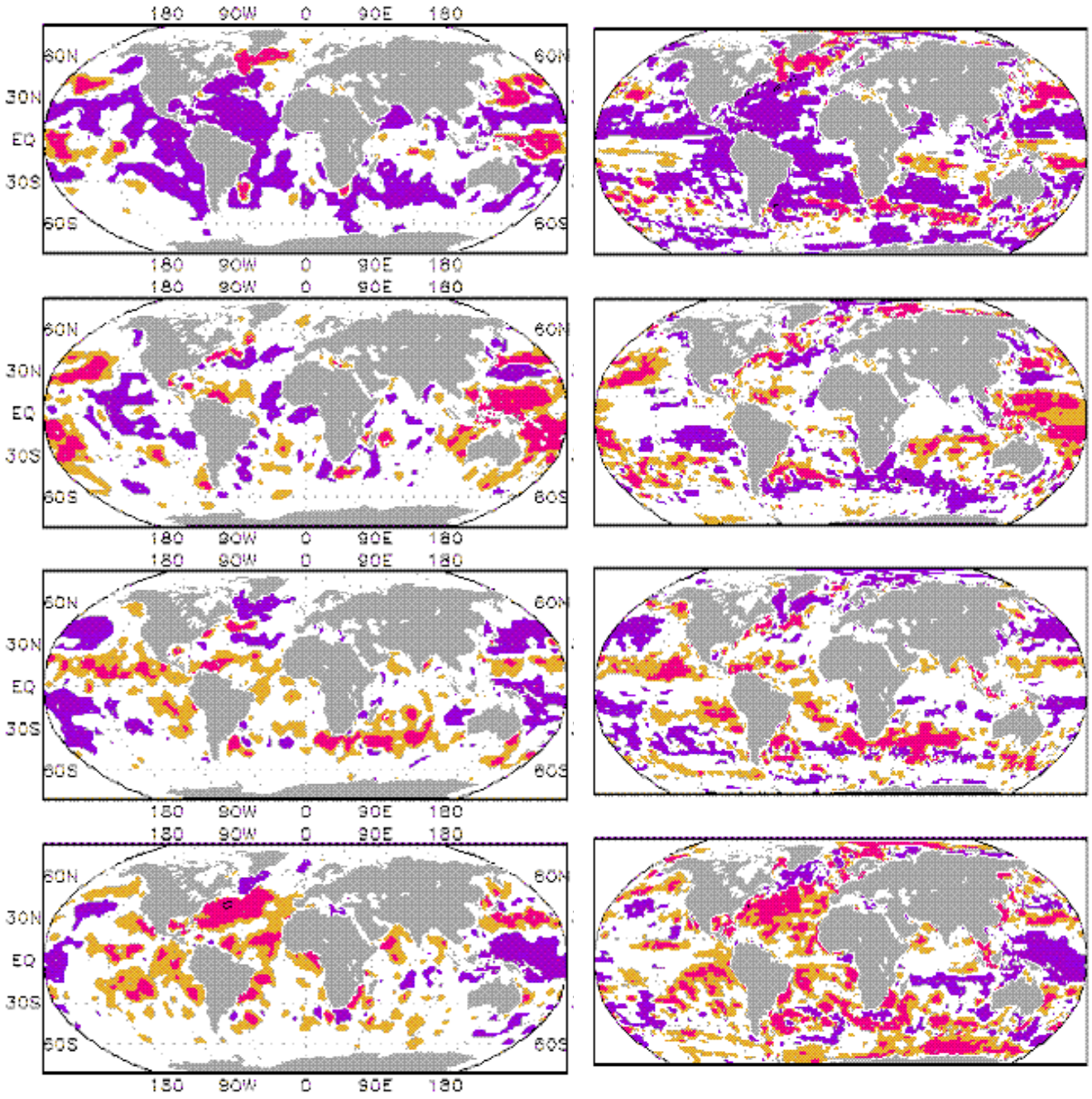

Fig. 4 Decadal anomalies of 0/700m averaged potential temperature (^{o}C) relative to the 40-year mean for: 1960-1969, 1970-1979, 1980-1989, and 1990-1999. The left four panels show results from the observational analysis of *Levitus et al. (2005).* The right four panels show results from this reanalysis. Note the gradual warming in the subtropical and midlatitude Atlantic, as well as the presence of strong decadal variability.

