# State Estimation In Modern Oceanographic Research

D. Stammer (SIO), C. Wunsch (MIT), I. Fukumori (JPL), J. Marshall (MIT)

On behalf of The ECCO Consortium:[1]

Rigorous global ocean state estimation methods can now be used to produce dynamically consistent time-varying model/data syntheses, producing results that begin to form the basis for studies of a variety of scientifically important problems. A schematic of a complete ocean observing and synthesis system is shown in Fig. 1. It includes global observations and state-of-the-art ocean general circulation models (OGCM) run on modern computer platforms. A global observing system is described in detail in *Smith and Koblinsky* (2001) and the present status of ocean modeling and anticipated improvements are addressed by *Griffies et al.* (2000). Here we focus on the third component of state estimation: the synthesis of the observations and a model into a unified, dynamically consistent, estimate. We summarize the current status of ocean state estimation efforts, describe the first applications, and discuss future directions in support of the wider oceanographic community. Both Climate Variability and Predictability (CLIVAR) and Global Ocean Data Assimilation Experiment (GODAE) programs are being designed and built to a great extent around the assumed availability of routine state estimates. A recent summary of ongoing international assimilation efforts in support of CLIVAR (http://www.clivar.org) and GODAE (http://www.bom.gov.au/GODAE) can be found in *Stammer et al.* (2001).

Fig. 1

The ocean is a major element in the climate system, a primary mechanism of meridional heat transport, a reservoir for carbon, and a focus of biological productivity, among other roles. Ongoing and planned oceanic and climate research activities are intended to measure, understand, and eventually predict, variations in the ocean and its interaction

[1]See http://www.ecco-group.org. Corresponding Author: Detlef Stammer, Scripps Institution of Oceanography, La Jolla, CA 92093, U.S.A; dstammer@ucsd.edu

with the atmosphere, land, and cryosphere. Substantial problems exist for any system of observing and simulating the rapidly changing flow field and associated properties such as temperature or carbon and their consequences. At the present time, partial ocean observing systems are now in place as a legacy of programs such as the World Ocean Circulation Experiment (WOCE) and the Tropical Ocean Global Atmosphere Program (TOGA). Satellite altimetry and the ARGO floats have been identified as the initial backbone of a sustainable global system.

In this context, ocean state estimation ("data assimilation") has as its goal obtaining the best possible description of the changing ocean by combining, in some suitably optimum way, all the diverse ocean observations, with theoretical knowledge of the ocean circulation as embodied in numerical ocean circulation models. If carried out properly, the result is a dynamically self-consistent estimate of the time-evolving ocean circulation, one which has greater information and forecast skill than does either model or data alone. To the extent that rigorous state estimation can be accomplished, it has many diverse and overlapping applications including:

- The synthesis of all observations of the present oceanic state for climate analysis and forecast purposes.
- A foundation for hypothesis testing about the ocean, and for model improvements.
- A platform for assessment of the global observing system and of the utility of new ocean data sets.
- Improved predictability of coastal, shelf, and regional models by providing suitable open-ocean boundary conditions.
- Improved open-ocean nowcasts and forecasts (with application to search and rescue, iceberg paths, oil spills, shipping routes, fisheries, etc.).
- An understanding of the role of the ocean in the carbon cycle.

- A consistent physical state estimate for use in biological models.
- A model, properly initialized, for coupling to equivalent atmospheric and cryospheric ones.

*A Consortium for State Estimation and Model Improvements*

The effort required to carry out rigorous global ocean state estimation is formidable. To marshal adequate resources, a consortium has been formed called "Estimating the Circulation and Climate of the Ocean" (ECCO) with funding provided by the National Aeronautics and Space Administration (NASA), the National Science Foundation (NSF), and the Office of Naval Research (ONR) under the National Oceanographic Partnership Program (NOPP). It builds on work that begun initially at the Massachusetts Institute of Technology (MIT), but now also involves scientists at the Scripps Institution of Oceanography (SIO) and the Jet Propulsion Laboratory (JPL). ECCO intends to transform existing ocean state estimation efforts from their present experimental status into quasi-operational applications that can support scientific, societal and navy needs alike. ECCO's ultimate goal is to employ rigorous methods of ocean/data syntheses in a sustained form to describe the global ocean circulation at time scales of days to decades, using data of any type and the best available models.

