DATA ASSIMILATION IN OCEAN MODELS

D L T Anderson

ECMWF

1. INTRODUCTION

Various techniques for data assimilation are used in oceanography as in meteorology, such as nudging, successive corrections, optimal interpolation (OI), Kalman Filtering and smoothing, variational methods in three or four dimensions (3D or 4D VAR). Although ocean assimilation is a relatively new field compared with its more mature sister discipline in meteorology, progress in testing techniques has been quite fast. By contrast, operational implementation of assimilation schemes has been much slower and much less effort has been expended in defining accurately the background error covariance functions.

The objectives in oceanography are broadly similar to meteorology as there is an increasing need to forecast accurately the future state of the oceans, for everything from climate prediction through pollution forecasting to fishery management. One major difference is that, in the atmospheric case one is invariably trying to predict the onset and evolution of instabilities, such as those causing the cyclones and anticyclones which determine weather in mid-latitudes. In oceanography one may sometimes be trying to predict instabilities, as for example meanders, rings and eddies in the Gulf Stream area, but there are many processes in the ocean which are much more deterministic. Deterministic here means independent of initial conditions, i.e. if you know the past history of the forcing field and use a good ocean model, the state of the ocean can be determined rather well without any data assimilation. However, ocean models are not perfect and the forcing fields are not adequately known and so assimilation is frequently used to compensate for forcing error. Whether data assimilation should be used to correct for systematic error in model or forcing field is debatable. In my opinion, it should not, but in oceanography it very frequently is.

Data assimilation is therefore a useful technique whenever there is insufficient data to define the state of the system from data alone. Curiously, it might also be useful if one is data rich, even so rich that one can measure anything and everything needed to produce an excellent analysis of the ocean state at any time from observations alone. Although it would appear that there is then no need to extrapolate data in time via a model, and thus no need to assimilate data, in general, analysis is not an end in itself but a step on the way to something else, usually a forecast. A good forecast requires a good model, and one way to test and improve a model is through data assimilation with the approaches which include the time domain (Kalman smoothing, 4-D Var) being particularly well suited. Of course in general the ocean is not well observed and so the data assimilation properties of filtering the data and using data from previous times are extremely important.

Measuring the ocean is very difficult: it is vast, inhospitible and is not easily observed by remote techniques, other than at the surface. Until recently there were insufficient data to make analyses of the instantaneous ocean state worthwhile other than in a few locations. The appreciation of the importance of the ocean to the climate system has however given new impetus to taking observations. There have been two major experiments, TOGA and WOCE, over the last decade or so which have raised the profile of physical oceanography considerably, to the extent that in some regions of the tropics as a result of TOGA, a substantial amount of in situ data is now routinely available in near-real-time and as a result of WOCE sea surface height is measured globally by satellite and available in near real time.

Most time scales in the ocean are much slower than those of the atmosphere so there is usually no need to produce analyses on an hourly or daily time scale as is required in meteorology. For most purposes weekly timescales are adequate and for some purposes monthly or longer are all that are needed, though there are also some applications for which a daily analysis is in demand: for example the depth of the upper ocean mixed layer, which can vary on atmospheric timescales, is of interest for military purposes.

Section (2) briefly describes ways of observing the ocean. Section (3) gives a brief review of techniques used in oceanography, ranging from the rather simple successive correction through to 4D variational. Section (4) concerns tropical analyses. The tropical Pacific is special, both because it is better observed in near real time and because the analyses are used as initial conditions for climate forecasts. Section (5) is devoted to the use of altimetric data. This type of observation is given a special section since altimetry provides global high resolution coverage. However the data are quite difficult to use since only deflections of the top surface can be measured. Various schemes to utilize them are described.

