Chapter 13
Variational Data Assimilation for the Global Ocean

James A. Cummings and Ole Martin Smedstad

Abstract A fully three dimensional, multivariate, variational ocean data assimilation system has been developed that produces simultaneous analyses of temperature, salinity, geopotential and vector velocity. The analysis is run in real-time and is being evaluated as the data assimilation component of the Hybrid Coordinate Ocean Model (HYCOM) forecast system at the U.S. Naval Oceanographic Office. Global prediction of the ocean weather requires that the ocean model is run at very high resolution. Currently, global HYCOM is executed at 1/12 degree resolution ($\sim$7 km mid-latitude grid mesh), with plans to move to a 1/25 degree resolution grid in the near future ($\sim$3 km mid-latitude grid mesh). These high resolution global grids present challenges for the analysis given the huge model state vector and the ever increasing number of satellite and in situ ocean observations available for the assimilation. In this paper the development and evaluation of the new oceanographic three-dimensional variational (3DVAR) data assimilation is described. Special emphasis is placed on documenting the capabilities built into the 3DVAR to make the system efficient for use in global HYCOM.

13.1   Introduction

Eddy-resolving global ocean prediction requires high resolution since the characteristic scale of ocean eddies is on the order of a few tens of kilometers. Only recently have sufficient data and computer power become available to nowcast and forecast the ocean weather at eddy-resolving scales, including processes that control the surface mixed layer, the formation of ocean eddies, meandering ocean

J.A. Cummings (✉)
Oceanography Division, Naval Research Laboratory, Monterey, CA, USA
e-mail: cummings@nrlmry.navy.mil

O.M. Smedstad
QinetiQ North America, Stennis Space Center, MS, USA

currents and fronts, and generation and propagation of coastally trapped waves. Hurlburt et al.(2008a) gives a good discussion of the requirements for an ocean model to be eddy-resolving. High resolution global ocean forecast models present challenges for the assimilation component of the forecasting system given the huge model state vector and the ever increasing number of satellite and in situ ocean observations available for the assimilation. Accordingly, the global analysis has to be both computationally efficient and accurate to account for the oceanographic features resolved by the high resolution model. At the same time the analysis must use all of the available observations and create and maintain dynamically adjusted corrections to the model forecast.

The purpose of this chapter is to provide an overview of a new variational ocean data assimilation system that has been developed as an upgrade to an existing multivariate optimum interpolation (MVOI) system (Cummings 2005). Compared to the MVOI the 3DVAR algorithm has several advantages. First, the 3DVAR performs a global solution that does not require data selection. In the MVOI, observations are organized into overlapping analysis volumes and the solution can depend on how the volumes are defined. This is not the case in the 3DVAR, as the global solve allows all observations to influence all grid points, a requirement for an optimum analysis. Second, through the use of observation operators, 3DVAR can incorporate observed variables that are different from the model prognostic variables. Examples of this in the ocean are integral quantities, such as acoustic travel time and altimeter measures of sea surface height, and direct assimilation of satellite radiances of sea surface temperature (SST) through radiative transfer modeling. Finally, 3DVAR permits more powerful and realistic formulations of the background error covariances, which control how information is spread from the observations to the model grid points and model levels. The error covariances also ensure that observations of one model variable produce dynamically consistent corrections in the other model variables.

The 3DVAR referred to in this paper is the Navy Coupled Ocean Data Assimilation (NCODA) system, version 3. NCODA 3DVAR is in operational use at the U.S. Navy oceanographic production centers: Fleet Numerical Meteorology and Oceanography Center (FNMOC) in Monterey, CA, and the Naval Oceanographic Office (NAVOCEANO) at the Stennis Space Center, MS. NCODA is truly a unified and flexible oceanographic analysis system. It is designed to meet all Navy ocean data analysis and assimilation requirements using the same code. In two-dimensional mode, NCODA provides SST and sea ice concentration analyses for lower boundary conditions of the Navy global and regional atmospheric forecast models. In three-dimensional mode, it is executed in a sequential incremental update cycle with the Navy ocean forecast models: the Hybrid Coordinate Ocean Model (HYCOM) on the global scale, and the Navy Coastal Ocean Model (NCOM) on the regional scale. Here, NCODA provides updated initial conditions of ocean temperature, salinity, and currents for the next run of the ocean forecast model. The analysis background fields, or first guess, are generated from a short-term ocean model forecast, and the 3DVAR computes dynamically consistent corrections to the first-guess fields using all of the observations that have become available

since the last analysis was made. Further, NCODA 3DVAR is globally relocatable and has been integrated into the Coupled Ocean Atmosphere Mesoscale Prediction System (COAMPS$^{\circledR}$[1]), which is used by Navy for rapid environmental assessment. In this mode of operation, the 3DVAR performs multi-scale analyses on nested, successively higher resolution grids. Finally, NCODA provides the data assimilation component for the WAVEWATCH wave model forecasting system at FNMOC (Wittmann and Cummings 2005). In this mode of operation, NCODA computes corrections to the model's two-dimensional wave spectra from assimilation of satellite altimeter and wave buoy observations of significant wave height.

The examples used in the paper are taken from NCODA 3DVAR analyses cycling with global HYCOM. Sections 13.2 and 13.3 of the paper describe the assimilation method and techniques used to specify the error covariances. Section 13.4 lists the ocean observing systems assimilated and outlines the data selection and data pre-processing that is done for the real-time global forecast. Section 13.5 gives an overview of the entire NCODA system, including the diagnostic suite. Section 13.6 presents some verification results from global HYCOM. Section 13.7 describes future capabilities and applications of the NCODA 3DVAR system, while Sect. 13.8 gives a summary.

13.2 Method

The method used in NCODA is an oceanographic implementation of the Navy Variational Atmospheric Data Assimilation System (NAVDAS), a 3DVAR technique developed for Navy numerical weather prediction (NWP) systems (Daley and Barker 2001). The oceanographic 3DVAR analysis variables are temperature, salinity, geopotential (dynamic height), and u, v vector velocity components. All ocean variables are analyzed simultaneously in three dimensions. The horizontal correlations are multivariate in geopotential and velocity, thereby permitting adjustments to the mass fields to be correlated with adjustments to the flow fields. The velocity adjustments (or increments) are in geostrophic balance with the geopotential increments, which, in turn, are in hydrostatic agreement with the temperature and salinity increments. The multivariate aspects of the 3DVAR assimilation are discussed further in Sect. 13.3.3.

The NCODA 3DVAR problem is formulated as:

$$x_a = x_b + P_b H^T (H P_b H^T + R)^{-1} [y - H(x_b)] \qquad (13.1)$$

where $\mathbf{x_a}$ is the analysis vector, $\mathbf{x_b}$ is the background vector, $\mathbf{P_b}$ is the positive-definite background error covariance matrix, $\mathbf{H}$ is the forward operator, $\mathbf{R}$ is the observation error covariance matrix, and $\mathbf{y}$ is the observation vector. At the present

---

[1]COAMPS$^{\circledR}$ is a registered trademark of the Naval Research Laboratory

time, the forward operator in NCODA is spatial interpolation performed in three dimensions by fitting a surface to a $4 \times 4 \times 4$ grid point target and evaluating the surface at the observation location. Thus, $\mathbf{HP_bH^T}$ is approximated directly by the background error covariance between observation locations, and $\mathbf{P_bH^T}$ directly by the error covariance between observation and grid locations. For the purposes of discussion, the quantity $[\mathbf{y} - \mathbf{H(x_b)}]$ is referred to as the innovation vector, $[\mathbf{y} - \mathbf{H(x_a)}]$ is the residual vector, and $\mathbf{x_a}$-$\mathbf{x_b}$ is the increment (or correction) vector.

The observation vector contains all of the synoptic temperature, salinity and velocity observations that are within the geographic and time domains of the forecast model grid and update cycle. Observations can be assimilated at their measurement times within the update cycle time window by comparison against time dependent background fields using the first guess at appropriate time (FGAT) method. An advantage of the FGAT method is that it eliminates a component of the mean analysis error that occurs when comparing observations and forecasts not valid at the same time. As will be described in Sect. 13.6, the use of FGAT in real-time HYCOM allows for assimilation of late receipt observations.

Equation (13.1) is the observation space form of the 3DVAR equation. A dual form of the 3DVAR is the analysis space algorithm, which is defined by the model state vector on some regular grid. Courtier (1997) has shown that the two formulations are equivalent and give the same solution. However, as discussed by Daley and Barker (2000, 2001), there are advantages to the use of an observation space approach in Navy ocean model applications. In the observation space algorithm the matrix to be inverted $(\mathbf{HP_bH^T + R})^{-1}$ is dimensioned by the number of observations, while in the analysis space algorithm the matrix to be inverted is dimensioned by the number of grid locations. Given the high dimensionality of global ocean forecast model grids, and the relatively sparse ocean observations available for the assimilation, an observation space 3DVAR algorithm will have a clear computational advantage. Further, an observation space system is more flexible and can more easily be coupled to many prediction models. As has been discussed, NCODA must work equally well with multiple atmospheric and oceanographic prediction systems in a wide variety of applications, as well as a wave model prediction system. Finally, due to the local nature of the observation space algorithm, the background error covariances are multivariate and can be formulated and generalized in a straightforward manner. As will be shown, this aspect of the 3DVAR is an important feature of NCODA. On the other hand, analysis space systems typically restrict the background error covariances to be sequences of univariate, one-dimensional digital filters, thereby ignoring the inherent multivariate nature of the background error correlations.

Solution of the observation space 3DVAR problem is done in two steps. First, the equation,

$$(HP_bH^T + R)z = [y - H(x_b)] \tag{13.2}$$

is solved for the vector $\mathbf{z}$. Next, a post-multiplication step is performed by left-multiplying $\mathbf{z}$ using,

$$x_a - x_b = P_bH^Tz \tag{13.3}$$

to obtain the correction field in grid point space. A pre-conditioned conjugate gradient descent algorithm is used to solve (13.2) using block diagonal pre-conditioners. The blocks are defined by decomposing the analysis grid into non-overlapping partitions of a regular quilt laid over the analysis domain in model grid point (i, j) space. The use of i, j blocks rather than latitude-longitude blocks allows the analysis to be completely grid independent. The flexibility of this approach is shown in Fig. 13.1 for the global HYCOM Atlantic basin analysis (see Sect. 13.6 and Fig. 13.9 for a discussion of the HYCOM basins). A total of 1,935, 2,436, and 1,147 blocks are defined for the global HYCOM Atlantic, Indian, and Pacific analysis regions, respectively, which use Mercator grid projections. Observations are sorted into the blocks and the pre-conditioning matrix is formed from a Choleski decomposition of the correlations between observations in the same block. The Choleski decomposed block matrices are calculated once and stored before application of the conjugate gradient descent algorithm. Solution of the pre-conditioned conjugate gradient for the vector **z** n (13.2) typically converges in 6–10 iterations. Determination of convergence is based on the norm of the gradient of the cost function estimated at each iteration step. This gradient is a vector the size of the number of observations and the norm is the square root of the sum of the elements, which are the residuals of the fit of the analysis to the innovations. In practice, convergence is reached when the norm of the gradient is reduced by 2 orders of magnitude. This is considered to be sufficient because an increase in the number of iterations only fits small-scale features. This may appear to be beneficial, but it must be noted that the post-multiplication step is a spatially smoothing operation when the background error correlations are applied. Thus, the extra iterations in the solver required to resolve small-scale features in the observations do not have much effect on the final analyzed increment field because of the smoothing effect of the post-multiplier.

Observation space 3DVAR algorithms converge quickly because very good pre-conditioners can be developed. In fact, the pre-conditioner used in NCODA is perfect. For example, NCODA is configured such that when the data count is less than 2,000 the observation data block is defined as the entire analysis domain. When this global pre-conditioned data block is presented to the conjugate gradient solver the algorithm converges in a single iteration. No further work by the solver is necessary. This sparse data pathway through the code is often encountered when NCODA 3DVAR is executed on nested grids in the relocatable coupled model system where the innermost grid represents a small geographic area.

As noted by Daley and Barker (2001), partitioning of the observations into blocks has no effect on the final solution. The NCODA 3DVAR formulation is guaranteed to include correlations between all observations in all blocks, thereby achieving a global solution. After the vector **z** is obtained it is post-multiplied by $\mathbf{P_b}\mathbf{H^T}$ to create the analysis correction fields for each analysis variable. This step is performed using blocks in grid space that are defined differently from the observation blocks used to compute the solution vector **z**. To accommodate high resolution ocean model forecast grids and minimize computer memory resource requirements for the analysis, the grid space blocks are defined smaller by simply
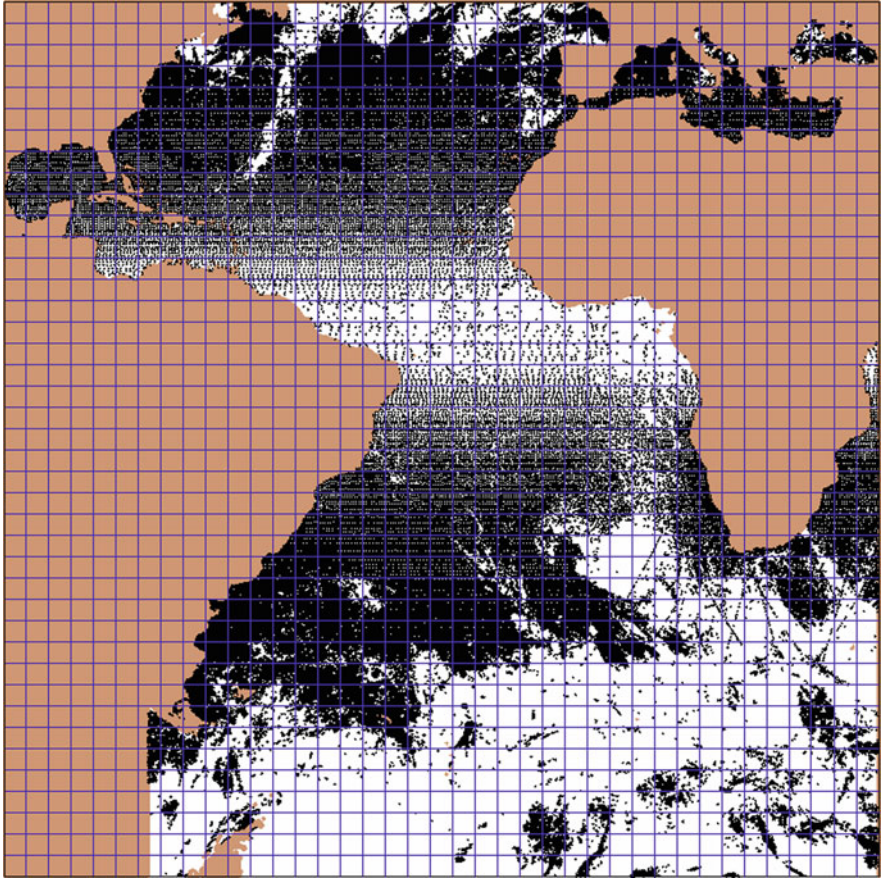
Fig. 13.1 Observation data blocks for HYCOM Atlantic basin grid. *Blue lines* give observation block edges; observation locations are indicated by *black dots*. A total of 1,935 data blocks are defined (43 in the X direction, 45 in the Y direction)

sub-setting the previously defined observation blocks. Again, it must be emphasized that partitioning the grid domain into blocks in the post multiplication does not affect the final results. The correction fields are guaranteed to contain the correlations between all observations and all grid points, thereby creating a seamless and continuous analysis.