ECCO activities are based on the MIT general circulation model ( *Marshall, et al.,* 1997a,b), including complete mixed layer physics and an eddy parameterization scheme. A release, with full documentation of both forward and adjoint ECCO codes, is available at http://mitgcm.org/sealion. The forward model code continues to evolve as part of ECCO, as does the adjoint counterpart; the community is enthusiastically encouraged to exploid both.

Results from all ECCO assimilation efforts can be accessed via the consortium's data server (Live Access Server at http://www.ecco-group.org/las). The server also provides the input observations and initial forcing fields as they were used during the various experiments. See *Lee et al.*, (2002), *Fukumori et al.* (2002) and *Stammer et al.* (2002a,b,c) or http://www.ecco.ucsd.edu for further details.

# Methodology

All data assimilation methods deal, in principle, with a vector $\mathbf{x}(t)$ representing a time-dependent model simulation of the state of the ocean. These state vectors include, typically on a regular grid, the values of temperature, salinity, pressure and velocity required by a general circulation model (GCM) to take one time-step into the future, if accompanied by appropriate boundary conditions. From the state vector, one can compute any derived quantity (e.g., enstrophy, potential vorticity or enthalpy flux, etc.) of interest. All oceanic measurements can be associated with its model equivalent through a functional relationship that can be written as,

$$\mathbf{E}(t)\,\mathbf{x}(t) + \mathbf{n}(t) = \mathbf{y}(t)\,, \tag{1}$$

where $\mathbf{E}$ is a matrix, $\mathbf{n}(t)$is the measurement error including data noise, and $\mathbf{y}(t)$ is the vector of observations. An estimated time-dependent solution, $\tilde{\mathbf{x}}(t)$, is then sought, that minimizes—usually, but not necessarily, in a least-squares sense—the model-data misfits, which then can be written as,

$$J = \sum_t (\mathbf{y}(t) - \mathbf{E}(t)\,\mathbf{x}(t))^T \mathbf{R}(t)^{-1} (\mathbf{y}(t) - \mathbf{E}(t)\,\mathbf{x}(t)), \tag{2}$$

subject to the model dynamics. Here $\mathbf{R}(t)$ is the observational error covariance.

Simplified methods (e.g., nudging, robust diagnostic, and objective mapping), intended to find approximations to the minimum of $J$ by relaxing the model constraints, are easy to set up and are computationally inexpensive. However, the temporal evolution of such ad hoc solutions are not necessarily consistent dynamically and usually imply internal sinks and sources of momentum, heat and freshwater. Rigorous methods, such as used by ECCO and others, are computationally much more demanding. We note, however, that they are needed to obtain dynamically self-consistent ocean estimates useful for understanding the physics of the system by exploiting all the information contained in the data.

## Ongoing ECCO Activities

Initially, two assimilation methods are being used by the ECCO consortium. One approach relies upon enforcing the model through what has come to be known as the "adjoint method" or the method of Lagrange multipliers. Details of how this has been done, choices for the elements of $\mathbf{R}(t)$, and the model error, as well as a full description of the model as run on a global 2° grid can be found in *Stammer et al.* (2002a). We refer to this calculation as a "preliminary WOCE synthesis." To bring the model into consistency with the observations, the initial-conditions for potential temperature and salinity fields are modified, as well as the surface forcing fields. Changes in those fields (often referred to as "control" terms) are determined as a best-fit, in a least-squares sense, of the model state to the observations and their uncertainties over the full data period. Observations, to which the model is fit include the daily TOPEX/POSEIDON (T/P) and ERS altimetric sea surface height (SSH) anomalies, the mean SSH, monthly mean SST fields, *Levitus et al.* (1994) monthly mean temperature and salinity climatologies over the entire water column as well as surface flux fields of momentum, heat and freshwater. The "first guess" estimate of the surface forcing is provided by the daily time-varying National Center for Environmental Prediction (NCEP) reanalysis fluxes of momentum, and daily estimates of heat, and freshwater (*Kalnay et al.*, 1996). Scatterometer results are used to estimate wind stress errors; see the references for a more general discussion of error estimates.

In parallel with the global 2° WOCE synthesis, a second global estimate employing a combination of the adjoint method and a reduced-state Kalman filter plus a *Rauch et al.* (1965, RTS) smoother are being used on a telescoping grid with 1° resolution in latitude and longitude in mid- and high latitudes. Meridional resolution decreases to 1/3° near the equator. This latter approach is intended to test the merits of the alternative minimization algorithm while focusing the science on the seasonal-to-interannual variability of the Pacific Ocean. We refer to this coimputation as a "preliminary near-real time analysis". Thus far, TOPEX/POSEIDON sea level anomalies, climatological mean temperature and salinity, and surface flux data from 1993 to present are being used. Results are described in detail by *Lee et al.* (2002), *Fukumori* (2002), *Fukumori et al.* (2002).

