

Simple Ocean Data Assimilation (SODA) reanalysis effort
(<http://www.atmos.umd.edu/~ocean>)

James Carton (Dept. Meteorology, UMD, College Park, MD, carton@atmos.umd.edu),
Ben Giese (Dept. Oceanography, TAMU, College Station, TX, b-giese@tamu.edu)

1. Background: The Simple Ocean Data Assimilation effort began at University of Maryland in the 1990s in order to develop an ocean reanalysis. In addition to providing an ocean reanalysis counterpart to the atmospheric reanalyses for application to climate studies our goal has been to examine the requirements needed for such an effort. Our period of interest, the past six decades, follows that covered by the atmospheric reanalyses. Our emphasis has been on adapting readily available models and assimilation approaches, the latter generally developed in the meteorological context, with minimum complexity and no attempts at generality. Our connection to the Washington DC area has helped this. Our goal is to keep the computational cost of the reanalysis down so that many experiments can be carried out in order to speed the rate of development. The initial efforts are described in two papers: *Carton et al. (2000a,b)*. This activity has continued. The latest release of this reanalysis is BETA25. Our support has come from the National Science Foundation and NASA, to whom we are extremely grateful.

2. Current reanalysis effort

Model and assimilation: A key aspect of the current effort is the adaptation of a much-improved eddy-permitting global model based initially on Parallel Ocean Program POP1.3 numerics (**Fig. 1**), and hopefully soon POP2.0. The use of POP numerics allows us to exploit developments and work collaboratively with the CCSM effort. The use of an eddy-permitting resolution model helps reduce model bias.

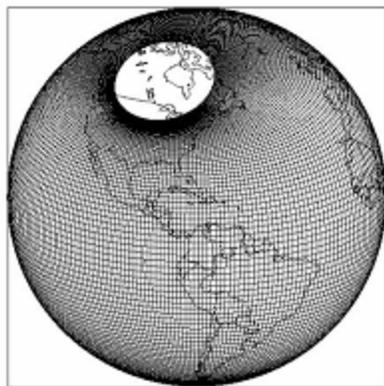


Fig. 1 Displaced pole grid. The horizontal resolution on the equator is 28km x 44km reducing to an approximately uniform 25km x 25 km in the western North Atlantic.

The sequential analysis begins with a state forecast, ω_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 ω_k^o . The sequential filter combines the forecast with the observations into an analysis, ω_k^a , which then provides the estimate of the true state. The analysis is then used as the initial condition for a subsequent forecast $\omega_{k+1}^f = \Omega \omega_k^a$.

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. Then at time t_{k+1} the bias and state analyses are computed in two stages:

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

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

where the bias-corrected forecast $\tilde{\omega}^f = \omega^f - \beta^a$ and 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. 2004* for the specification of \mathbf{K} and \mathbf{L}). The forecast error covariances are a function of depth, location, and the local flow field.

Computational requirements: These reanalysis activities are being carried out at NCAR and at NAVOCEANO (through an HPC proposal) on IBM multiprocessor computers. On these multiprocessor machines the code is extremely efficient mainly because of the efficiency of the basic model.

Ocean Data: Two-thirds of the basic subsurface temperature and salinity data set has been obtained from the World Ocean Database 2001. The WOD2001 has been extended by including the National Oceanographic Data Center/NOAA monthly updates, as well as operational profile observations from the Global Temperature-Salinity Profile Program archive (including observations from the TAO/Triton mooring thermistor array and ARGO drifters). This extended data is particularly important in improving the data coverage during the past decade, during which entries into the WOD2001 decline. Additional mixed layer temperature observations are obtained from the COADS surface marine observation set.

Substantial quality control is already included in the WOD2001 (see *Conkrite, 2002*). Additional data checking for the SODA reanalysis 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. Our additional quality control (including buddy-checking, examination of forecast-minus-observation differences, and vertical stability) eliminates an additional 5% of the profiles. This data set is available for comparison to our reanalyses. Altimetry has been obtained from the NASA Pathfinder program (for earlier reanalyses) and from

AVISO, combining multiple altimeters, including GEOSAT, for SODA1.25 (coming soon).

Winds and fluxes: We have focused on two surface wind products that span most of the period of interest, are global, and have daily resolution, the NCEP/NCAR reanalysis (1948-2004) and the ECMWF ERA40 reanalyses (1958-2001). The latter requires correction of the mean stress, which is particularly a problem in the tropics. The former behaves quite differently before and after the mid-1970s, when satellite observations were introduced. Two reanalyses have been produced to date using these wind products (see **Table 1**). Both atmospheric reanalyses suffer errors in surface heat and freshwater fluxes. Instead, we have chosen to use data-based analyses, either Xie-Arkin, and more recently, GPCP rainfall. Model net heat and freshwater fluxes are based on bulk parameterizations.

Table 1 SODA eddy-permitting reanalyses

	period	Winds	Rain	Heat
SODA1.0	1948-2003	NCEP/NCAR	GPCP	bulk
SODA1.2	1958-2001	ERA40	GPCP	bulk

User Support/Data distribution: The primary data set is monthly average fields of all state variables, including winds, on the complete 40-levels with the horizontal grid mapped onto a more convenient uniform $1/2^\circ \times 1/2^\circ$ grid in netcdf format. For SODA1.2 this data set is approximately 70Gb. Some additional data sets, e.g. dynamic height, depth of isopycnal surfaces, as well as a larger 5-day average data set (420Gb) are available upon request. We rely on web-based distribution (either through University of Maryland or our friends at GFDL, IRI, IPRC). In addition we attempt to respond to individual requests. A disadvantage of this approach is that we have little control over the different uses of our data sets. We do not have a good mechanism for warning users about data problems.

Error estimation: we conduct comparison to independent observations as well as participate in intercomparison with other reanalyses. We will provide some of the observation comparisons on our website.

References

- Carton, J.A., G. Chepurin, X.H. Cao, and B.S. Giese, 2000a: A Simple Ocean Data Assimilation analysis of the global upper ocean 1950-95. Part I: Methodology, *J. Phys. Oceanogr.*, **30**, 294-309.
- Carton, J.A., G. 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.
- Chepurin, G.A., J.A. Carton, and D. Dee, 2004: Forecast model bias correction in ocean data assimilation, *Mon. Wea. Rev.*, accepted.
- Conkright, M. E., J. I. Antonov, O. Baranova, T. P. Boyer, H. E. Garcia, R. Gelfeld, D. Johnson, R. A. Locarnini, P. P. Murphy, T. D. O'Brien, I. Smolyar, and C. Stephens, 2002: NOAA Atlas NESDIS 42, WORLD OCEAN DATABASE 2001 Volume 1: Introduction, *U.S. Gov. Printing Office, Wash., D.C.*, 160pp.