Parallelization of the 3DVAR algorithm is achieved in three ways. The first parallelization is done over the observation-defined blocks in the pre-conditioner, the second parallelization is done over observation-defined blocks in the conjugate gradient solver, and the third parallelization is done over grid point-defined blocks in the post-multiplication step (mapping from observation space to grid space). The number of processors used to execute the 3DVAR can be as few as one or as many as the maximum number of observation or grid node blocks. A load balancing

algorithm is used to spread the work related to the block-dependent calculations out evenly across the processors. In the conjugate gradient descent step, the work load for an observation block is calculated as the sum of the observation-observation interactions. In the post-multiplication step, the work estimate is based on the sum of the observation-grid point interactions. Observation and grid point blocks are determined to be close enough to contribute to the solution if the block centers are within 8 correlation length scales. Thus, for a given block size, the number of observation-observation and observation-grid point block interactions varies with the horizontal correlation length scales and will be more numerous where length scales are long. Further efficiency is achieved by keeping communication among the processors minimal. To do this matrix elements are calculated, stored, and used on each processor, they are never passed between processors. Only elements of the solution and correction vectors scattered across the processors have to be communicated and reassembled and, in the case of the solution vector, broadcast for the next iteration. Note that memory utilization for the conjugate gradient solver in the 3DVAR is reduced as the number of processors is increased. This feature allows the 3DVAR to scale very well across many processors on large machines, and run equally well on small platforms with limited memory.

13.3 Error Covariances

Specification of the background and observation error covariances in the assimilation is very important. As previously noted, the background error covariances control how information is spread from the observations to the model grid points and model levels, but they also ensure that observations of one model variable produce dynamically consistent corrections in the other model variables. The background error covariances in the NCODA 3DVAR are similar to the error covariances defined for the MVOI, but with some notable exceptions. As in the MVOI, the error covariances in the 3DVAR are separated into a background error variance and a correlation. The correlation is further separated into a horizontal (C_h) and a vertical (C_v) component. Correlations are modeled as either second order auto-regressive (SOAR) functions of the form,

$$C_h = (1 + s_h) \exp(-s_h)$$
$$C_v = (1 + s_v) \exp(-s_v) \tag{13.4}$$

or Gaussian functions of the form,

$$C_h = \exp(-s_h^2)$$
$$C_v = \exp(-s_v^2) \tag{13.5}$$

where s_h and s_v are the horizontal and vertical distances between observations or observations and grid points, normalized by the arithmetic mean of the horizontal or

the vertical correlation length scales at the two locations. The horizontal correlation length scales vary with location and the vertical correlation length scales vary with depth and location in the analysis. As described in the subsequent sections, both correlation components evolve with time in accordance with information obtained from the model forecast background valid at the update cycle interval.

13.3.1 Horizontal Correlations

The horizontal correlation length scales are set proportional to the first baroclinic Rossby radius of deformation using estimates computed from the historical profile archive by Chelton et al. (1998). Rossby length scales qualitatively characterize scales of ocean variability and vary from 10 km at the poles to greater than 200 km in the tropics. The Rossby length scales increase rapidly near the equator which allows for stretching of the zonal scales in the equatorial wave guide. Flow-dependence is introduced in the analysis by modifying the horizontal correlations with a tensor computed from forecast model sea surface height (SSH) gradients. The flow-dependent tensor spreads innovations along rather than across the SSH contours, which are used as a proxy for the circulation field. Flow dependence is a desirable outcome in the analysis, since error correlations across an ocean front are expected to be characteristically shorter than error correlations along the front. Note that other gradient fields can be used as a flow-dependent tensor in the analysis, such as SST or potential vorticity (Martin et al. 2007). The flow dependent correlation tensor (C_f) is computed using either a SOAR or Gaussian model defined in (13.4) and (13.5), where the SSH difference between two locations is normalized by a scalar that defines the strength of the flow dependence. Because the flow dependent correlations are computed directly from the forecast SSH fields they depend strongly on the accuracy of the model forecast. This dependence may prove not to be very useful in practice if the forecast model fields are inaccurate. Accordingly, the normalization scalar can be set to a relatively large value in order to reduce the strength of the flow dependence in the analysis and prevent a model with systematic errors from adversely affecting the analysis. Alternatively, the flow dependence can be switched completely off. Figure 13.2 shows a zoom of the analysis increments off South Africa from a global high resolution SST analysis executed using a 6-h update cycle. The analysis has 12-km resolution at the equator, 9-km mid-latitude, and is a FNMOC contribution to the Group for High Resolution SST (GHRSST). Background SST gradients are used as the flow dependent tensor, with the result that the SST analysis increments are constrained by the meanders and eddies associated with the Agulhas retroflection current. The increments are both positive and negative along the front and eddy locations, indicating that application of the flow dependent tensor is a relatively weak constraint and the strength and position of features can change from one update cycle to the next in the analysis.

    To account for the discontinuous and non-homogeneous influence of coastlines in the analysis a second tensor is introduced (C_l) that rotates and stretches horizontal
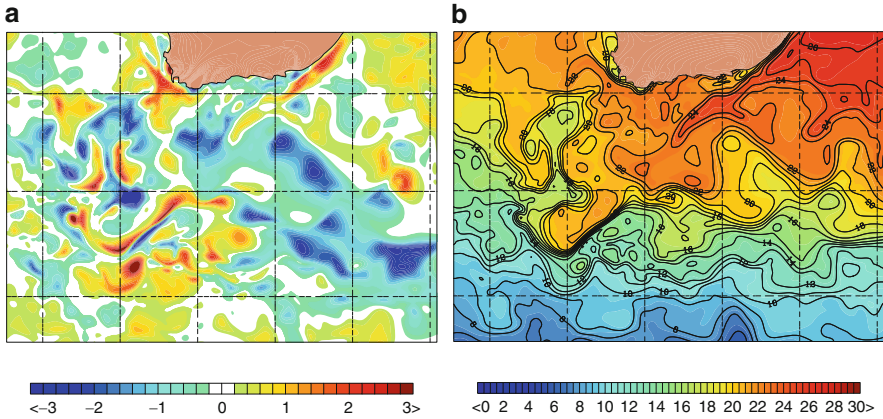
Fig. 13.2 Analysis example of flow dependent tensor based on SST gradients in Agulhas Current region; scalar value defining gradient strength of flow dependence set to 0.5°C. (**a**) analyzed increments; (**b**) analyzed SST field

correlations along the coast while minimizing or removing correlations into the coast. First, all observations and model grid points are assigned an orthogonal distance to land value based on a 1-km global coastline database. Land distances greater than some minimal value (say, 20 km) are set to the minimal value. This operation results in land distance gradients greater than zero along coastlines and zero elsewhere. Similar to the flow dependence tensor, the coastline tenor is then calculated using the difference in land distance between two locations normalized by a scalar that specifies the strength of the coastline dependence. Away from the coast (>20 km) this difference is zero resulting in no modification of the horizontal correlations. However, in the vicinity of the coast (<20 km) land distance differences are non-zero, resulting in $C_l < 1$ and a modification of the horizontal correlations. Background error correlations close to the coast are expected to be anisotropic because horizontal advection from coastal currents will elongate the corrections and spread the information along the coast. Figure 13.3 illustrates the coastline tensor applied to an observation ∼5 km from the coast in Monterey Bay. In this example, the horizontal correlations are specified as homogenous with a length scale of 30 km. The effect of the coastline tensor is clearly seen as the correlations adjust to prominent coastal features like the Monterey peninsula to the south and the rotation of the coastline to an east–west orientation north of the observation location.

The total horizontal background error correlation (C_b) is then computed as the product of the two correlation components and the two correlation tensors according to,

$$C_b = C_h C_v C_f C_l \tag{13.6}$$

Fig. 13.3 Example of land
distance correlation tensor for
point 4.8 km from coast in
Monterey Bay, California,
USA. Observation point is
given by white X mark.
Horizontal length scales are
assumed homogenous at
30 km. The land distance
tensor spreads the
correlations from the
observation point along the
contours of the Monterey Bay
coastline
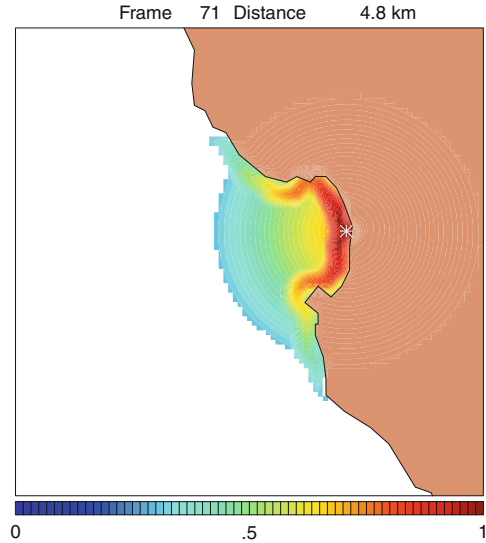
13.3.2 Vertical Correlations

Vertical correlation length scales vary with location and depth and evolve from one
analysis cycle to the next in the 3DVAR. They are defined on the basis of either:
(1) background density vertical gradients in pressure space, or (2) background
density differences in isopycnal space. In the vertical density gradient option, a
change in density stability criterion is used to define a well-mixed layer. The change
in density criterion is then scaled by the background vertical density gradient at each
grid location and grid level according to,

$$h_v = \rho_s/(\partial\rho/\partial z) \tag{13.7}$$

where h_v is the vertical correlation length scale, ρ_s is the change in density criterion
($\sim 0.15 \text{ kg m}^{-3}$), and $\partial\rho/\partial z$ is the vertical density gradient. Surface mixed layer
depths, calculated at each grid point using the same change in density criteria
(Karra et al. 2000), are spliced onto the three-dimensional vertical length scale
field computed using (13.7). With this modification, surface-only observations
decorrelate at the base of the spatially varying mixed layer. The vertical density
gradient correlations are computed each update cycle from the background den-
sity fields, thereby allowing the vertical scales to evolve with time and capture
changes in mixed layer, thermocline depths, and the formation of mode waters.
Overall, the method produces vertical correlation length scales that vary with depth
and location, and are long when the water column stratification is weak and short
when the water column is strongly stratified.

In the isopycnal option, observation or grid point differences in density are scaled by ρ_s to form a correlation. This procedure essentially derives the vertical correlations relative to a density vertical coordinate. Observations are more correlated along an isopycnal than across an isopycnal, which introduces considerable flow dependence into the correlations. The procedure is cost free and does not require a transformation of the model background to isopycnal coordinates. All that is needed is knowledge of the density for any point of interest, which can be obtained from the observation itself or the model forecast. Use of the isopycnal vertical correlation option is ideally suited for HYCOM, since each coordinate surface in the model is assigned a reference isopycnal. Vertical correlation defined along isentropic surfaces is well known in atmospheric data assimilation (e.g., Riishøjgaard 1998). Note that vertical correlations in the analysis are calculated either via a SOAR, (13.4) or Gaussian, (13.5) function using lengths scales derived from either the vertical density gradient or isopycnal formulations.

Figure 13.4 gives cross sections through the vertical correlation length scale field and the model density field for the HYCOM Pacific domain (Sect. 13.6). The length scales were computed using the vertical density gradient option with $\rho_s = 0.15$. The cross sections extend from the coast of Japan at 42°N, 140°E along a great circle path to the equator at 0°N, 160°E. Figure 13.4a shows vertical correlation length scales shorter near the surface and longer at depth in agreement with the density stratification (Fig. 13.4b). The influence of the Kuroshio front is clearly seen, with longer length scales at increasingly shallower depths as the permanent thermocline shoals towards the equator. Relatively longer length scales are also seen in the 17–19°C mode-water layer immediately south of the Kuroshio, which has relatively uniform density at depths of 200–400 m.

