APPENDIX A

# Data Assimilation

**Coordinating Lead Authors:** James Carton, University of Maryland; Eugenia Kalnay, University of Maryland

Data assimilation is an exercise in the calculation of conditional probabilities in which short-term model forecasts are combined with observations to best estimate the state of, for example, the atmosphere. Since there are limitations in model resolution and errors associated with parameterization of unresolved physical processes, and the behavior of the atmosphere is chaotic, forecast accuracy is described by a probability distribution, as is observation accuracy. These probability distributions are combined to form conditional probabilities, which are simplified by assuming these distributions are Gaussian (normally distributed). The conditional probabilities are used to create a more accurate *analysis* than can be obtained solely from either the forecasts or the observations. The same approach can be applied to the ocean, land surface, or cryosphere.

Atmospheric data assimilation proceeds through a succession of (typically) six hour *analysis cycles*. At the beginning of each cycle, a six-hour model forecast is carried out starting from initial conditions of atmospheric pressure, temperature, humidity, and winds provided by the previous analysis cycle, with observed boundary conditions such as sea surface temperature and snow cover. At the end of each cycle all available current observations are quality controlled, and the differences between the observations and the model forecast of the same variables, referred to as observational increments or innovations, are computed. The observations may include the same variables observed with different systems (*e.g.*, winds measured from airplanes or by following the movement of clouds). They may also include observations of variables that do not directly enter the forecast such as satellite radiances, which contain information about both temperature and moisture.

If the evolving probability distributions of the model forecasts and observations are known, then it is possible to construct an analysis that is optimal because the expected error variance, which is the difference between the analysis of a variable and its true value, is minimized. In practice, the probability distributions are unknown. In addition, it is not possible to solve the computational problem of minimizing the error variance for realistic complex systems. In order to address these problems, several simplifying assumptions are needed. The observational increments are generally assumed to be Gaussian. With this assumption a cost function can be constructed whose minimization, which provides us with the optimal analysis, leads to the Kalman Filter equations. A bigger assumption that the probability distribution of the forecast errors does not depend on time, gives rise to the widely used and more simplistic three-dimensional variational type of data assimilation (3DVAR). Four-dimensional variational data assimilation (4DVAR) is a generalization of the cost function approach that allows the forecast initial conditions (or other control variables such as diffusive parameters) to be modified based on observations within a time window.

Despite the use of simplifying assumptions, the Kalman Filter and 4DVAR approaches still lead to challenging computational problems. Efforts to reduce the magnitude of the computational problems and exploit physical understanding of the physical system have led to the development of Monte Carlo approaches known as Ensemble Kalman Filter (EnKF). EnKF methods, like 4DVAR, can be posed in such a way that the analysis at a given time can be influenced by past, present, and future observations. This property of time symmetry is especially desirable in reanalyses since it allows the analysis at past times to benefit to some extent from future enhancements of the observing system.

**Table A.1 Characteristics of some existing global ocean model-based reanalyses of ocean climate (extracted from: <http://www.clivar.org/data/synthesis/directory.php>)**

| Organization/System | Model | Analysis Method | Time Period | Weblinks |
|---|---|---|---|---|
| NCEP, Météo France, CERFACS | OPA8.2,2°x2°x31 Lev (~0.5°x2° tropics) ERA40 forcing | Multivariate 3D-Var (OPAVAR) for T & S profiles | 1962 to 2001 | <.fr/globc/overview.html> |
| ECMWF | HOPE, 1°x1°x29 Lev (1/3°x1° tropics) | OI | 1959 to 2006 | <ecmwf.int/products/forecasts/d/charts/ocean/reanalysis/> |
| ECCO-GODAE | MITgcm 1°x1° | 4DVAR | 1992 to 2004 | <www.ecco-group.org> |
| ECCO-JPL | MITgcm and MOM4 1°x1°x50 lev | Kalman filter and RTS smoother | 1993 to present | <ecco.jpl.nasa.gov/external/> |
| ECCO-SIO | 1°x1° | 4DVAR | 1992 to 2002 | <ecco.ucsd.edu> |
| ECCO2 | MITgcm, 18kmx50 Lev | Green's functions | 1992 to present | |
| ENACT consortium | | | 1962 to 2006 | <www.ecmwf.int/research/EU_projects/ENACT/> |
| FNMOC/GODAE | | | | <www.usgodae.org> |
| GECCO | | | 1950 to 2000 | <www.ecco-group.org> |
| GFDL | | | 1960 to 2006 | <www.gfdl.noaa.gov/> |
| UK Met Office GloSea | GloSea OGCM 1.25°x1.25°x40 Lev (0.3°x1.25° tropics) daily ERA40 fluxes with corrected precipitation | OI | 1962 to 1998 | <www.metoffice.gov.uk/research/seasonal/glosea.html> |
| NASA Goddard GMAO | Poseidon, 1/3°x5/8° | MVOI, Ensemble KF | 1993 to present | <gmao.gsfc.nasa.gov> |
| INGV | OPA8.2 2°x2°x31 Lev (0.5°x2° tropics) ERA40 and operational ECMWF fluxes | Reduced Order MVOI with bivariate T and S EOFs | 1962 to present | |
| MEXT K-7 | MOMv3 1°x1°x36 Lev NCEP2 reanalysis, ISCCP data | 4DVAR | 1990 to 2000 | <www.jamstec.go.jp/frcgc/k7-dbase2/eng/> |
| MERCATOR-3 | OPA8.2 2°x2°x31 Lev (~0.5° meridional at the tropics) | Single Evolutive Extended Kalman (SEEK) filter | 1993 to 2001 | <www.mercator-ocean.fr.html/systemes_ops/psy3/index_en.html> |
| JMA MOVE/MRI.COM | | | 1949 to 2005 | <www.mri-jma.go.jp/Dep/oc/oc.html> |
| NOAA/NCEP GODAS | MOMv3 1°x1°x40 Lev (1/3°x1° tropics) NCEP Reanalysis2 | 3DVAR | 1980 to present | <www.bom.gov.au/bmrc/ocean/JAFOOS/POAMA/> |
| BoM, CSIRO, POAMA | ACOM2 (based on MOM2), 2°x2°x27 Lev (0.5°x2° at high latitudes) ERA40 | MVOI, ensemble KF | 1980 to 2006 | <www.atmos.umd.edu/~ocean/> |
| SODA | POP1.4, POP2.01, global ave | MVO1 with evolving error | 1958 to 2005 | |