A substantial body of work is not fully discussed: for example Kalman filtering or the techniques of feature modelling as applied, for example, to Gulf Stream ring vortices. See Robinson *et al* (1989), Miller (1986, 1989), Dee(1991), Cane *et al* (1996), Hurlburt *et al* (1990), Mariano (1990) and Fu *et al* (1991) and the papers of De Mey and Miller (this volume) for further information on these topics. A number of texts go more thoroughly into data assimilation than is done here. The interested reader is referred to the book edited by Anderson and Willebrand, 1989 which contains a number of articles on the theory and practice of data assimilation in ocean models and the closely related topic of inverse modelling. The review article by Ghil and Malanotte-Rizzoli (1991) is recommended, as is the special issue of the Journal of the Meteorological Society of Japan (1996). The book by Daley, (1991), gives a good exposé of techniques used in meteorology a few years ago. The more theoretical book by Bennett (1992) will be helpful for the serious reader. The book edited by Malanotte-rizzoli 1996 covers many topics in ocean data assimilation in greater depth than is done here. The review article by Anderson *et al* (1996) expands on a number of topics discussed in this article at a level suitable for new recruits to ocean data assimilation.

2. OBSERVING THE OCEAN

The ocean is an inhospitable environment, making it difficult to make measurements, particularly of the sub-surface. Measurements of surface temperature (SST) have been made routinely for a long time, mainly by merchant and Navy ships, and in recent times SST has been measured by satellite using infrared or microwave radiometers, to the point that when combined with ship measurements an analysis of SST is possible on a monthly and even weekly basis over much of the globe using an optimal interpolation technique (Reynolds *et al* 1995).

While SST is an important variable for ocean atmosphere interaction it is not much use for determining the ocean structure beneath the surface. In the past this has been obtained by CTD (Conductivity Temperature and Depth) instruments. These are lowered over the stern of a ship and can make high quality measurements of temperature and salinity all the way to the bottom of the ocean (5km). Unfortunately such instruments can only be deployed from special oceanographic ships. As there are only a few such ships, measurements can only be made in a small number of places at any given time: there is no possibility of creating a real-time global picture of the ocean from such instruments. In fact one of the objectives of the World Ocean Circulation Experiment (WOCE) is to survey a large part of the ocean for the *first* time.

Observing the upper ocean is easier. A device called an XBT (eXpendable BathyThermograph) can measure temperature in the upper few hundred meters of water, not to the accuracy of the CTD but good enough for many purposes. The XBT consists of a probe connected through a long thin wire to recording equipment on the ship. As the probe drops through the water it measures temperature which is relayed back along the wire. The probe is assumed to have a certain fall rate and so a time record of temperature can be transformed into a depth record of temperature. The advantage of an XBT is that it is a relatively small instrument which can be mounted on a merchant ship and measurements taken while the ship is underway without alteration to the ship's schedule. Someone on board has to be responsible for the loading and firing of the instrument. As a result of the Tropical Ocean Global Atmosphere experiment (TOGA), many merchant ships are instrumented with a transmitter so that the data can be communicated to a satellite and are available within a few hours of being taken. Typical resolution is about 100km along the shipping lanes. An AXBT (Airborne XBT) is a variant which can be deployed from an aircraft, but tends to be used only in limited areas.

Another way to observe the ocean is to have moorings anchored to the bottom. Technology has advanced to the stage where such moorings can survive for many months before needing recovery and refurbishment. These have been mostly developed in the tropical Pacific ocean, one of the remotest places on earth, as a result of TOGA. These moorings (called TAO) can measure the temperature of the top 500 m of the ocean and relay the data via satellite back to shore in near real time. In addition they measure the surface wind and the atmospheric humidity. Tests have been made of their ability to carry automatic rain gauges and salinity sensors.

Measuring the velocity field directly is more difficult than measuring the thermal field and so often it is derived from T using geostrophy. Direct velocity measurements can be obtained from moored current meters, but these are expensive. For example there are only 5 current meter moorings in the whole tropical Pacific. Fortunately in much of the tropical ocean the thermal field is the more important for data assimilation (Moore *et al* 1987, Anderson and Moore 1989). Nonetheless velocity data are important for understanding the movement of properties by advective processes and considerable effort has gone into designing methods to obtain velocity information. One promising approach is to use drifters. These are best developed for measuring the surface velocity. The drifter is drogued at a depth of 15m but has a surface float which can transmit its position to a satellite. By monitoring successive positions and provided the buoy is drogued so that it follows the water, one can infer velocity in a Lagrangian sense.