13.3.3 Multivariate Correlations

The horizontal and vertical correlation functions described above are used in the analysis of temperature, salinity, and geopotential. Temperature and salinity are analyzed as uncorrelated scalars, while the analysis of geopotential is multivariate with velocity. Geopotential is computed in the analysis from vertical profiles of temperature and salinity by integrating the specific volume anomaly (Fofonoff and Millard 1983) from a level of no motion (2,000 m depth) to the surface. The multivariate correlations require specification of a parameter γ, which measures the divergence permitted in the velocity correlations, and a parameter φ, which specifies the strength of the geostrophic coupling of the velocity/geopotential correlations. Typically, γ is set to a small, constant value ($\gamma = 0.05$) that produces weakly divergent velocity increments and assumes that the divergence is not correlated with changes in the mass field. The geostrophic coupling parameter φ varies with location from 0 to 1. It is scaled to zero within 1° of latitude from the equator, where geostrophy is not defined, and in shallow water (<50 m deep), where friction rather than pressure gradient forces control ocean flow. The multivariate correlations
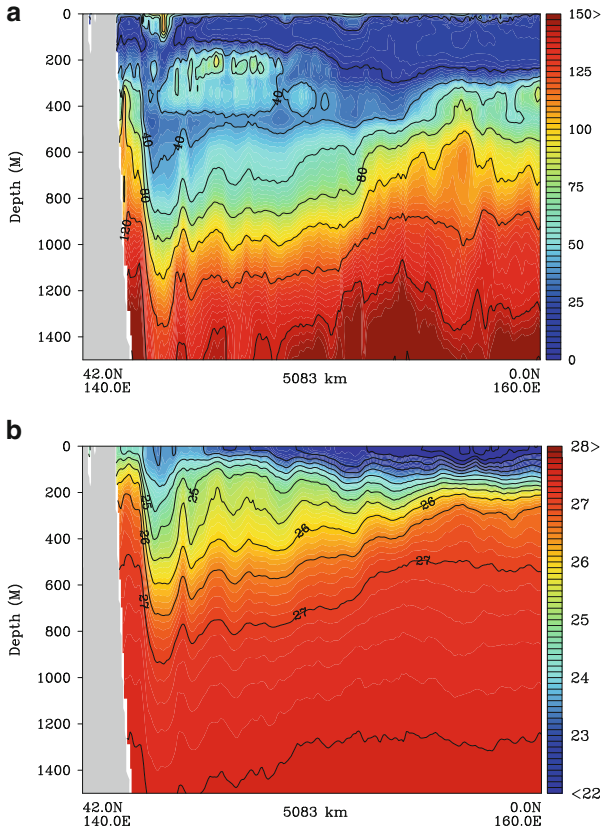
Fig. 13.4 Cross sections of vertical correlation length scales and density from Pacific basin run of global HYCOM. (**a**) Vertical length scales (m); (**b**) Density (kg/m³)

also include auto- and cross-correlations of the u, v vector velocity components. However, at the present time, there are no operational sources of ocean current observations available for the assimilation, although the capability to assimilate velocity data is built into the 3DVAR system. A full derivation of the multivariate horizontal correlations is provided in Daley (1991). The multivariate correlations are derived from the first and second derivatives of the SOAR (or Gaussian) model function and require precise calculation of the angles between any two locations in order to guarantee a symmetric correlation matrix.
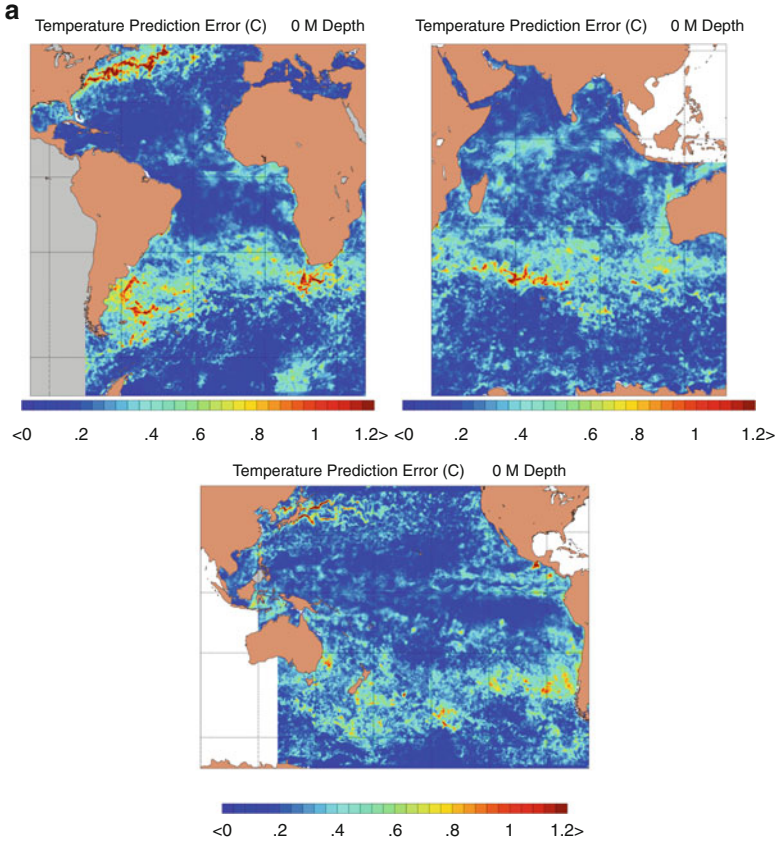
13.3.4 Background Error Variances

Background error variances are poorly known in the ocean and are likely to be strongly dependent on model resolution and other factors, such as atmospheric

model forcing errors and ocean model parameterization errors. In the analysis, the background error variances ($\mathbf{e_b^2}$) vary with location, depth, and analysis variable. The variances are computed prior to an analysis from a weighted time history of differences in forecast fields valid at the update cycle interval and issued from a series of analyzed states according to,

$$e_b^2 = \sum_{k=1}^{n} w_k (x_k - x_{k-1})^2 \tag{13.8}$$

where $\mathbf{x_k} - \mathbf{x_{k-1}}$ are the differences in model forecasts (indices indicating grid location and depth are omitted for clarity), k is the update cycle index, n is the number of update cycles into the past to use in the summation, and w_k is a weight vector computed using a geometric series, $w_k = (1 - \phi)^{k-1}$, where ϕ is typically set to 0.1. The background error variances computed according to (13.8) are normalized such that the weighted averages are unbiased. In practice, the background error variances tend to evolve to a quasi-steady state over time. The model forecast difference fields include the influence of observations from the assimilation, so in well observed areas the background errors are consistent with the innovations (model-data errors at the update cycle interval). However, in the case of poorly observed or strong flow areas the background error variances are more likely dominated by model variability arising from atmospheric forcing and baroclinic and barotropic instabilities. Figure 13.5 shows background temperature error standard deviation computed using Eq. (13.8) for different vertical levels in the global HYCOM analysis domains (see Sect. 13.6). Figure 13.6 shows the background salinity error standard deviation and Fig. 13.7 the background velocity error standard deviation at the surface. Relatively high background errors are evident at all depths in boundary current areas: Gulf Stream, Kuroshio, Agulhas, Brazil-Malvinas, East Australia. Surface salinity error levels are also large near some river outflow areas, in tropical regions, and in the marginal ice zone around Antarctica during the Austral summer. Surface velocity error standard deviations tend to be large in western boundary currents and in the inter-tropical convergence zone (ITCZ) due to the variable wind and solar forcing in that area.

The adaptive scheme implemented here is designed to provide background errors that: (1) are appropriate for the time interval at which data are inserted into the model; (2) are coherent with the variance of the innovation time series; (3) reflect the variable skill of the different ocean forecast models that are used with the analysis system; and (4) adjust quickly to new ocean areas when the analysis is re-located in a rapid environmental assessment mode of operation. One difficulty with this approach is that differences in model fields contain a mixture of forecast and analysis error. Forecast errors result from initial condition, model, and atmospheric forcing deficiencies, while analysis errors result from the fact that the statistical parameters used in the analysis represent expected values and are unlikely to be correct at all places and at all times.

Fig. 13.5 Temperature (°C) background error standard deviations valid 20 January 2012 in global HYCOM analysis domains: Atlantic, Indian, and Pacific. (**a**) 0 M depth, (**b**) 150 M depth, (**c**) 300 M depth

13.3.5 Observation Error Variances

The observation errors and the background errors are assumed to be uncorrelated, and errors associated with observations made at different locations and at different times are also assumed to be uncorrelated. As a result of these assumptions, the observation error covariance matrix **R** is set equal to $1 + E_o^2$ along the diagonal and zero elsewhere. Note that E_o^2 represents observation error variances (e_o^2) normalized by the background error variances interpolated to the observation location ($E_o^2 = e_o^2/e_b^2$). Observation errors are computed as the sum of a measurement error and a representation error. Measurement errors reflect the accuracy of the instruments and the ambient conditions in which the instruments operate. These errors are fairly well known for many ocean observing systems, although the errors can change in time due to calibration drift of the instruments and other
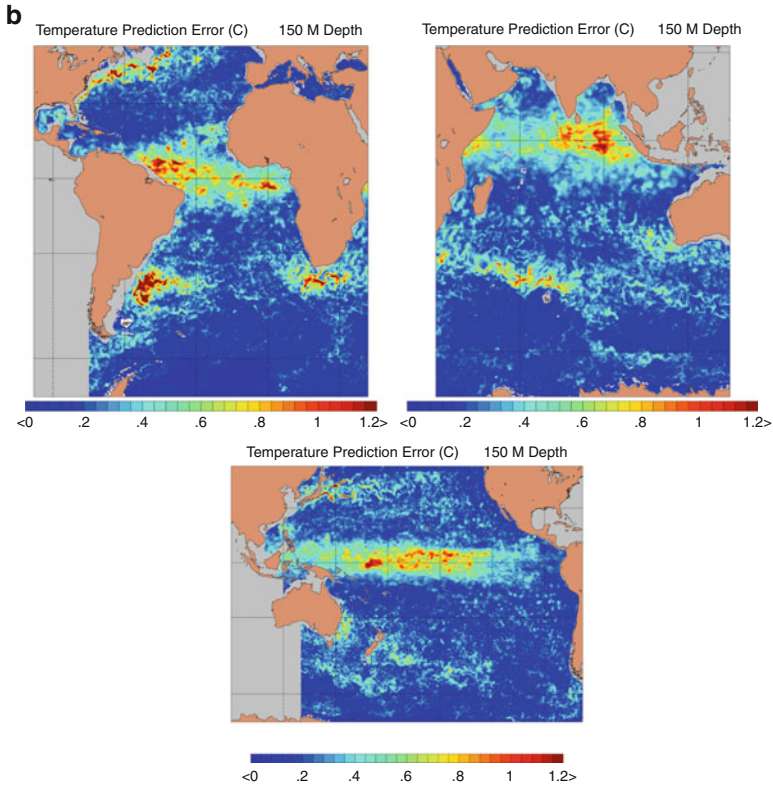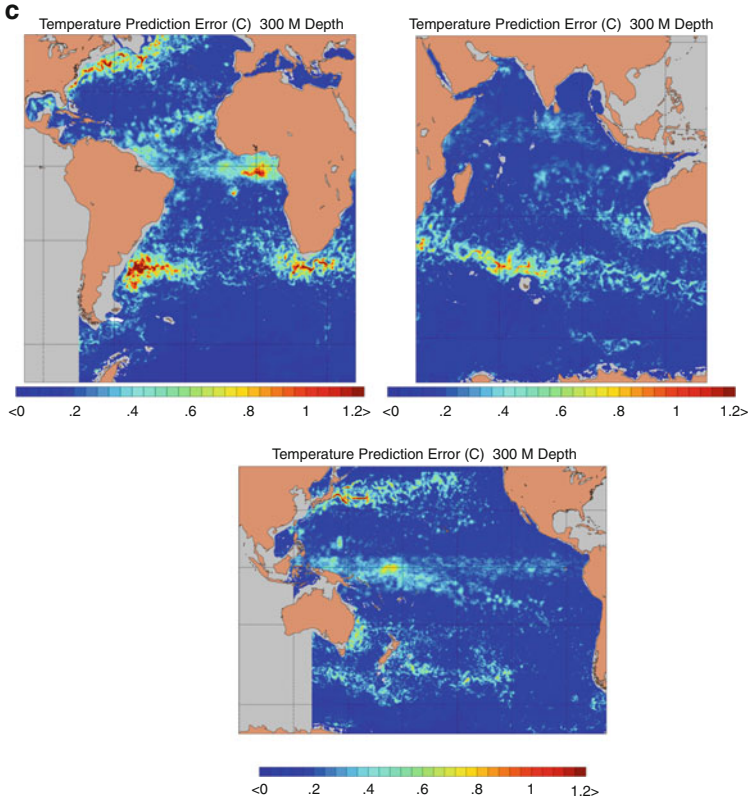
Fig. 13.5 (continued)

factors. Representation errors, however, are a function of the resolution of the model and the resolution of the observing network. For satellite retrievals with known measurement footprints, representation errors are set equal to the gradient of the background field at the observation location when the retrieval footprint exceeds the model grid resolution. Representation error of profile observations consists of two additive components. The first component is set proportional to the observed profile vertical gradients of temperature and salinity as a proxy for uncertainty associated with internal waves. The second component is estimated from the variability of multiple observed profile level data averaged into layers defined by the model vertical grid (see Sect. 13.4.2).

13.4 Ocean Observations

The analysis makes full use of all sources of the operational ocean observations. Ocean observing systems assimilated by the 3DVAR are listed in Table 13.1, along
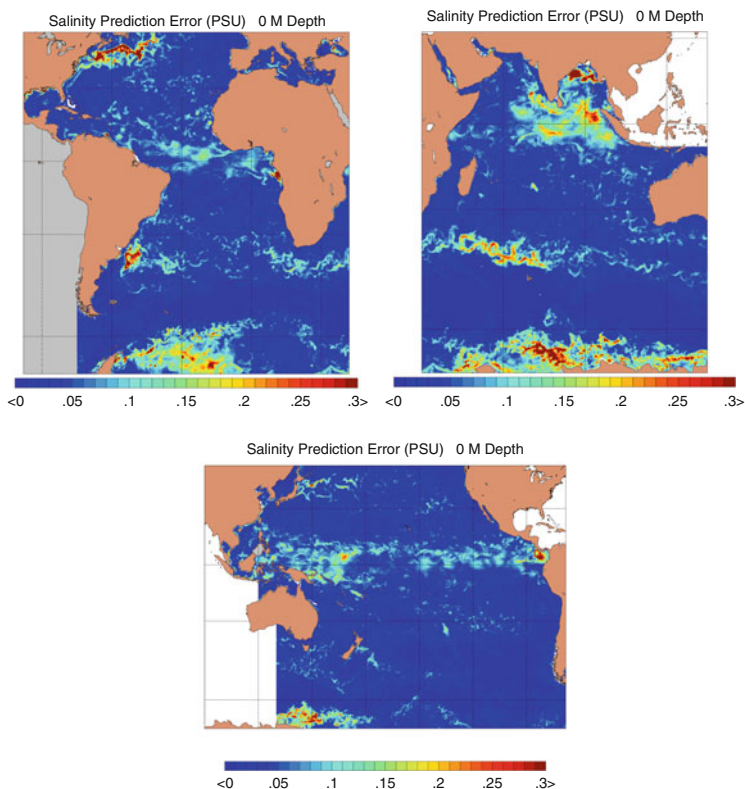
Fig. 13.5 (continued)

with typical global data counts per day. All ocean observations are subject to data quality control (QC) procedures prior to assimilation. The need for quality control is fundamental to a data assimilation system. Accepting erroneous data can cause an incorrect analysis, while rejecting extreme, but valid, data can miss important events. The NCODA 3DVAR analysis was co-developed and is tightly coupled to an ocean data QC system. Cummings (2011) provides an overview of the NCODA ocean data quality control procedures.