In addition to the global state estimates, ECCO is also using regional models that are nested into the global results and that make estimates with significantly higher resolution than is now possible globally. For example, we model the North Atlantic and the tropical Pacific at 1/6° horizontal resolution or higher. Both open and closed boundary conditions are used.

Ongoing efforts of the ECCO Consortium are now aiming to produce two major, sustained analysis products: (1) A full global synthesis of all available data at the highest practical resolution. (2) A near-real time product, whose delay is to be no longer than about two weeks, using most of the routinely available continuing data, and an estimation method compromising the need for optimality with that for computational efficiency. Both products will be routinely distributed through the project data server.

# First Science Applications

Both global state estimation efforts have produced results that are sufficiently acceptable for first scientific applications and analyses. Here we can only provide a few examples of what is ongoing or can be expected to emerge over the next few months and years in many fields of oceanographic research.

Within ECCO, a particular science focus is on the determination of transports and the budgets of mass, heat, freshwater, and energy in various regions of the global domain. The state estimates are also being examined to assess the accuracies of various estimates of air-sea fluxes of momentum, heat, and freshwater; to quantify the relative impact of different observing systems; to study ocean dynamics; and to improve forecasting skills. Fig. 1 shows (right panel) the mean flow field at 27m and 1975 m depth from the 2° WOCE synthesis calculation, together with the mean sea surface height and the temperature field at 1975 m. All major circulation structures are simulated, but are overly smooth due to the present low model resolution.

The mean net surface heat flux field resulting from the the optimization is displayed in the upper panel of Fig.2. The adjustment relative to the prior NCEP estimate is also Fig.2 shown that is required to bring the model state into consistency with the ocean data. Modifications of the net NCEP heat fluxes are of the order of $\pm 20$ W $m^{-2}$ over large parts of the interior oceans. Maximum changes occur along the boundary currents in the Northern Hemisphere where changes of up to $\pm 100$ W $m^{-2}$ can be found. Most of the eastern boundary currents now show a significant heat uptake. The same is true in the Arabian Sea, where the off-shore Ekman transport brings up cold water, and which is then heated by the atmosphere. Strong warming occurs over Flemish Cap, in the North Pacific, and along most of the Antarctic Circumpolar Current. Note that the optimization also removes some of the small-scales visible in the eastern Pacific in the initial NCEP estimates of buoyancy flux fields; these originate from Gibbs effects in the spectral representation of mountain ranges such as the Andes. Limited space here prohibits a detailed discussion of the skill in the estimated net flux changes, but *Stammer et al.* (2002c) compare the large-scale ECCO-estimated momentum, heat and fresh water fluxes with estimates obtained from satellite data and regional bulk formula estimates. They conclude that combined ocean observations and dynamics provide a route to improving air-sea flux estimates, complementary to those from atmospheric models, or from direct observational campaigns.

An example of the use of the results for inferring physical processes in the ocean is given in Fig. 3, taken from the *Fukumori et al.*, 2002 run. The figure shows (top-panel) a temperature timeseries at 100m depth from the constrained model, and compares it with observations (not used as constraints) from one of the TOGA-TAO buoys at that location, as well as with the unconstrained model. In contrast to the unconstrained run, the constrained results simulate the temperature observations quite closely. Taking the figure as an indication of the constrained model skill in simulating the observed flow, temperature and salt fields, one can than start using the model output to compute unobserved quantities, such as ocean transports or heat storage. The figure also shows (lower two panels) a diagnosis of the model mixed-layer temperature evolution in the eastern equatorial Pacific (Nĩno-3: 5S-5N, 150W-90W) during 1997-2000. The tendency of advection and horizontal diffusion to warm the mixed layer during 1997-2000, is counterbalanced by a loss of heat to the atmosphere. Vertical advection—one of the dominant agents in regulating mixed-layer temperature here—changes its sign in 1999 as El Niño switches to La Niña (see *Lee et al.*, (2002) for details). The examples illustrate what can be accomplished by rigorous state estimation: inference is made about unobserved quantities from sparse observations, with the estimated state evolution being dynamically consistent in the sense that changes can be attributed to a sum of explicit physical processes. Fig. 3