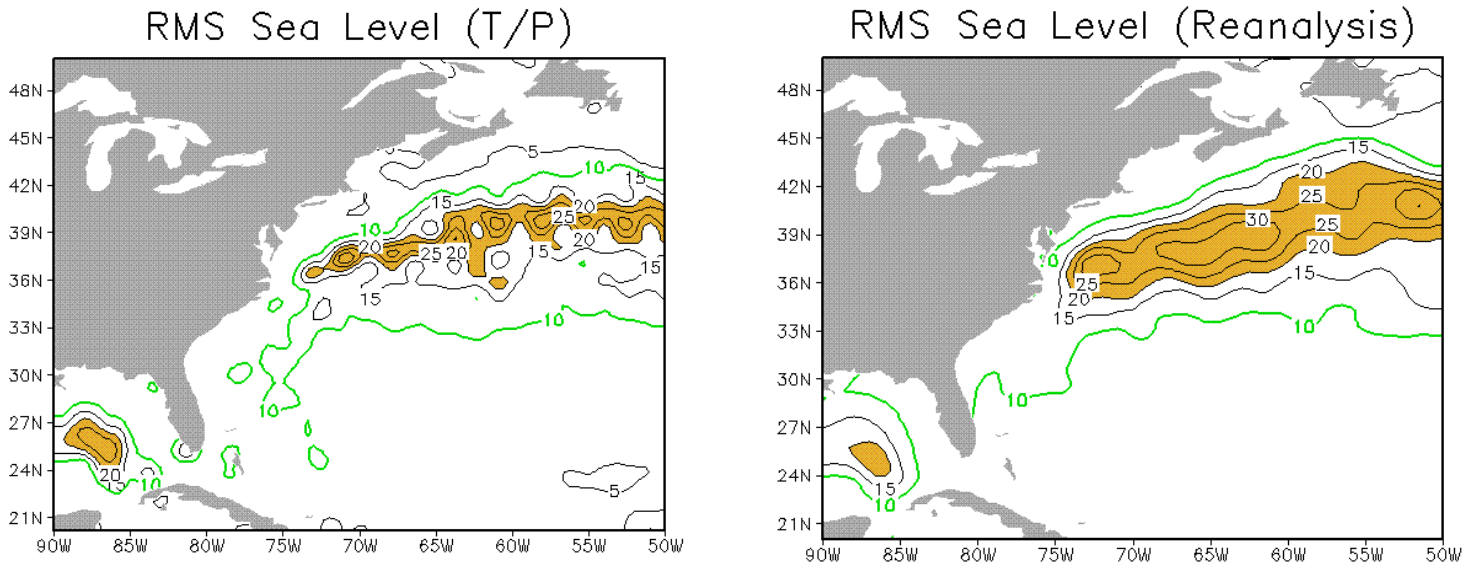

Fig. 5 Root-mean-square sea level variability from nine years (1993-2001) of monthly smoothed TOPEX/Poseidon altimetry and reanalysis sea level. When the recent combined altimetry from multiple satellites is used observed sea level variability increases by 50%.