13.4.1   Surface Observations

Table 13.1 indicates that there are many high volume sources of satellite and in situ SST, SSH, and sea ice observations. It is not uncommon to assimilate ~40 million satellite SST retrievals, ~2 million sea ice concentration retrievals, and ~500,000

Salinity Prediction Error (PSU) 0 M Depth

Salinity Prediction Error (PSU) 0 M Depth

Salinity Prediction Error (PSU) 0 M Depth

Fig. 13.6 Surface salinity (PSU) background error standard deviations valid 20 January 2012 in global HYCOM analysis domains: Atlantic, Indian, and Pacific

altimeter SSH observations in a single day. These high-density, surface-only, data types must be thinned prior to the analysis to remove redundancies in the data and minimize horizontal correlations among the observations. The data thinning is achieved by averaging innovations into bins with spatially varying sizes defined using the ratio of horizontal correlation length scales and horizontal grid resolution. Innovations are inversely weighted based on observation error in the data thinning process, and in the case of SST observations the water mass of origin is maintained (see Cummings 2005 for a discussion of the Bayesian water mass classification scheme). The length scale to grid mesh ratio bin sizes automatically adjust to changes in the spatially varying horizontal correlation length scales, but are never smaller than the underlying model grid mesh. As a result, fewer data are thinned as the grid resolution decreases or as the correlation length scales shorten. This adaptive feature of the data thinning process can be used to decrease (increase) the amount of data thinning by artificially shortening (lengthening) the horizontal correlation length scales given a fixed model grid. Note that simply increasing data
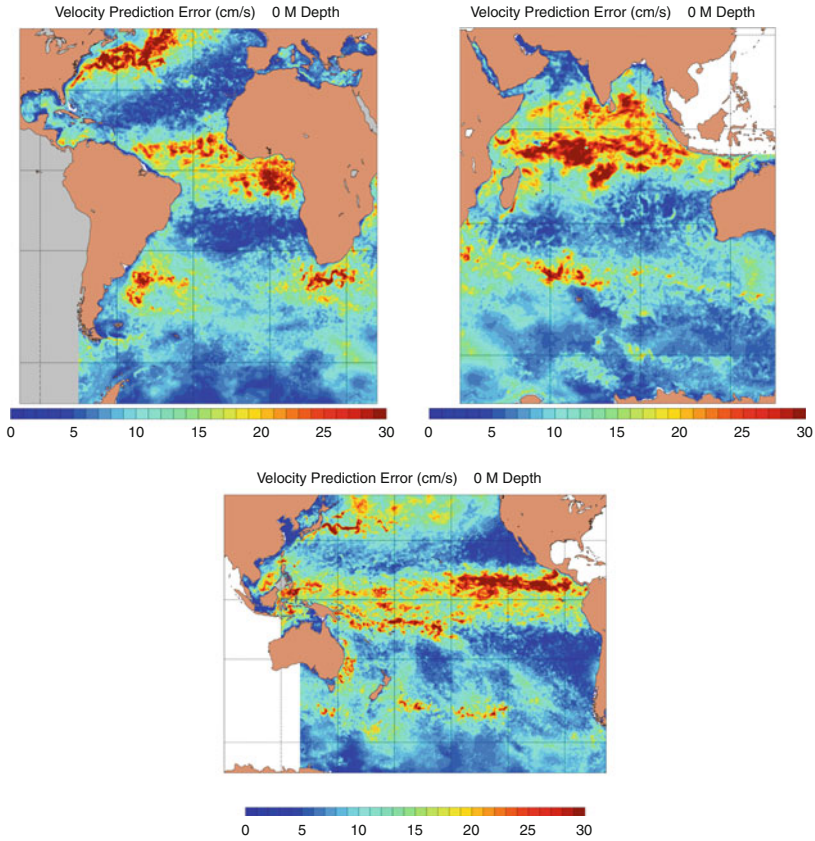
Fig. 13.7 Surface velocity (cm/s) background error standard deviations valid 20 January 2012 in global HYCOM analysis domains: Atlantic, Indian, and Pacific

density does not necessarily improve the analysis. More data will require more conjugate gradient iterations while, more importantly, it may not noticeably alter the results given the smoothing operation of the post-multiplication step (see discussion in Sect. 13.2). Figure 13.8 shows an example of data thinning results for 6 h of satellite SST observations in the FNMOC GHRSST analysis. Even with just 6 h of SST data the various satellite missions and in situ sources show a high degree of spatial overlap. The data thinning removes this data redundancy and creates a sampling pattern consistent with the horizontal correlation length scales defined for the analysis. In this case, length scales are based on Rossby radius of deformation, which varies significantly across the grid. As a result, there is increased data thinning near the equator where length scales are ∼200 km. Elsewhere, especially at high latitude, the data thinning is much less, and satellite retrievals with footprint resolutions of 2 km and 8 km are directly assimilated without any spatial averaging.

Table 13.1 Data types assimilated in NCODA 3DVAR with typical daily data counts. Note that the profile data counts are for the entire profile. Profiles typically contain hundreds of levels that are assimilated as unique latitude, longitude, level observations

Data type	Data source	Specifications	Number daily obs
Satellite SST	NOAA-18 NOAA-19	Infrared 2-km day, night retrievals	4,800,000
	NOAA-18 NOAA-19	Infrared 8-km day, night, relaxed day retrievals	800,000
	AMSR-E	Microwave 25-km day, night retrievals	3,600,000
	METOP-A	Infrared 2-km day, night retrievals	15,000,000
	METOP-A	Infrared 8-km day, night, relaxed day retrievals	450,000
	GOES E/W	Infrared 12-km day, night retrievals	2,000,000
	MeteoSat-2	Infrared 8-km day, night retrievals	220,000
	AATSR	Infrared 1-km day, night retrievals	12,000,000
In Situ SST	Ships	Engine room intake	6,500
		Hull contact sensor	1,000
		Bucket temperature	100
		CMAN Station	100
	Drifting Buoy		34,000
	Fixed Buoy		7,000
Satellite altimeter	Jason 1, 2 Envisat	SSHA	150,000
		SWH	180,000
Sea ice concentration	DMSP F13, F14, F15	SSM/I 25-km retrievals	900,000
	DMSP F16, F17, F18	SSMIS 25-km retrievals	1,200,000
Profiles	Drifting buoy	Temperature	50
	Fixed buoy		1,200
	Argo		600
	XBT		100
	TESAC (CTD)		3,500
	Drifting buoy	Salinity	50
	Fixed buoy		800
	Argo		600
	TESAC (CTD)		3,000

13.4.2 Profile Observations

Preparation of profile observations for the assimilation consists of several steps. First, observed profiles are extended to the bottom using the model forecast. The
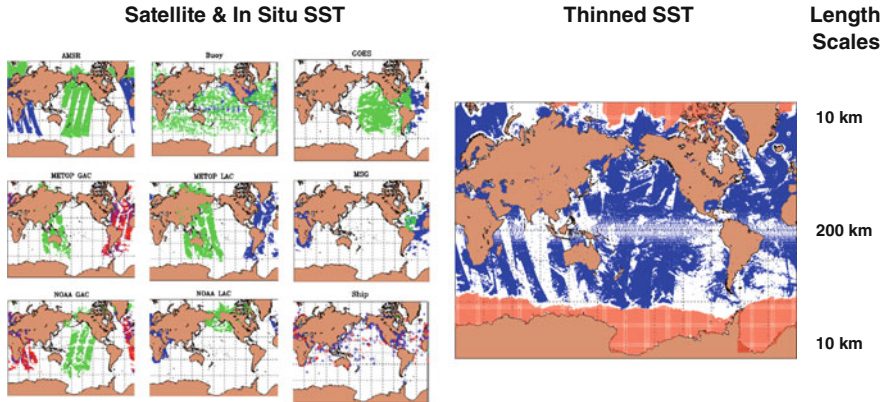
Fig. 13.8 Data thinning of global SST data. Satellite and in situ sources SST show in *left panel* (blue daytime, green nighttime, red relaxed day satellite retrieval types). The SST data sources are (in order from top to bottom): AMSR-E, Drifting and Fixed Buoy, GOES E/W, METOP-A GAC, METOP LAC, MeteoSat-2, NOAA 18,19 GAC, NOAA 18,19 LAC, Surface Ship (engine room intake, bucket, hull contact sensor). Thinned data for assimilation is show in *middle panel* (*blue*— SST observation; *red*—freezing sea water under ice covered seas). Schematic of how correlation lengths vary as a function of latitude shown on *right*

observed profile is merged to the forecast profile by selecting the depth at which the merge is complete based on the shape of the extracted forecast model profile. This target depth is set to be the second zero crossing of the forecast profile curvature. Note that the merge can fail if a suitable target depth is not found or if the difference between the observed and model profile at the merge depth is too large (>3°C for temperature; >0.1 PSU for salinity). Second, similar to the high density surface-only data, profile observations are thinned in the vertical to remove redundant data. The profile thinning is done by averaging temperature and salinity observations at observed levels within vertical layers defined by the mid-points of the model vertical grid. Since the ocean circulation models interfaced with the 3DVAR have very different vertical coordinates (NCOM uses a sigma/z vertical grid; HYCOM uses a z/isopycnal/sigma hybrid vertical grid), model vertical levels at the grid point closest to the profile location are used to define layer thicknesses. Third, in cases where profile vertical sampling is inadequate to resolve the local vertical correlation length scales, the profile is expanded in the vertical by linearly interpolating data to interleaving levels in order to form a more vertically dense profile. This scheme ensures vertically smooth analysis increments at all model levels even when vertical correlations are short due to strong density stratification. This situation routinely occurs in the tropics with the sparse vertical sampling in profiles received from the Tropical Atmosphere Ocean (TAO), Triangle Trans-Ocean Buoy Network (TRITON), and Prediction and Research Moored Array in the Atlantic (PIRATA) buoys. It is clear that the vertical sampling of the tropical mooring arrays needs to be improved.

13.4.3 Altimeter Sea Surface Height

Table 13.1 shows that most ocean observations are remotely sensed and measure ocean variables only at the surface. The lack of synoptic real-time data at depth places severe limitations on the ability of the data assimilation system to resolve and maintain an adequate representation of the ocean mesoscale. Subsurface properties in the ocean, therefore, must be inferred from surface-only observations. The most important observing system for this purpose is satellite altimetry, which measures the time varying change in SSH. Changes in sea level are strongly correlated with changes in the depth of the thermocline in the ocean, and the ocean dynamics generating sea level change are for the most part the mesoscale eddies and meandering ocean fronts. The SSH data are provided as anomalies relative to a time-mean field. The time mean removes the unknown geoid, but it also removes the mean dynamic topography (MDT), which needs to be added back in order to allow the data to be compared with model fields. The 3DVAR determines the satellite altimeter SSH sampling locations in two alternative ways: (1) direct assimilation of the along-track data at the observed locations, or (2) by first performing a 2D horizontal analysis of SSH and then generate a sampling pattern of synthetic profile locations within contours of sea level change that exceed a prescribed noise level threshold (see Cummings 2005 for details). Once the altimeter sampling locations are known there are two alternative methods available in the 3DVAR to project the SSH data to depth in the form of synthetic temperature and salinity profiles. One method is the Modular Ocean Data Assimilation System (MODAS) database, which models the time averaged co-variability of dynamic height vs. temperature at depth and temperature vs. salinity at a fixed location from an analysis of historical profile data (Fox et al. 2002). The MDT used in the MODAS method is derived from historical hydrographic data. Note that an upgrade to the MODAS synthetic profile capability, the Improved Synthetic Ocean Profile (ISOP) system (Helber et al. 2012), is currently being evaluated. The second "direct" method adjusts the model forecast density field to be in agreement with the difference found between the model forecast sea surface height field and the SSH measured by the altimeter (Cooper and Haines 1996). The MDT used in the direct method is the mean SSH from the model derived from a hindcast run. Output of the direct method is in the form of innovations of temperature and salinity from the forecast model background field, which are directly input into the assimilation. An advantage of the direct method is that it relies on model dynamics for its prior information rather than statistics fixed at the start of the assimilation. However, a disadvantage is that it cannot explicitly correct for forecast model errors in stratification due to model drift in the absence of any real data constraints. MODAS does not suffer from these limitations, although MODAS may have marginal skill due to: (1) sampling limitations of the historical profile data, (2) non-steric signals in the altimeter data, or (3) weak correlations between steric height and temperature at depth due to other factors, such as the influence of salinity on steric height at high latitudes. Needless to say, neither of the methods available for assimilating altimeter SSH data is ideal. A new method under

development assimilates altimeter SSH by conversion of the along-track SSH slopes to geostrophic velocity profiles. This method is described briefly in Sect. 13.7.

While having the potential of adding important information in data-sparse areas, the number of altimeter-derived synthetic observations computed can greatly exceed and overwhelm the in situ observations in the analysis. Accordingly, the synthetic observations are thinned prior to the analysis in four ways. First, it is assumed that directly observed temperature and salinity profiles are a more reliable source of subsurface information wherever such observations exist. Altimeter-derived synthetic profiles, therefore, are not generated in the area surrounding an in situ profile observation. Second, the observed SSH from the along-track data or the analyzed incremental change in sea level must exceed a threshold value, defined as the noise level of the satellite altimeters, to trigger the generation of a synthetic observation. This value is typically set to 4 cm. Third, projection of the SSH signal onto the model subsurface density field can produce unrealistic results when the vertical stratification is weak. In the absence of specific knowledge about how to partition SSH anomaly into baroclinic and barotropic structures in these weakly stratified regions, synthetic profiles are rejected for assimilation when either of the following occurs: (1) the top-to-bottom temperature difference of the MODAS synthetic profile is less than 5°C; or (2) the maximum value of the Brunt-Väisälä frequency calculated from the model density profile in the direct method is less than 1.4. Fourth, there are problems with the SSH data in shallow water due to contamination of the altimeter signal by tides. Accordingly, SSH data are not assimilated in water depths less than 400 m.