On a global scale, meridional heat fluxes integrated across basins equal the air-sea heat divergence over the respective region (see *Stammer et al.*, 2002b). Meridional heat transports for the South Atlantic and Indo-Pacific Oceans, show good agreement (within error bars) with the wholly independent estimates from the Ganachaud/Wunsch global hydrographic inversion. Significant discrepancies do exist over the North Atlantic however, where the present result is only 50% of the inverse estimate at mid-latitudes. This result is expected in a relatively viscous model with 2° lateral resolution, in which the boundary currents (Gulf Stream and deep western boundary current) are sluggish and diffuse.

Output from the optimizations are also being used in studies outside physical oceanography, including simulations of biogeochemical tracer transport. Fig. 4 shows the 6-year Fig. 4

mean (1993-98) air-sea CO2 flux obtained by driving a biogeochemical model with the JPL constrained circulation fields. Outgassing to the atmosphere occurs in the tropics where carbon-rich abyssal waters upwell to the surface: carbon is taken up by the ocean in high latitudes where cold waters are mixed to depth. Fig. 4 shows the mean air-sea flux, but the model exhibits much interannual variability with ENSO being a major player in modulating the flux in the tropics. The model is being used to study interannual variability in air-sea CO2 and O2 fluxes and compared with estimates based on atmospheric inversions (*McKinley et al.*, 2002).

Another example of applications outside physical oceanography is the determination of the ocean's angular momentum (OAM) fluctuations and their contribution to observed polar motion and Earth rotation. *Ponte et al.* (1998), using forward model runs without any assimilation, showed the effects of OAM variability on the Earth's rotation. Subsequently, *Ponte et al.* (2001) demonstrated that the optimized solution produces a significantly increased fraction of the observed rotation variance. This type of integral comparison provides a strong independent consistency check on the estimated ocean state estimate and underlines the importance of such estimates for study of Earth rotation and polar motion. Ultimately, measurements of polar motion and rotation will be used to constrain the ocean flow field, as will be measurements of gravity fluctuations and changes in the ocean mass field expected to originate from the upcoming Gravity Recovery and Climate Experiment (GRACE) mission (http://www.csr.utexas.edu/grace/).

A by-product of the adjoint method is the estimate of the adjoint solution. This solution exists in a space dual to that of the forward model (e.g., *Wunsch,* 1996), and represents a nearly complete description of estimate sensitivity to data, and of the flow of information in the model. Use of this solution, to examine the controlling factors on oceanic meridional heat flux, was described by *Marotzke et al.* (1999).

# Outlook

Physical oceanography is entering a new era in which there will be a far greater reliance on global observations and ocean state estimation as a synthesis tool to provide the community with estimates of the time-evolving ocean and climate state. The preliminary results reported here show that the existing data-base and the available modeling and computing capability have advanced to the point where true three dimensional, skillful estimates of the global time-evolving general circulation are practical. To a large extent, this statement is a vindication of the vision which drove the World Ocean Circulation Experiment—that such estimates could become possible by about the year 2000, and that they would be necessary for the advancement of the science.

We strongly emphasize however, that we are at the very beginning of what will be a long process of improvements and enlargement of the scope of oceanic state estimation. The analogous effort in numerical weather prediction has already spanned several decades. With expected advances in computer power and numerical algorithms, models with greatly increased spatial resolution, and more sophisticated and complete physical parameterizations will come. Regional estimates at extremely high resolution with open boundaries are already underway. Global estimates at $1/4°$ are envisioned as being feasible in two-to-three years. A much fuller exploration of the control vector space will be undertaken, including adjustments to internal parameters such as diffusion and viscosity coefficients, and external parameters such as the bottom topography and sidewall boundary condition relationships. The information contained in adjoint solutions remains to be more fully exploited.

Looking further into the future, one can envision ongoing global estimates of coupled ocean/atmosphere/cryosphere models with sufficient skill that true climate forecasting (that is, with skill estimates) will be possible. At the same time, we envision a vigorous interaction between estimation activities and the biogeochemical and physical observation efforts now being designed. A critical element for enhanced understanding of biological processes in the ocean is the availability of a realistic flow field such as that which becomes available from ocean state estimation. The greatest challenge of all to the oceanographic community may well be to find a way to sustain the long term observing systems, model development and synthesis activities which are all essential to reaching our goal.