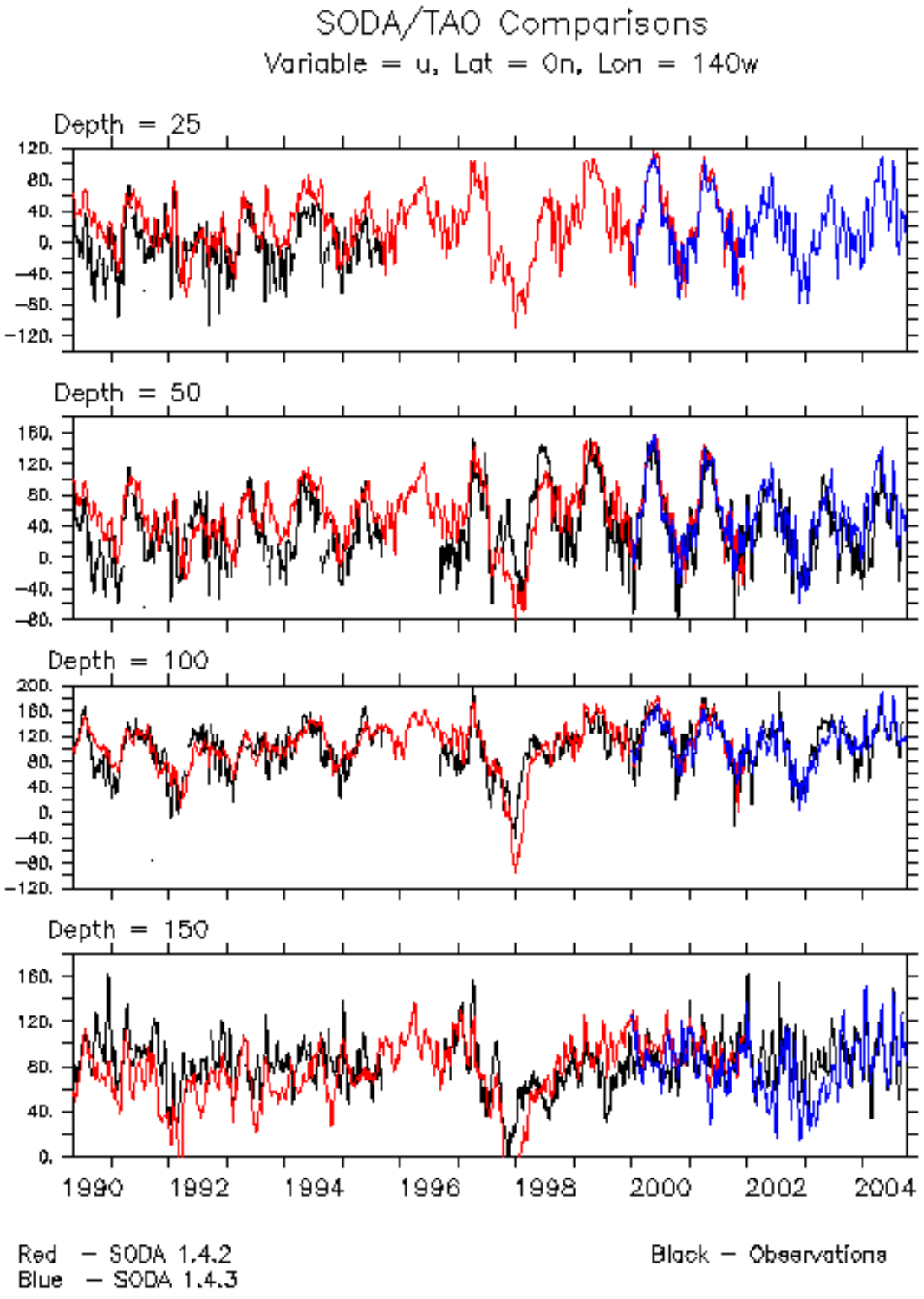

Fig. 6 Zonal velocity at 0°N, 140°W from upward looking Acoustic Doppler Profiler observations (black) compared to the reanalysis (red) and reanalysis experiment #3 using QuikSCAT winds (blue). Data is averaged at 5 day intervals. The correlations of the 5day time series between the observations and the reanalysis are, from top to bottom, 0.55, 0.69, 0.79, and 0.58.