13.5 NCODA System

NCODA is a comprehensive ocean data assimilation system. In addition to the 3DVAR it contains other components that perform functions useful for many applications. These component capabilities are briefly summarized in this section.

13.5.1 Analysis Error Covariance

The analysis error covariance $\mathbf{P_a}$ is estimated from the equation,

$$P_a = P_b - P_b H^T (H P_b H^T + R)^{-1} H P_b \qquad (13.9)$$

where $\mathbf{P_b}$ and $\mathbf{R}$ are the background and observation error covariances previously defined for (13.1). Unlike (13.1), which involves matrix–vector operations, (13.9) requires the use of matrix-matrix operations and is computationally expensive to perform. The NCODA 3DVAR provides an estimate of the analysis error variance (the diagonal of the second right-hand term) in the form of a normalized reduction

of the forecast error ranging from 1 (0 % reduction) to 0 (100 % reduction) for each analysis variable at all model grid points. The analysis error solution is a local approximation performed within the grid decomposition blocks that is improved upon though the use of halo regions to bring in the influence of additional observations. The analysis error estimation uses the same data inputs as the 3DVAR other than the innovations. In this way the analysis error calculation can be done at the same time as the analysis, albeit on a different set of processors, to improve throughput of the entire data assimilation system. The primary application of the analysis error covariance program is as a constraint in the Ensemble Transform technique (Sect. 13.5.3).

13.5.2 Adjoint

Adjoint-based observation sensitivity provides a feasible all at once approach to estimating observation impact. Observation impact is calculated in a two-step process that involves the adjoint of the forecast model and the adjoint of the assimilation system. First, a cost function ($\mathbf{J}$) is defined that is a scalar measure of some aspect of the forecast error. The forecast model adjoint is used to calculate the gradient of the cost function with respect to the forecast initial conditions ($\partial \mathbf{J}/\partial \mathbf{x_a}$). The second step is to extend the initial condition sensitivity gradient from model space to observation space using the adjoint of the assimilation procedure ($\partial \mathbf{J}/\partial \mathbf{y} = \mathbf{K}^T \partial \mathbf{J}/\partial \mathbf{x_a}$), where $\mathbf{K} = \mathbf{P_b}\mathbf{H^T}[\mathbf{HP_b}\mathbf{H^T} + \mathbf{R}]^{-1}$ is the Kalman gain matrix of (13.1) and the adjoint of $\mathbf{K}$ is given by $\mathbf{K^T} = [\mathbf{HP_b}\mathbf{H^T} + \mathbf{R}]^{-1}\mathbf{HP_b}$. The only difference between the forward and adjoint of the analysis system is in the post-multiplication of going from the solution in observation space to grid space. The pre-conditioned, conjugate gradient solver $[\mathbf{HP_b}\mathbf{H^T} + \mathbf{R}]$ is symmetric or self-adjoint and operates the same way in the forward and adjoint directions. The NCODA 3DVAR adjoint was coded directly from the forward 3DVAR by transposition of the post-multiplier to a pre-multiplier that is invoked first to convert adjoint sensitivities from grid space to observation space prior to execution of the solver for calculation of observation sensitivities and data impacts.

13.5.3 Ensemble Transformation

The ensemble transform (ET) ensemble generation technique (Bishop and Toth 1999) transforms an ensemble of forecast perturbations into an ensemble of analysis perturbations. The method ensures that the analysis perturbations are consistent with the analysis error covariance matrix ($\mathbf{P_a}$), computed using (13.9). To compute the required transform matrix an eigenvector decomposition is performed,

$$(X_f^T P_a^{-1} X_f)/n = C \lambda C^T \tag{13.10}$$

where $\mathbf{X_f}$ is the matrix of ensemble forecast perturbations about the ensemble forecast mean, $\mathbf{P_a}$ is the analysis error covariance matrix, n is the number of model variables (state vector), and C are the eigenvectors and λ the eigenvalues of the left hand side of (13.10). Superscript T indicates matrix transpose. Given the eigenvector decomposition the transformation matrix $\mathbf{T}$ is given by $\mathbf{T} = \mathbf{C}\lambda^{-1/2}\mathbf{C^T}$, which is used to transform a matrix of forecast perturbations to a matrix of analysis perturbations according to $\mathbf{X_a} = \mathbf{X^f T}$. If the ensemble size is large enough it can be shown that the covariance of the analysis perturbations equals the prescribed analysis error covariance $\mathbf{P_a}$ (McLay et al. 2008). Thus the analysis error covariance is an effective constraint in the ET, ensuring that the ensemble generation system is consistent with the data assimilation system.

The NCODA ET is multivariable and computes the transformation matrix for temperature, salinity, and velocity simultaneously. As a result the NCODA ET perturbations are balanced and flow dependent. In an ET ensemble generation scheme the control run is the only ensemble member that executes the 3DVAR. This results in a considerable savings in computational time as compared to a perturbed observation approach where the analysis must be executed by all of the ensemble members. Given a 3DVAR control run analysis and its corresponding analysis error covariance estimate, the system calculates the ET analysis perturbations and adds the perturbations to the control run to form new initial conditions for each ensemble member. The forecast model is then integrated creating a new set of ensemble forecasts for the next cycle of the ET. The NCODA ET and 3DVAR have been successfully implemented in a coupled ocean atmosphere mesoscale ensemble prediction system (Holt et al. 2011).

13.5.4 Residual Vector

The residual vector $[\mathbf{y} - \mathbf{H(x_a)}]$ is very useful in assessing the fit of the analysis to specific observations or observing systems. It is usually calculated at the end of the analysis after the post-multiplication step by horizontally and vertically interpolating the analysis vector $(\mathbf{x_a})$ to the observation locations and application of the nonlinear forward operators $\mathbf{H}$ to obtain $\mathbf{H(x_a)}$ in observation space. This result is then subtracted from the observations to form the residual vector. The problem here is that horizontal and vertical interpolations of the analysis grid to the observation locations and subsequent application of the $\mathbf{H}$ operator introduces error into the residual vector, which may change interpretation of the quality of the fit of the analysis to an observing system. A better approach is to estimate the analysis result, and the residual vector, while still in observation space, that is, before application of the post-multiplication (13.3). Daley and Barker (2000) show that a good approximation of the true residuals while in observation space can be obtained from $\mathbf{y_a} = \mathbf{y} - \mathbf{Rz}$, where $\mathbf{y}$ is observation vector, $\mathbf{y_a}$ the residual vector, $\mathbf{R}$ is the observation error covariance matrix, and $\mathbf{z}$ is defined in (13.2). Using

this formulation to calculate residuals gives a better indication of the performance of the 3DVAR assimilation algorithm and how best to tune the background and observation error statistics to improve the analysis. The NCODA 3DVAR system routinely computes residual vectors while still in observation space and saves the residual and innovation vectors for each update cycle in a diagnostics file. As noted, a time history of the innovations and the residuals is the basic information needed to compute *a posteriori* refinements to the 3DVAR statistical parameters. Analysis of the innovations is the most common, and the most accurate, technique for estimating observation and forecast error covariances and the method has been successfully applied in practice (e.g. Hollingsworth and Lonnberg 1986). Similarly, a spatial autocorrelation analysis of the residuals is used to determine if the analysis has extracted all of the information in the observing system. Any spatial correlation remaining in the residuals at spatial lags greater than zero represents information that has not been extracted by the analysis and indicates an inefficient analysis (Hollingsworth and Lonnberg 1989).

13.5.5 Internal Data Checks

Internal data checks are those quality control procedures performed by the analysis system itself. These data consistency checks are best done within the assimilation algorithm, since it requires detailed knowledge of the background and observation error covariances, which are available only when the assimilation is being performed. The first step is to scale the innovations $(\mathbf{y} - \mathbf{H}(\mathbf{x_b}))$ by the diagonal of $(\mathbf{HP_bH^T} + \mathbf{R})^{1/2}$, the symmetric positive-definite covariance matrix of (13.1). The elements of this scaled innovation vector $(\mathbf{d}\hat{})$ should have a standard deviation equal to 1 if the background and observation error covariances have been specified correctly. Assuming this to be the case, set a tolerance limit (T_L) to detect and reject any observation that exceeds it. Since $\mathbf{P_b}$ and $\mathbf{R}$ are never perfectly known, it is best to use a relatively high tolerance limit $(T_L = 4.0)$ to identify marginally acceptable observations.

The second part of the internal data check is a consistency check. It compares the marginally acceptable observations with all of the observations. The procedure is a logical extension of the tolerance limit check described above. In the data consistency test, the innovations are scaled by the full covariance matrix (not just the diagonal). The elements of this scaled innovation vector $(\mathbf{d^*})$ are also dimensionless quantities normally distributed. However, because the scaling in $\mathbf{d^*}$ involves the full covariance matrix, it includes correlations between all of the observations. By comparing the vectors $\mathbf{d}\hat{}$ and $\mathbf{d^*}$ it can be shown (Daley and Barker 2000) which marginally acceptable observations are inconsistent with other observations and can therefore be rejected. The $\mathbf{d^*}$ metric should increase (decrease) with respect to $\mathbf{d}\hat{}$ when that observation is inconsistent (consistent) with other observations, as specified by the background and observation error statistics.

13.6   Global HYCOM

As mentioned in the introduction, the NCODA 3DVAR analysis is currently cycling with global HYCOM in real-time at NAVOCEANO. The 3DVAR is expected to replace the MVOI as the data assimilation component in the operational HYCOM, which is referred to as the Global Ocean Forecast System (GOFS) version 3.

As configured within GOFS v3, HYCOM has a horizontal equatorial resolution of .08° or ∼1/12°(∼7 km mid latitude) resolution. This makes HYCOM eddy resolving. Eddy-resolving models can more accurately simulate western boundary currents and the associated mesoscale variability and they better maintain more accurate and sharper ocean fronts. In particular, an eddy resolving ocean model allows upper ocean topographic coupling via flow instabilities, while an eddy-permitting model does not because fine resolution of the flow instabilities is required to obtain sufficient coupling (Hurlburt et al. 2008b). The coupling occurs when flow instabilities drive abyssal currents that in turn steer the pathways of upper ocean currents (Hurlburt et al. 1996 in the Kuroshio; Hogan and Hurlburt 2000 in the Japan/East Sea; and Hurlburt and Hogan 2008 in the Gulf Stream). In ocean prediction this coupling is important for ocean model dynamical interpolation skill in data assimilation/nowcasting and in ocean forecasting, which is feasible on time scales up to about a month (Hurlburt et al. 2008a).

The global HYCOM grid is on a Mercator projection from 78.64°S to 47°N and north of this it employs an Arctic dipole patch where the poles are shifted over land to avoid a singularity at the North Pole. This gives a mid-latitude (polar) horizontal resolution of approximately 7 km (3.5 km). This version employs 32 hybrid vertical coordinate surfaces with potential density referenced to 2,000 m and it includes the effects of thermobaricity (Chassignet et al. 2003). Vertical coordinates can be isopycnals (density tracking), often best in the deep stratified ocean, levels of equal pressure (nearly fixed depths), best used in the mixed layer and unstratified ocean, and sigma-levels (terrain-following), often the best choice in shallow water. HYCOM combines all three approaches by choosing the optimal distribution at every time step. The model makes a dynamically smooth transition between coordinate types by using the layered continuity equation. The hybrid coordinate extends the geographic range of applicability of traditional isopycnic coordinate circulation models toward shallow coastal seas and unstratified parts of the world ocean. It maintains the significant advantages of an isopycnal model in stratified regions while allowing more vertical resolution near the surface and in shallow coastal areas, hence providing a better representation of the upper ocean physics. HYCOM is configured with options for a variety of mixed layer sub-models (Halliwell 2004) and this version uses the K-Profile Parameterization (KPP) of Large et al. (1994). A more complete description of HYCOM physics can be found in Bleck (2002). The ocean model uses 3-hourly Navy Operational Global Atmospheric Prediction System (NOGAPS) forcing from FNMOC that includes: air temperature at 2 m, surface specific humidity, net surface short-wave and long-wave radiation, total (large scale plus convective) precipitation, ground/sea temperature,

zonal and meridional wind velocities at 10 m, mean sea level pressure and dew-point temperature at 2 m. The first six fields are input directly into the ocean model or used in calculating components of the heat and buoyancy fluxes while the last four are used to compute surface wind stress with temperature and humidity based stability dependence. Currently the system uses the 0.5° degree resolution application grid NOGAPS products (i.e. already interpolated by FNMOC to a constant 0.5° latitude/longitude grid); however HYCOM can also (and preferably) use the NOGAPS T319 computational grid (i.e. a Gaussian grid—constant in longitude, nearly constant in latitude) products. Typically atmospheric forcing forecast fields extend out to 120 h (i.e. the length of the HYCOM/NCODA forecast). On those instances when atmospheric forecasts are shorter than 120 h, an extension is created based on climatological products. The last available NOGAPS forecast field is then gradually blended toward climatology to provide forcing over the entire forecast period. The current version of the global HYCOM forecast system includes a built-in energy loan, thermodynamic ice model. In this non-rheological system, ice grows or melts as a function of SST and heat fluxes. For an extensive validation of the global forecast system see Metzger et al. (2008, 2010a,b).

The NCODA 3DVAR analysis system consists of three separate programs that are executed in sequence. The first program does the analysis and data preparation, including computation of the innovation vector. The second program performs the 3DVAR, where it reads the innovation vector and outputs the analysis increment correction fields. The third program performs several post-processing tasks, such as updating the background error fields and computing some diagnostic and verification statistics. The global HYCOM 3DVAR analysis is split into seven overlapping regions covering the global ocean (Fig. 13.9). The Atlantic, Indian and Pacific Ocean regions cover the Mercator part of the model grid. The remaining four regions cover the irregular part of the model domain, one region in the Antarctic, one each in the northern part of the Atlantic and Pacific and the last region covering the Arctic Ocean. The boundary between the different regions follows the natural boundary of the continents. The regions overlap to ensure that the analyses will be smooth across the boundaries that fall over the ocean. At present the forecast system is running on 624 Cray XT5 processors. The processors are split among the sub-regions so that each regional analysis can run in parallel and finish at about the same time. Note that performing the 3DVAR in sub-regions is a holdover from the old MVOI system. There are no limitations in the 3DVAR that prevent the analysis from being executed on the full global HYCOM grid. However, at the present time, memory limitations in the data prep program do not allow the system to be executed globally. This problem is being addressed.