*Acknowledgments* The authors acknowledge support of the ECCO project through the National Oceanographic Partnership Program, with funding from NASA, NSF and ONR. The ECCO consortium also acknowledges computational support through NSF (NPACI, NCAR), ONR and NASA.

# References


[1] Fukumori I., A partitioned Kalman filter and smoother, Monthly Weather Review, (in press), 2002.

[2] Fukumori, I., B. Tang, D. Menemenlis, Z. Xing, T. Lee, B. Cheng, and L.-L. Fu, Evolution of the global oceanic state from 1993 to 2000 estimated by altimetric data assimilation, (in preparation), 2002.

[3] Griffies, S.M., C. Böning, F.O. Bryan, E.P. Chassignet, R. Gerdes, H. Hasumi, A. Hirst, A. Trguir, and D. Webb, 2001: Developments in ocean climate modelling, *Ocean Modelling*, 2, 123–192.

[4] Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, A. Leetmaa, R. Reynolds, and R. Jenne, The NCEP/NCAR reanalysis project. *Bull. Amer. Meteor. Soc.*, 77, 437-471, 1996.

[5] Lee, T., B. Cheng, I. Fukumori, D. Menemenlis, L.-L. Fu, Z. Xing, and B. Tang, Assimilation of TOPEX/Poseidon data into a near-global OGCM: application to analysis of upper-ocean heat budget, (in preparation), 2002.

[6] Marotzke, J., R. Giering, Q. K. Zhang, D. Stammer, C. N. Hill, and T. Lee, Construction of the adjoint MIT ocean general circulation model and application to Atlantic heat transport sensitivity. *J. Geophys. Research*, 104, 29,529 - 29,548, 1999.

[7] Marshall, J., A. Adcroft, C. Hill, L. Perelman, and C. Heisey, A finite-volume, incompressible navier-stokes model for studies of the ocean on parallel computers. *J. Geophys. Res.*, 5753–5766, 1997a.

[8] Marshall, J., C. Hill, L. Perelman, and A. Adcroft, Hydrostatic, quasi-hydrostatic, and nonhydrostatic ocean modeling, J. Geophysical Res., 102(C3), 5733-5752, 1997b.

[9] McKinley, G, Follows, M. and J. Marshall, Internnual variability in air sea fluxes of CO2 and O2. in preparation, 2002.

[10] Ponte, R., D. Stammer and C. Wunsch, 2001: Improved ocean angular momentum estimates using an ocean model constrained by large-scale data, *Geophys. Res. Letters*, 28, 1775–1778.

[11] Rauch, H.E., F. Tung, and C.T. Streibel, 1965: Maximum likelihood estimates of linear dynamic systems, *AIAA J.*, 3, 1445–1450.

[12] Smith, N. and C. Koblinsky, Observing the Ocean in the 21st Century, Bureau of Meteorology, Melbourne, Australia, 2001.

[13] Stammer, D., C. Wunsch, R. Giering, C. Eckert, P. Heimbach, J. Marotzke, A. Adcroft, C.N. Hill, and J. Marshall, The global ocean circulation during 1992 –1997, estimated from ocean observations and a general circulation model, *J. Geophys. Res.*, in press., 2002a.

[14] Stammer, D., C. Wunsch, R. Giering, C. Eckert, P. Heimbach, J. Marotzke, A. Adcroft, C.N. Hill, and J. Marshall, Ocean transports during 1992 –1997, estimated from ocean observations and a general circulation model, *J. Geophys. Res.*, submitted for publication, 2002b.

[15] Stammer, D., K. Ueyoshi, W.B. Large, S. Josey and C. Wunsch, Global sea surface estimates obtained through ocean data assimilation, *Bulletin of the American Meteorological. Society*, submitted for publication, 2002c.

[16] Stammer, D., R. Bleck, C. Böning, P. DeMey, H. Hurlburt, I. Fukumori, C. LeProvost, R. Tokmakian, Wenzel, 2001: Global ocean modeling and state estimation

in support of climate research. In: *Observing the Ocean in the 21st Century*, C.J. Koblinsky and N.R. Smith (eds), Bureau of Meteorology, Melbourne, Australia, 511–528.

[17] Wunsch, C., *The Ocean Circulation Inverse Problem*, 442 pp., Cambridge Univ. Press, New York, 1996.