Two assimilative runs of the 3DVAR cycling with global HYCOM on a daily basis (24-h update cycle) are reported here. Both runs were initialized from a non-assimilative spin-up of the model. The run initialized on 1 May 2010 was executed in hindcast mode and has the advantage of assimilating synoptic ocean observations. The run initialized on 29 November 2011 is a real-time run and must deal with data latency issues associated with some of the ocean observing systems. Satellite altimeter and profile observations have the longest time delays before the data are
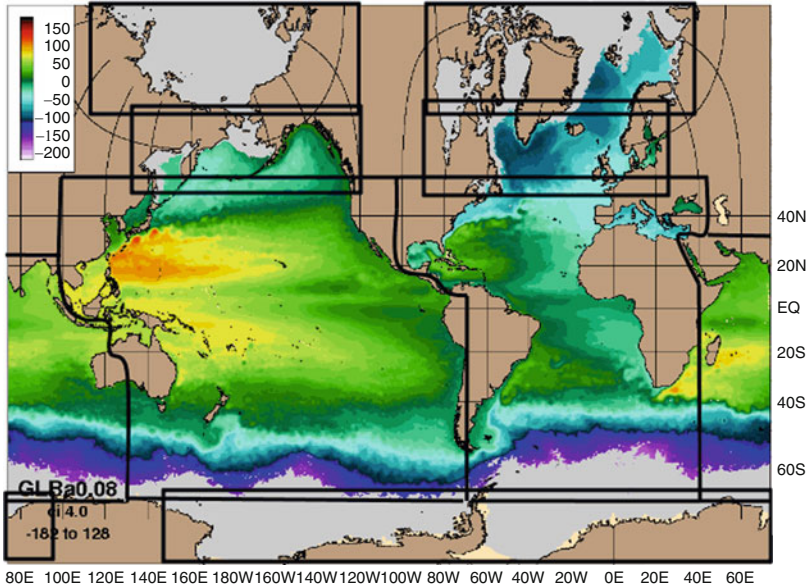
Fig. 13.9 NCODA 3DVAR analysis regions for global HYCOM. The three regions in the Atlantic, Indian and Pacific Ocean cover the Mercator projection part of the global model grid. The three regions in the Arctic Cap cover the irregular bi-polar part of the global grid: northern part of the Atlantic, northern part of the Pacific, and a region covering the Arctic Ocean. A spherical grid projection is used in the vicinity of Antarctica

available for assimilation in real-time. The delays in the altimeter data are at least 7296 h due to orbit corrections that have to be applied to improve the accuracy of the measurements. Profile data can be delayed up to ∼72 h. Since ocean data are so sparse it is important to use all of the data in the assimilation. Accordingly, in real-time applications the 3DVAR has the capability to select data for the assimilation based on receipt time (the time the observation is received at the center) instead of observation time. In this way all data *received* since the previous analysis are used in the next real-time run of the 3DVAR. However, data selected this way will necessarily contain non-synoptic measurement times. This source of error in the analysis is reduced by comparing observations against time dependent background fields using FGAT. Hourly forecast fields are used in the FGAT for assimilation of SST observations in order to maintain a diurnal cycle in the model. Daily averaged forecast fields are used in FGAT for profile data types (both synthetic and real). SSH data are assimilated in global HYCOM using the MODAS synthetic profile approach. The 3D temperature, salinity, and u, v velocity analysis increments are incrementally inserted into the model over a 6 h time period using the incremental analysis update procedure (Bloom et al. 1996). A separate 2D ice concentration analysis is used to update the ice concentration in the thermodynamic ice model.

Figures 13.10, 13.11, and 13.12 give time series of innovation and residual error statistics in the Pacific domain of the hindcast run. The statistics are computed in
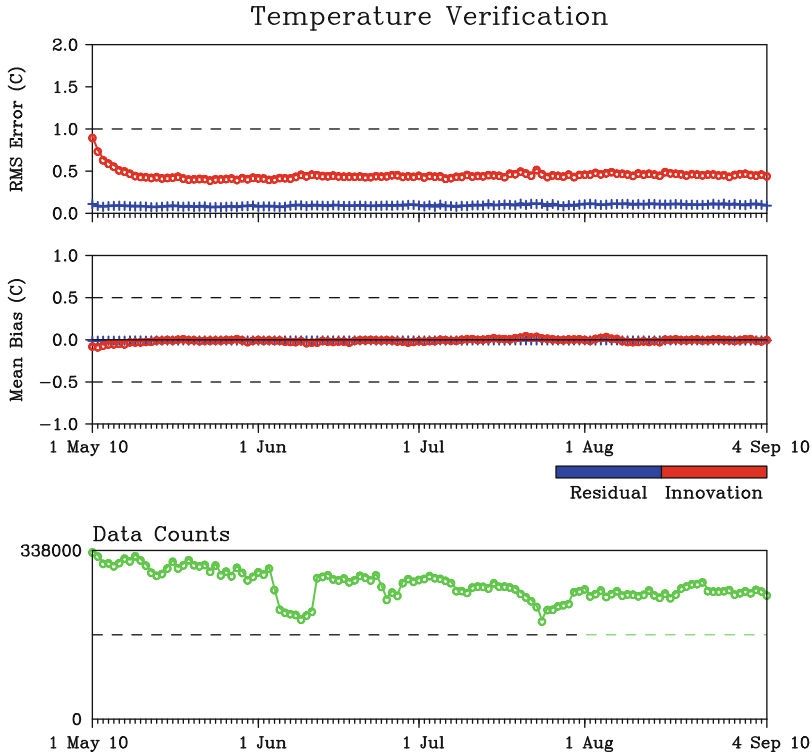
Fig. 13.10 Time of RMS and mean bias error statistics for temperature observations in HYCOM Pacific basin. *Upper panel* reports RMSE, middle panel reports mean bias, and bottom gives temperature data counts. Tick marks along time axis indicate 24-h update cycle periods

observation space and represent averages across all data assimilated for a particular analysis variable. Innovation RMS errors for temperature (Fig. 13.10) and salinity (Fig. 13.11) show increased errors for the first few update cycles while the free running model adjusts to the data. After this initial adjustment time, RMS errors are very stable, with temperature errors ∼0.4°C and salinity errors ∼0.1 PSU. The model innovations are remarkably unbiased in both temperature and salinity. The 3DVAR analysis produces a reduction in error from the innovations to the residuals of about 60 %, which is clearly seen in both temperature and salinity. However, the time series of the layer pressure error statistics (Fig. 13.12) are the most interesting. When cycling with HYCOM, the 3DVAR includes a sixth analysis variable, layer pressure. Layer pressure innovations are computed as differences in the depths of density layers in the observations and the model forecast. The layer pressure correction fields are then used to correct isopycnal layer depths in the model. Unlike the fairly rapid response of the free-running model to the assimilation of temperature and salinity observations, bias in the layer structure of the model spin-up takes about a month to adjust to the data. Layer pressure RMS errors remain
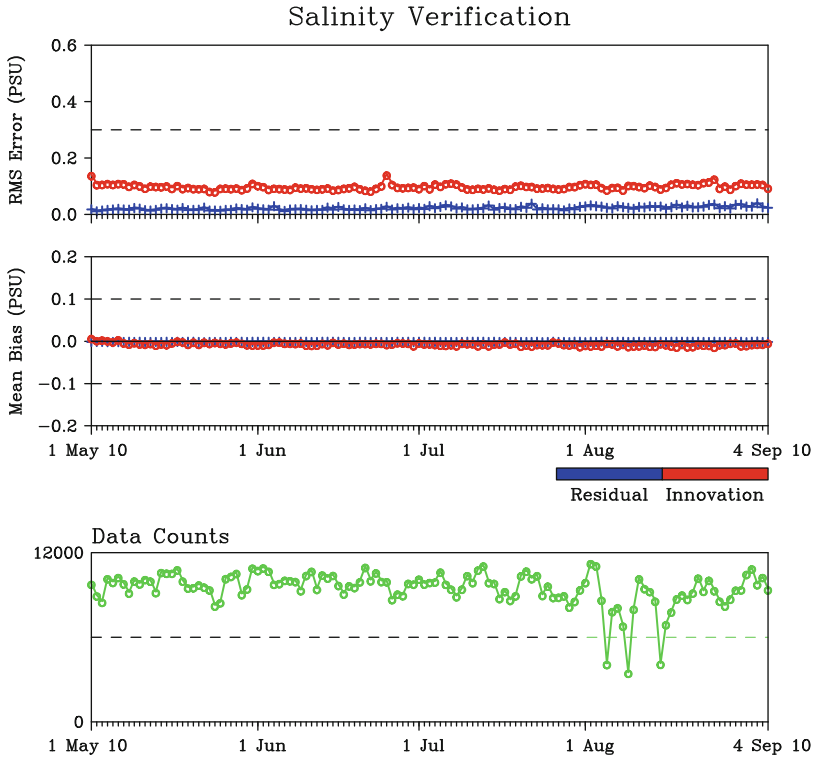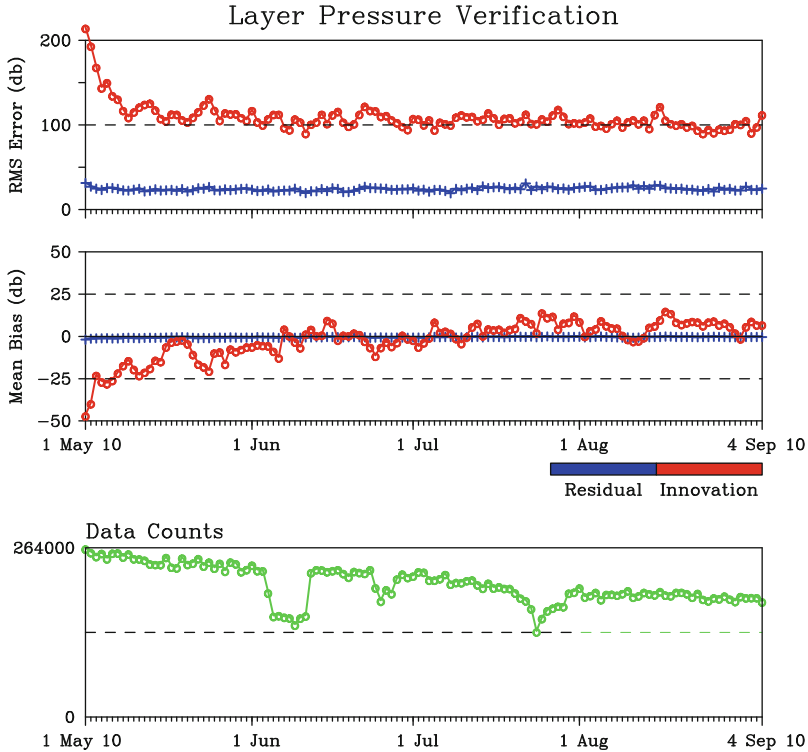
Fig. 13.11 Same as Fig. 13.10, except for salinity observations

high (∼100 db) after the adjustment time period due to the assimilation of MODAS synthetic profiles at high latitudes. MODAS synthetics were not thinned based on stratification (Sect. 13.4.3) in these model runs. Layer pressure RMS errors are reduced more than 50 % when weakly stratified MODAS synthetics are rejected (not shown).

Figure 13.13 shows a verification result from the real-time run for sea surface height in the Kuroshio region on 12 January 2012. The assimilation of SSH anomalies is crucial to accurately map the circulation in these highly chaotic regions dominated by flow instabilities. The white (black) line overlain is an independent analysis of available infrared observations of the north edge of the current system performed at the Naval Oceanographic Office. The frontal analysis clearly indicates that the forecast system is able to accurately map the mesoscale features in the western boundary current.

Table 13.2 gives run times for the 3DVAR conjugant gradient solver and post-multiplication steps. The run times are listed for a typical day (28 January 2012) in six of the global HYCOM analysis subdomains. A total of 2.2 million observations were assimilated into the HYCOM grid that contained more than 520 million grid

Fig. 13.12 Same as Fig. 13.10, except for layer pressure observations. Layer pressure is computed from density using temperature and salinity profiles (see text for details)

points. The total time of the 3DVAR step in the NCODA analysis system is the maximum time needed to complete any of the subdomains—in this case 14.2 min to complete the Indian Ocean analysis. Efficiency of the 3DVAR is clear, especially in the large Pacific basin where >1 million observations were assimilation into 195.2 million grid points in ∼9.8 min wall clock time. Table 13.2 also shows how well the analysis scales using different numbers of processors. Reduction of the Indian Ocean run time, and thus speed-up of the 3DVAR analysis step in global HYCOM analysis/forecast system, can easily be achieved by simply increasing the number of processors. In general, the post-multiplication step of the analysis is more computationally expensive than the observation space solver. Accordingly, the analysis contains an option to perform the post-multiplication step on a reduced resolution grid. The innovations are always formed from the full resolution model grid, and the solution vector is calculated using all of the observations, but now the solution is mapped to every other (or any multiple) horizontal grid point. This option results in a considerable saving in computational time with no loss of information when analysis correlation length scales generally exceed the model grid resolution.
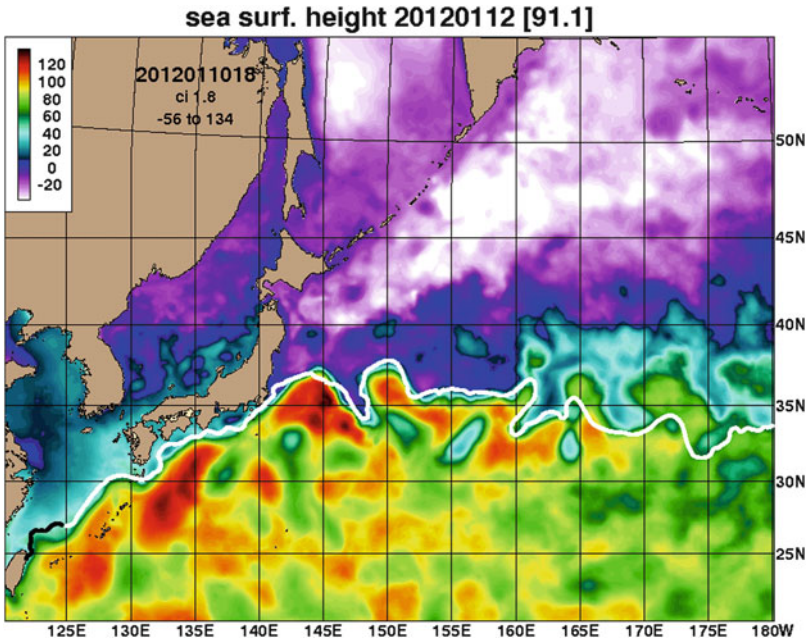
Fig. 13.13 Sea surface height in the Kuroshio region from the $1/12°$ global HYCOM/NCODA forecast system on January 12, 2012. An independent infrared (IR) analysis of the north edge of the current system performed by the Naval Oceanographic Office is overlain. A *white* (*black*) line means the IR analysis is based on data less (more) than four days old

Table 13.2 3DVAR run times for six of the seven global HYCOM analysis domains on 28 January 2012

Domain	Grid size	Number procs	Number obs	Solver (min)	Post proc (min)	Total (min)
Atlantic	$1{,}751 \times 1{,}841 \times 42$	88	613,525	4.8	5.6	10.7
Indian	$1{,}313 \times 1{,}569 \times 42$	64	468,828	6.6	7.3	14.2
Pacific	$2{,}525 \times 1{,}841 \times 42$	416	1,028,369	6.7	2.6	9.8
Arctic Ocean	$1{,}630 \times 551 \times 42$	16	11,879	0.1	0.2	1.7
Arctic Atlantic	$1{,}490 \times 551 \times 42$	16	82,137	0.1	0.6	2.3
Arctic Pacific	$1{,}335 \times 551 \times 42$	16	17,630	0.4	0.2	1.6
Totals	520,250,556[a]	616	2,222,368			

[a]Total for grid size is the total number of grid points

Full resolution correction fields for the model update are produced for each analysis variable in the NCODA 3DVAR post-processing step by interpolation. This reduced resolution grid option is used in global HYCOM where the solution vector is mapped to every other model grid point.

13.7   Future Capabilities

The NCODA 3DVAR and Navy global ocean forecasting systems continue to be developed and improved. These new developments and capabilities are summarized in this section.

13.7.1   HYCOM GOFS

The present $1/12°$ global HYCOM/NCODA system is the first step towards a $1/25°$ global forecast system. The first phase of the upgrade will continue to use the $1/12°$ model. In this phase the simple thermodynamic ice model will be replaced by the Los Alamos Community Ice CodE (CICE). CICE is the result of an effort to develop a computationally efficient sea ice component for a fully coupled forecast system. CICE has several interacting components: a thermodynamic model that computes local growth rates of snow and ice due to vertical conductive, radiative and turbulent fluxes, along with snowfall; a model of ice dynamics, which predicts the velocity field of the ice pack based on a model of the material strength of the ice; a transport model that describes advection of the areal concentration, ice volumes and other state variables; and a ridging parameterization that transfer ice among thickness categories based on energetic balances and rates of strains. HYCOM and CICE will be fully coupled via the Earth System Modeling Framework (ESMF: Hill et al. 2004). An interim, fully coupled, real time Arctic Cap HYCOM/CICE/NCODA-3DVAR forecast system has been set up until CICE is implemented in the global model (Posey et al. 2010). The second phase of the upgrade includes the implementation of a fully coupled $1/25°$ HYCOM/CICE model that includes tidal forcing and uses NCODA 3DVAR as the data assimilation component for both HYCOM and CICE. Preliminary experiments with the assimilative $1/25°$ model are under way. This model will have $\sim$3 km mid latitude resolution.

13.7.2   Satellite SST Radiance Assimilation

At the present time, SST retrievals are empirically derived using stored regressions between cloud cleared satellite SST radiances and drifting buoy SSTs. The regressions are global, calculated once, and held constant. The coefficients represent a very broad range of atmospheric conditions with the result that subtle systematic errors are introduced into the empirical SST when the method is uniformly applied to new radiance data. In the 3DVAR, work is underway to develop an observation operator for direct assimilation of satellite SST radiances. This new physical SST algorithm uses an incremental approach. It takes as input prior estimates of SST and short-term predictions of air temperature and water vapor profiles from NWP. The algorithm is

forced by differences between observed and predicted top-of-the-atmosphere (TOA) brightness temperatures (BTs) for the different satellite SST channel wavelengths. Calculation of the TOA-BTs requires use of a fast radiative transfer model. For this purpose the Community Radiative Transfer Model (CRTM; Han et al. 2006) is being integrated into the 3DVAR. In addition to the TOA forward model, CRTM provides the tangent linear radiance sensitivities (Jacobians) with respect to the prior SST, water vapor, and atmospheric temperature predictor variables as a function of the infrared satellite 3.5, 11 and 12 μm wavelengths. The physical SST inverse model for a given channel is given by,

$$
\begin{bmatrix} \delta BT \cdot J_{sst} \\ \delta BT \cdot J_t \\ \delta BT \cdot J_q \end{bmatrix} = \begin{bmatrix} \varepsilon_{sst}^{-1} \cdot J_{sst} \cdot J_{sst} & J_{sst} \cdot J_t & J_{sst} \cdot J_q \\ J_t \cdot J_{sst} & \varepsilon_t^{-1} \cdot J_t \cdot J_t & J_t \cdot J_q \\ J_q \cdot J_{sst} & J_q \cdot J_t & \varepsilon_q^{-1} \cdot J_q \cdot J_q \end{bmatrix} \begin{bmatrix} \delta T_{sst} \\ \delta T_a \\ \delta Q_a \end{bmatrix} \quad (13.11)
$$

where δBT are the TOA-BT innovations, J_{sst}, J_t, and J_q are the radiative transfer model Jacobians for SST, atmospheric temperature, and water vapor, respectively, ε_{sst}, ε_t, and ε_q are the errors of the priors, and δT_{sst}, δT_{atm}, and δQ_{atm} are the corrections output for each of the priors that take into account the variable SST and temperature and water vapor content of the atmosphere at the time and location of the radiance measurement. The prior corrections are calculated and summed over the SST channels (3 channels at night, 2 channels during the day). With this approach, coefficients that relate radiances to SST in the observation operator are dynamically defined for each atmospheric situation observed. The method removes atmospheric signals in the radiance data and extracts more information on the SST, which improves the time consistency of the SST estimate, especially in the tropics where water vapor variations create unrealistic sub-daily variations in the empirically derived SST. However, the physical SST method requires careful consideration of biases and error statistics of the NWP fields. Biases are expected since the NWP information may represent areas that are both cloudy and clear, while the satellite radiance data, by definition, are only available in clear-sky, cloud free conditions. Accordingly, a bias correction step is under development following the ideas developed by Merchant et al. (2008). Proper specification of the error statistics of the priors is also required to correctly partition the observed TOA-BT differences into the various sources of variability (atmospheric temperature, water vapor, or sea surface temperature). Sensitivity experiments are underway to evaluate situation dependent error statistics for the atmospheric temperature and water vapor priors using the 96-member global NWP ensemble operational at FNMOC.

Implementation of the physical SST method via an observation operator will have many advantages in the 3DVAR. First, in a coupled model forecast, the prior SST will come from the coupled ocean model forecast and differences between observed and predicted TOA-BTs will be computed using the coupled model atmospheric state. This is a true example of coupled data assimilation: an observation in one fluid (atmospheric radiances) creates an innovation in a different fluid (ocean SST). Second, the method can easily be extended to incorporate the

effects of aerosols; the presence of which tends to introduce a cold bias in infrared estimates of SST. To do this prior information on the microphysical properties of dust and its amount and vertical distribution is obtained from the Navy Aerosol Analysis Prediction System (NAAPS; http://www.nrlmry.navy.mil/aerosol/). The contribution of NAAPS aerosol information to the TOA-BTs is determined using CRTM, which contains aerosol Jacobians defined for 91 wavelengths and 6 aerosol species. Equation (13.11) is then expanded to a 4×4 matrix to further partition differences between observed and simulated TOA BTs into an additional aerosol source of variability. Third, the method can be applied to radiances from ice covered seas to determine ice surface temperature (IST). Knowledge of IST is important since it controls snow metamorphosis and melt, the rate of sea ice growth, and modification of air–sea heat exchange. IST has been added as an analysis variable in the 3DVAR and is analyzed simultaneously with SST to form a seamless depiction of surface temperature from the open ocean to ice covered seas. This capability will be used in the coupled HYCOM/CICE system (Posey et al. 2010).

13.7.3 SSH Velocity Assimilation

An alternative to assimilating SSH information referenced to the along-track mean is to assimilate the dynamically important along-track SSH slope. Altimeter SSH slopes provide the cross-track component of the vertically averaged geostrophic current. As noted in Sect. 13.4.3, current methods for assimilating altimeter SSH data via synthetic temperature and salinity profiles have known deficiencies. One major difficulty is the need to specify a reference MDT matching that contained in the altimeter data; a non-trivial problem. The mean height of the ocean includes the Geoid (a fixed gravity equipotential surface) as well as the MDT, which is not known accurately enough relative to the centimeter scales of variability contained in the dynamic topography. The use of SSH slopes obviates the need for a MDT.

To derive geostrophic currents from SSH slopes appropriate for the ocean mesoscale, noise in the along-track altimeter data must be suppressed. For this purpose a quadratic LOESS smoother (LOcally wEighted Scatterplot Smoother: Cleveland and Devlin 1988; Schlax and Chelton 1992) with varying cutoff wave lengths is applied. The wave lengths are adjusted in accordance with the Rossby radius of deformation to account for the varying eddy length scales. The advantage of this method is that noise in the data, the SSH slope derivative, and the u, v vector velocity components are all computed in a single operation. Figure 13.14 shows the LOESS smoothing of the altimeter SSH data along two tracks; track 109 across the Gulf Stream (Fig 13.14a) and track 106 across the Kuroshio (Fig. 13.14b). The quality of the LOESS filter is clearly seen when the altimeter data exhibit considerable noise (distance points 1,000–3,000, track 109; distance points 1,200–2,440, track 106), and when the altimeter data show strong signals from crossing the Gulf Stream and Kuroshio fronts (distance points 3000–3800, track 109; distance points 400–1,000, track 106). Figure 13.15 shows the Atlantic and Pacific basin
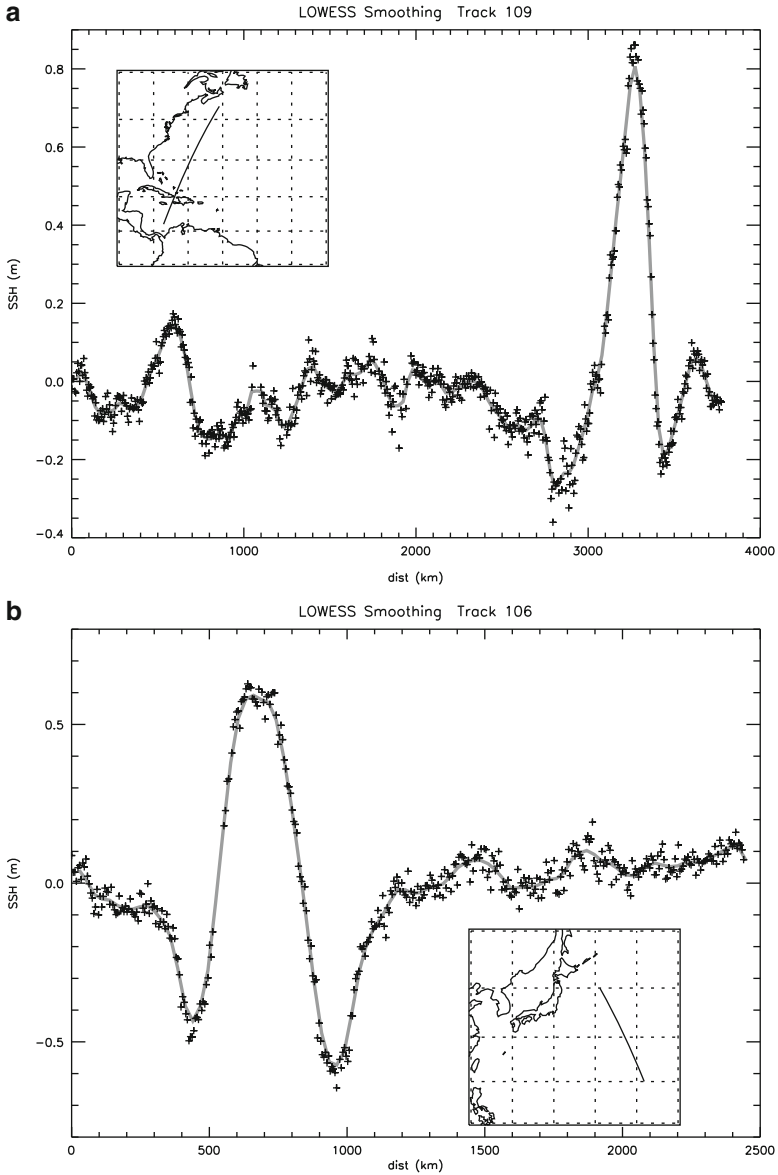
Fig. 13.14 Smoothed along-track SSH computed using LOESS filter. All data from 10 January 2012. (**a**) LOESS filter fit to altimeter SSH data along track 109 in the Gulf Stream area; (**b**) LOESS filter fit to altimeter SSH data along track 106 in the Kuroshio area. Plus marks give raw altimeter SSH data values, solid line gives LOESS fit