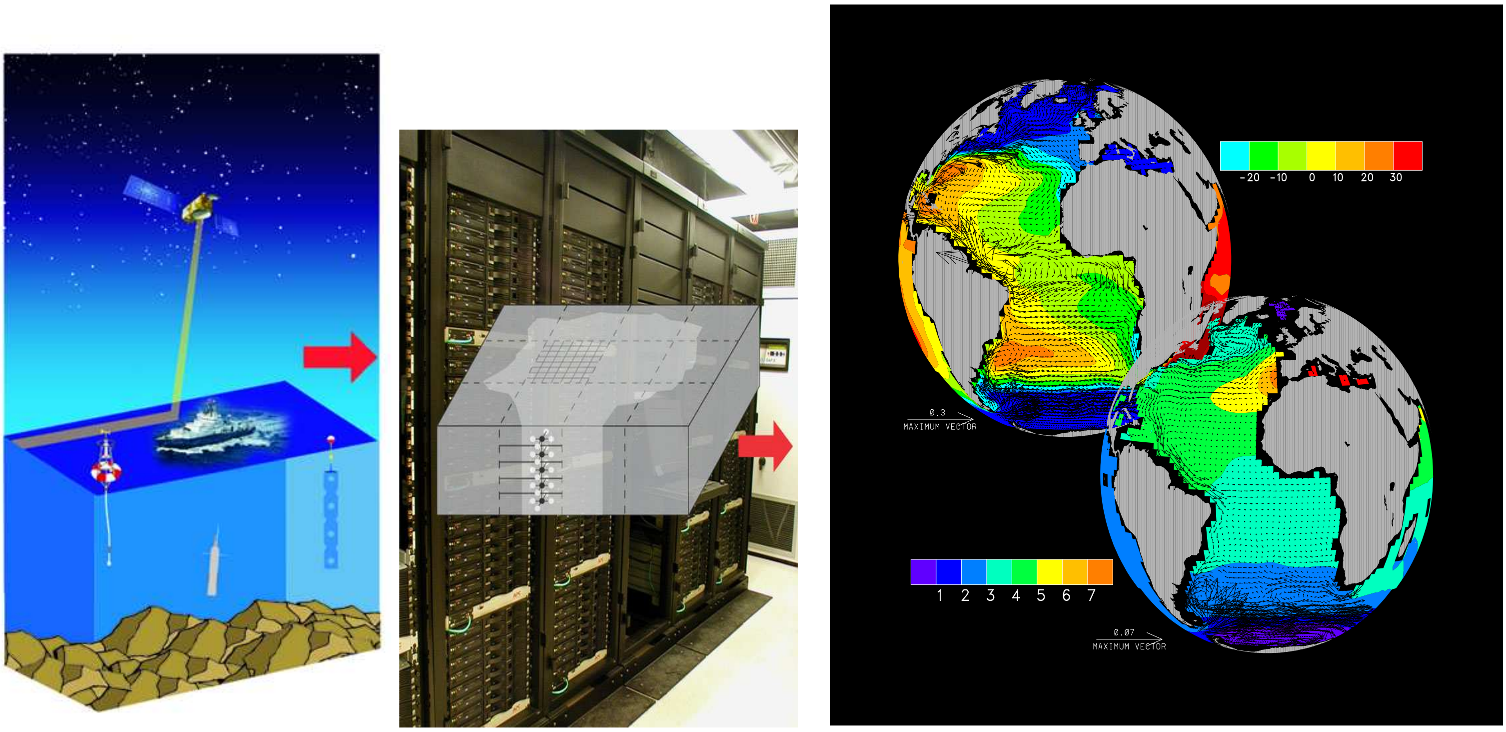


Figure 1: Schematic of ocean observing and assimilation system. The observing system on the left is providing near-realtime ocean observations. Combined with dynamics embedded in state-of-the-art circulation models that are run on super-computers (middle panel) ocean syntheses are obtained that form the basis for studies of the ocean circulation. Shown here in the right panel is an example of the ongoing WOCE synthesis for the near surface and the deep circulation. See Stammer et al. (2002a,b,c) for details.