Fig. 13.15 Basin scale geostrophic velocity data calculated from smoothed along-track altimeter SSH data using LOESS filter. All data from 10 January 2012. Top part of each Figure gives basin scale results, lower left gives LOESS filter results, lower right gives zoom on geostrophic velocity along tracks intersecting the Kuroshio and Gulf Stream fronts. (**a**) HYCOM Pacific basin; (**b**) HYCOM Atlantic basin
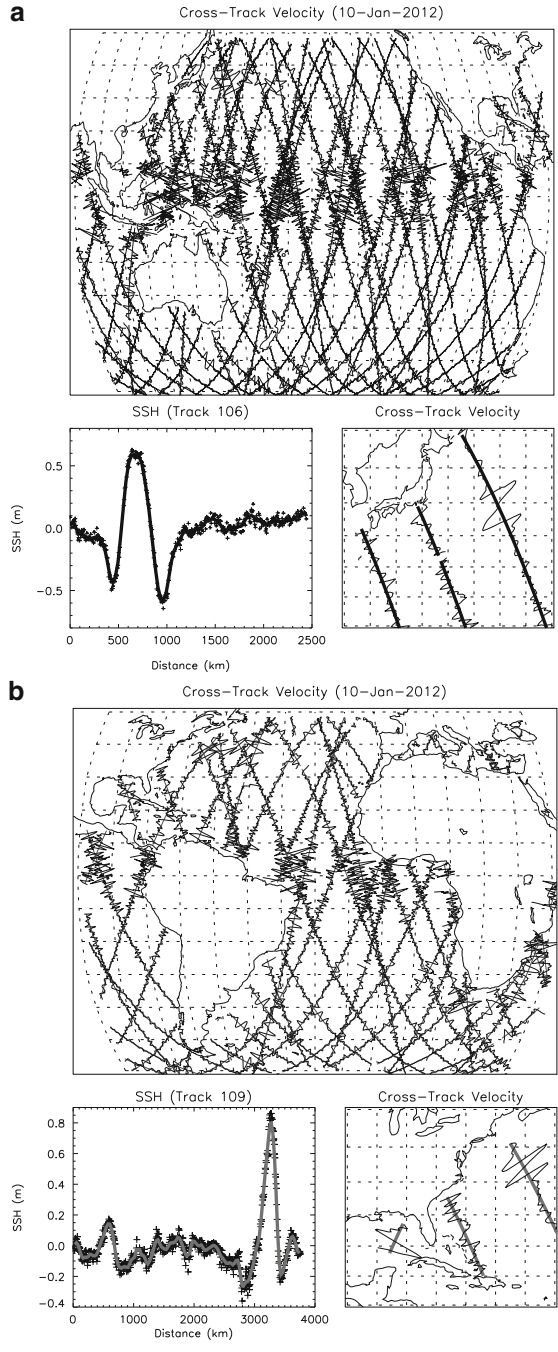
cross-track geostrophic velocities computed using the LOESS filter for one day of along-track altimeter data (10 January 2012). It is readily apparent that a tremendous amount of mesoscale oceanographic information is contained in the geostrophic velocities derived from the along-track altimeter data.

Once the altimeter SSH along-track geostrophic currents are calculated the model equivalents are determined. Cross-track geostrophic velocity relative to a deep level of no motion (2,000 m) is computed from the model using dynamic height differences at points adjacent to the along-track estimate of the SSH slope. The difference between the vertically averaged model and altimeter cross-track geostrophic velocities is used to correct the relative geostrophic shear from the model and form the velocity profile $u_a(z)$ for the assimilation according to:

$$u_a(z) = u_g(z) - \overline{u_g} + c \qquad (13.12)$$

where $u_g(z)$ is the model relative geostrophic shear profile, $\bar{u}_g$ is its vertical average, and c is the integral cross track velocity component calculated from the altimeter slope. Assimilation of the u, v velocity vectors formed this way via the multivariate correlations in the 3DVAR provide balanced geopotential increments, which in turn are decomposed into balanced temperature and salinity increments using a linearized equation of state. The velocity profiles in this scheme are very sensitive to the reference level of no motion. One option here is to use Argo trajectory data to infer a time dependent geopotential field at the float parking depth (cf. Davis 2005). A dynamic geopotential field would go a long way in solving a long-standing problem of hydrography: properly referencing geostrophic shear.

13.7.4 Hybrid Ensemble Four Dimensional Data Assimilation

A four-dimensional (4D) ensemble-enhanced data assimilation scheme for global HYCOM is being developed to better deal with the late receipt, temporally distributed observations than the current 3DVAR methodology. As previously noted, a crucial aspect of all ocean data assimilation schemes is the way in which the background error covariances are specified. The data assimilation process is optimal if the background error covariances are perfectly known, which is never the case. A major challenge then is to find ways to estimate accurate and comprehensive background error covariances. Ensemble methods provide a method for doing this, including the ability to provide a flow-dependent estimate of the background error covariances.

When ensemble covariances are used in a variational data assimilation framework to augment the existing background-error covariance, analyses are further improved. This method is called a hybrid ensemble variational method. In comparison with conventional ensemble-based data assimilation, a hybrid scheme is attractive for the following reasons. First, the hybrid schemes build upon existing variational systems enabling the ensemble information to be incorporated relatively

easily. Existing variational ocean data assimilation technology and capabilities are not lost. Second, when ensemble variances are imperfect the optimal error variance estimate is a linear combination of a climatological covariance and an ensemble covariance. The superiority of hybrids over conventional ensemble assimilation schemes is particularly marked when the ensemble size is small or the model error is large.

A static 4D ensemble covariance data base will be computed from an ensemble of mesoscale anomalies using the long term integration of global HYCOM in the 1993–2009 reanalysis product, which includes NCODA 3DVAR assimilation. Covariances calculated in this way have clear physical meanings and represent 4D model climate flow dependence and model variable interactions. Existing 3DVAR initial covariances will be extended to 4D by assuming that the error covariances between variables are a separable function of space and time. The computational overhead of imparting this 4D aspect to the 3DVAR covariances is expected to be very small. The 4D extension of the NCODA covariances will then be linearly combined with the 4D localized HYCOM static ensemble covariances forming a fully 4D hybrid data assimilation scheme. Optimum values for weighting the ensemble and extended 3DVAR covariances in the hybrid are determined from model statistics.

13.8   Summary

This paper describes the development, implementation, and validation of a new oceanographic 3DVAR assimilation system. The system is unified and flexible and a key component of many Navy ocean and atmosphere applications. It is run globally or regionally, where it can be applied to nested, successfully higher-resolution grids, providing analyses on a range of scales. NCODA 3DVAR provides the assimilation component for both ocean and wave model prediction systems as well as multiple atmospheric prediction systems, where it is used to provide sea ice and SST lower boundary conditions. It assimilates a wide range of ocean data types and it contains numerous diagnostic features for assessing and tuning the statistics needed for the assimilation as well as quality control. The background error covariance formulation permits considerable anisotropy with adaptive horizontal and vertical length scales and error variances that vary with location and evolve with time. It is shown to be efficient for very large scale, high resolution global ocean model grids, assimilating millions of observations a day. The intelligent, adaptive data thinning algorithm permits all sources of the high density surface data types to be assimilated with minimal loss of information. The parallel implementation has minimal communication overhead, with granularity of the code (important for load balancing) easily controlled by the number and size of the observation data blocks. The NCODA 3DVAR system is operational at the Navy oceanographic production centers and is in the final phase of pre-operational testing as the data assimilation component for the global HYCOM forecasting system.

References

Bishop CH, Toth Z (1999) Ensemble transformation and adaptive observations. J Atmos Sci 56:1748–1765

Bleck R (2002) An oceanic general circulation model framed in hybrid isopycnic- Cartesian coordinates. Ocean Model 4:55–88

Bloom SC, Takacs LL, Da Silva AM, Ledvina D (1996) Data assimilation using incremental analysis updates. Mon Weather Rev 124:1256–1271

Chassignet EP, Smith LT, Halliwell GR, Bleck R (2003) North Atlantic simulations with the HYbrid coordinate ocean model (HYCOM): impact of the vertical coordinate choice, reference pressure, and thermobaricity. J Phys Oceanogr 33(12):2504–2526

Chelton DB, DeSzoeke RA, Schlax MG, Naggar KE, Siwertz N (1998) Geographical variability of the first baroclinic Rossby radius of deformation. J Phys Oceanogr 28:433–460

Cleveland WS, Devlin SJ (1988) Locally weighted regression: an approach to regression analysis by local fitting. J Am Stat Assoc 83:596–610

Cooper M, Haines KA (1996) Altimetric assimilation with water property conservation. J Geophys Res 24:1059–1077

Cummings JA (2005) Operational multivariate ocean data assimilation. Q J R Met Soc 131: 3583–3604

Cummings JA (2011) Ocean data qaulity control. In: Schiler A, Brassington GB (eds) Operational oceanography in the 21st century. Springer, Dordrecht, pp 91–121

Courtier P (1997) Dual formulation of four-dimensional variational assimilation. Q J R Meteorol Soc 123:2449–2461

Daley R (1991) Atmospheric data analysis. Cambridge University Press, Cambridge, p 457

Daley R, Barker E (2000) The NAVDAS source book. Naval Research Laboratory NRL/PU/ 7530-00-418, Monterey, 151pp

Daley R, Barker E (2001) NAVDAS formulation and diagnostics. Mon Weather Rev 129:869–883

Davis R (2005) Intermediate-depth circulation of the Indian and South Pacific Oceans measured by autonomous floats. J Phys Oceanog 35:683–707

Fofonoff NP, Millard RC (1983) Algorithms for computation of fundamental properties of seawater. Tech Pap Mar Sci UNESCO 44:53

Fox DN, Teague WJ, Barron CN, Carnes MR, Lee CM (2002) The modular ocean data assimilation system. J Atmos Ocean Technol 19:240–252

Halliwell GR (2004) Evaluation of vertical coordinate and vertical mixing algorithms in the HYbrid Coordinate Ocean Model (HYCOM). Ocean Model 7(3–4):285–322

Han Y, van Delst P, Liu Q, Weng F, Yan B, Treadon R, Derber J (2006) JCSDA Community Radiative Transfer Model (CRTM)—version 1, NOAATechnical Report. NESDIS 122:40

Helber RW, Carnes MR, Townsend TL, Barron CN, Dastugue JM (2012) Validation test 1 report for the Improved Synthetic Ocean Profile (ISOP) system, Part I: Stand-alone capability (In preparation)

Hill C, DeLuca C, Balaji V, Suarez M, da Silva A (2004) The architecture of the earth system modeling framework. Comp Sci Eng 6:18–28

Hogan PJ, Hurlburt HE (2000) Impact of upper ocean—topographic coupling and isopycnal outcropping in Japan/East Sea models with $1/8°$ to $1/64°$ resolution. J Phys Oceanogr 30:2535–2561

Hollingsworth A, Lonnberg P (1986) The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. Tellus 38A:111–136

Hollingsworth A, Lonnberg P (1989) The verification of objective analyses: diagnostics of analysis system performance. Meteor Atmos Phys 40:3–27

Holt TR, Cummings JA, Bishop CH, Doyle JD, Hong X, Chen S, Jin Y (2011) Development and testing of a coupled ocean–atmosphere mesoscale ensemble prediction system. Ocean Dynam 61:1937–1954

Hurlburt HE, Hogan PJ (2008) The Gulf Stream pathway and the impacts of the eddy-driven abyssal circulation and the deep western boundary current. Dynam Atmos Oceans 45:71–101

Hurlburt HE, Wallcraft AJ, Schmitz WJ Jr., Hogan PJ, Metzger EJ (1996) Dynamics of the Kuroshio/Oyashio current system using eddy-resolving models of the North Pacific Ocean. J Geophys Res 101(C1):941–976

Hurlburt HE, Chassignet EP, Cummings JA, Kara AB, Metzger EJ, Shriver JF, Smedstad OM, Wallcraft AJ, Barron CN (2008a) Eddy-resolving global ocean prediction. In: Hecht M, Hasumi H (eds) Ocean modeling in an eddying regime. Geophysical monograph, vol 177. American Geophysical Union, Washington, DC, pp 353–381

Hurlburt HE, Metzger EJ, Hogan PJ, Tilburg CE, Shriver JF (2008b) Steering of upper ocean currents and fronts by the topographically constrained abyssal circulation. Dynam Atmos Oceans 45:102–134. doi:10.1016/j.dynatmoce.2008.06.003

Large WG, Mc Williams JC, Doney SC (1994) Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. Rev Geophys 32:363–403

Karra B, Rochford PA, Hurlbut H (2000) An optinal definition for mixed layer depth. J Geophys Res 105:16803–16821

Martin MJ, Hines A, Bell MJ (2007) Data assimilation in the FOAM operational short-range ocean forecasting system: a description of the scheme and its impact. Q J R Meteorol Soc 133:981–995

McLay J, Bishop CH, Reynolds CA (2008) Evaluation of the ensemble transform analysis perturbation scheme at NRL. Mon Weather Rev 136:1093–1108

Merchant CJ, Le Borgne P, Marsouin A, Roquet H (2008) Optimal estimation of sea surface temperature from split-window observations. Rem Sens Environ 112:2469–2484

Metzger EJ, Smedstad OM, Thoppil PG, Hurlburt HE, Wallcraft AJ, Franklin DS, Shriver JF, Smedstad LF (2008) Validation Test Report for the Global Ocean Prediction System V3.0— $1/12°$ HYCOM/NCODA: Phase I. NRL Memo. Report NRL/MR/7320–08- 9148

Metzger EJ, Smedstad OM, Thoppil PG, Hurlburt HE, Franklin DS, Peggion G, Shriver JF, Townsend TL, Wallcraft AJ (2010a) Validation Test Report for the Global Ocean Forecast System V3.0 − $1/12°$ HYCOM/NCODA: Phase II. NRL Memo. Report NRL/MR/7320–10-9236

Metzger EJ, Thoppil PG, Smedstad OM, Franklin DS (2010b) Global Ocean Forecast System V3.0 Validation Test Report Addendum: addition of the Diurnal Cycle. NRL Memo. Report NRL/MR/7320–10-9305

Posey PG, Metzger EJ, Wallcraft AJ, Smedstad OM, Phelps MW (2010) Validation of the $1/12°$ Arctic Cap Nowcast/Forecast System (ACNFS). NRL Memo. Report NRL/MR/7320–10-9287

Riishøjgaard LP (1998) A direct way of specifying flow-dependent background error correlations for meteorological analysis systems. Tellus 50A:42–57

Schlax MG, Chelton DB (1992) Frequency domain diagnostics for linear smoothers. J Am Stat Assoc 87:1070–1081

Wittmann P, Cummings J (2005) Assimilation of altimeter wave measurements into WAVE-WATCH III. In: Proceedings of the 8th international workshop on wave hindcasting and forecasting, North Shore Oahu, Hawaii 16 pp