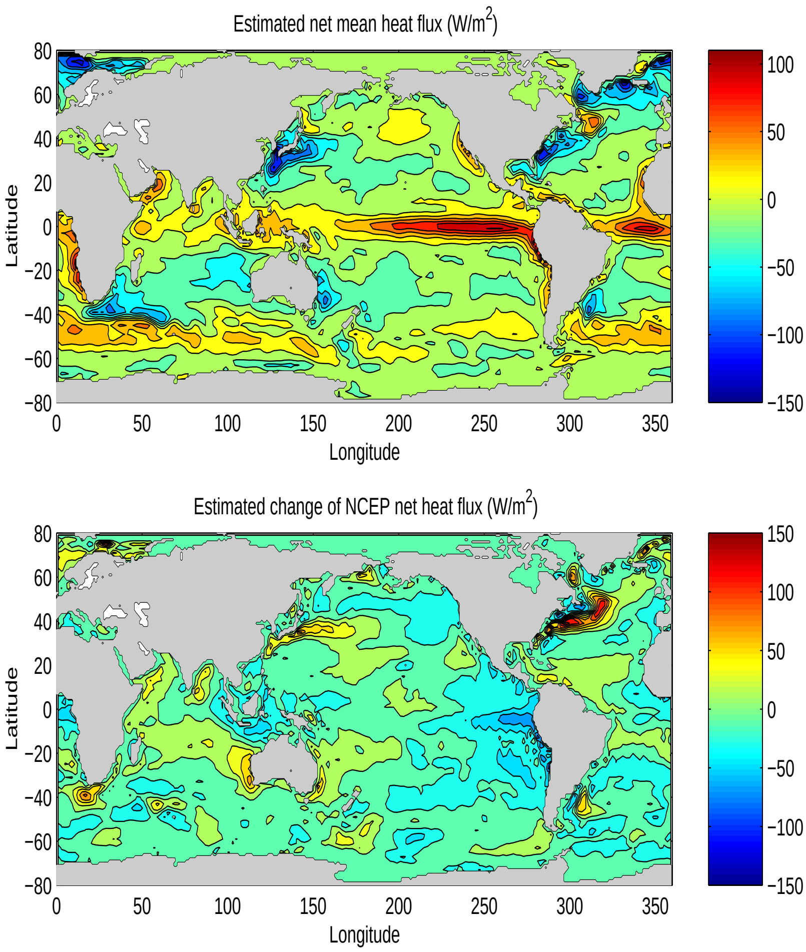


Figure 2: (upper panel) Mean estimated net surface heat flux required to bring the model into consistency with data (in Watts/m$^2$). (lower panel) Mean estimated adjustment of the National Center for Environmental Prediction (NCEP) net surface heat flux required to bring the model into consistency with data (countour increment is 20 Watts/m$^2$). See *Stammer et al.* (2002a,c) for details.

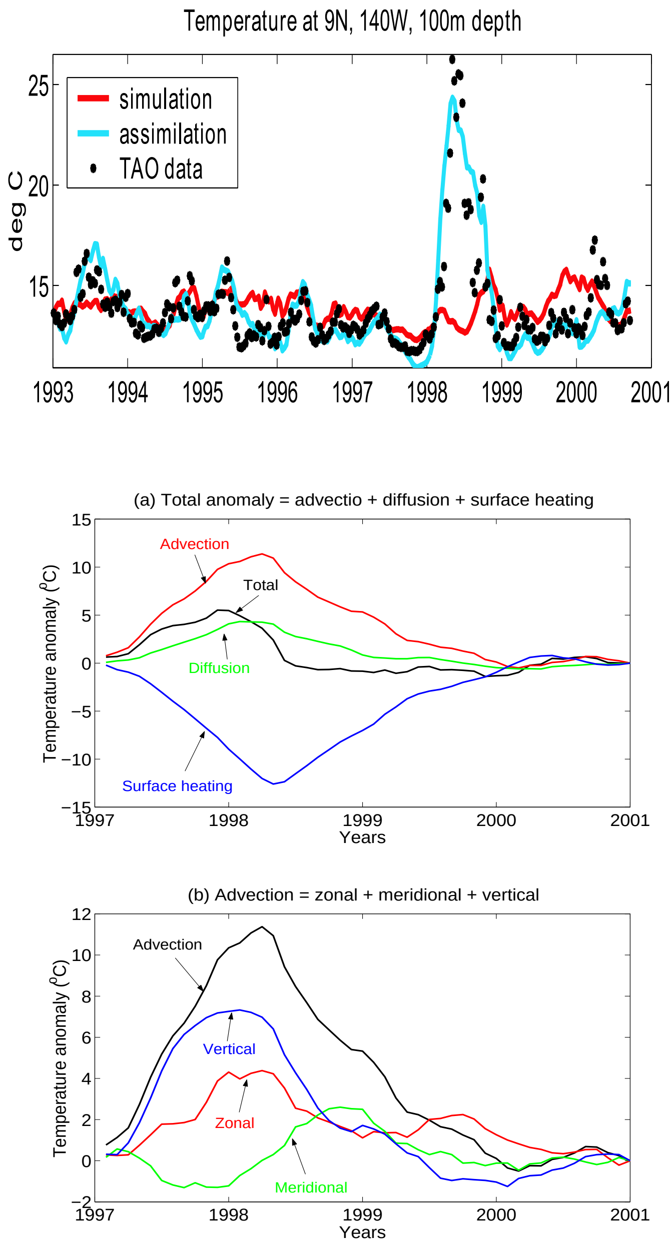


Figure 3: top: (top): Temperature time series at 9° N, 140° W at 100 m depth. Shown are results from the constrained near-realtime run discussed by *Fukumori et al.* (2002) in blue and from an unconstrained run in red. Observations from a TOGA TOA buoy at the same location are shown as bold dots. (Middle and bottom panels): Contributions of various terms to the tendency of the mixed layer temperature heat balances in the eastern tropical Pacific as described by *Lee et al.* (2002). Individual terms are specified in the panels.

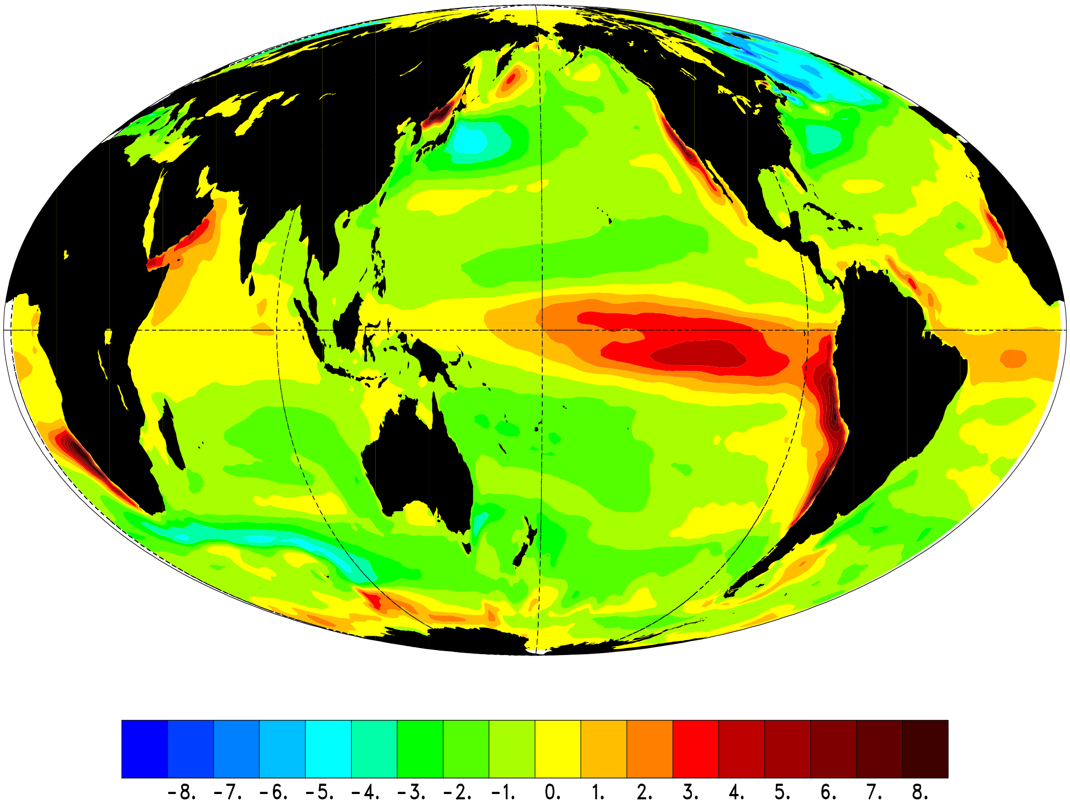


Figure 4: The 6-year mean (1993-98) air-sea CO2 flux obtained by driving a biogeochemical model with a constrained circulation field. Fluxes are in mol/m**2/yr and positive indicates a flux from the ocean to the atmosphere.