Drifters can be deployed at other depths. The first deeper floats (known as SOFAR floats) were deployed in the sound channel in the ocean, at a depth of 600-800 m where the sound speed is a minimum, making this region a wave guide for sound waves. They generate a sound which was detected at listening stations hundreds of kilometers away and by triangulation the buoy position can be determined. This strategy was later reversed to have the listening device on the buoy and the pingers on land and the buoys renamed RAFOS floats. It is also possible to have these off the sound channel and still be able to detect the signals. More recently floats which do not require a ground based tracking array have been developed, known as ALACE (Autonomous LAgrangian Circulation Explorer) floats. They drift at a preset depth but every so often surface and transmit their position to a satellite, then re-submerge to their operating depth. The surfacing frequency is approximately once per month with an operating depth of 1000m, but these parameters are variable. Because they require no tracking array, they can be used in remote places. Assimilating this type of circulation data has not so far been attempted. An interesting extension is that floats measure the temperature profile as they ascend which can then be transmitted in real time to satellite. These are known as PALACE or Profiling ALACE floats.

Electromagnetic radiation does not penetrate the ocean more than a few mm. So unlike the atmosphere, this cannot be used for remotely sensing the interior of the ocean but sound can travel through the ocean quite well and is the basis for a tomographic method of ocean remote sensing. The sound speed varies with temperature and pressure, but only slightly with salinity. The basis of the method is then to accurately measure transmission times of sound waves from an array of sound generators to an array of receivers. Information on the thermal structure of the ocean in the interior region spanned by the array can then be obtained. Tomography is still in its development phase and has only been used in limited regions of the ocean. The field is well covered in Knox 1989.

On a horizontal scale of kilometers and larger, the top surface of the ocean can move up or down, either by the movement of mass or by thermal readjustments. The former process takes place on fast time scales largely in response to the wind via the barotropic mode. The second process can occur either by heating or cooling of the water column, or by internal readjustments, for example replacing warmer water by colder water or vice versa. This latter process tends to be the more important for understanding and predicting interannual climate variability. From space it is possible to detect these movements to very high accuracy: the instrument, called an altimeter, can measure to an accuracy of $\sim$ 2cms from a satellite in an orbit at a height of $\sim$ 1000km!, although a number of corrections have to be applied such as ionospheric correction, corrections for water vapour and ocean tides, before a useful oceanic parameter can be obtained. The most difficult correction is often for the Earth's geoid, the constant gravitational potential surface which is unknown at smaller scales. Deviations of the sea surface from the geoid imply surface currents through geostrophic balance. Satellites which have carried operationally useful altimeters include SEASAT, GEOSAT, ERS1, ERS2 and TOPEX/POSEIDON (TP). The task is to use these data to infer the subsurface state. From one set of measurements this is clearly impossible since an infinity of solutions is possible. By taking account of the temporal evolution of the data however the number of solutions is greatly reduced. The measurements are only made sub-satellite on a swath $\sim$ 10km wide so this instrument can observe only a small part of the ocean on one orbit, but over a period of ten days can observe the ocean globally.

The surface displacement can also be measured by tide gauge. This is a low cost, low maintenance instrument which can be positioned on islands and the measurements either recorded and then collected later, or transmitted by satellite. As part of TOGA, many islands in the tropics were instrumented and the data transmitted in real time. As tide gauges can only be deployed where there are islands only parts of the world can be observed in this way. The data can be useful for calibrating altimeter orbits as well as giving information on the ocean circulation.

3. ASSIMILATION METHODS

Data assimilation methods combine a model with data to obtain a better estimate of the ocean state. The techniques can, in a broad sense, be divided into sequential and model-trajectory methods. Sequential methods combine model and data at given (analysis) times. Once the analysis is carried out, a forward integration is made to the next assimilation time using the analysis as initial conditions for the model, and the process is repeated. In the model-trajectory approach, data gathered during a time interval is fitted, seeking the best model trajectory that fits the data over the whole interval. If the model is assumed to be perfect, the model trajectory is determined by its initial and/or boundary conditions, and so the problem reduces to a search for the optimal values of these "control variables".

3.1. SEQUENTIAL METHODS

The analysis step in sequential assimilation methods can be summarised by the equation

$$\mathbf{x}_a = \mathbf{x}_f + \mathbf{K}(\mathbf{y}_o - \mathbf{H}\mathbf{x}_f) \tag{3.1}$$

where $\mathbf{x}$ represents the model state variables (e.g temperature, velocities, etc), the subscript a indicates analysis and subscript f the previous forecast; $\mathbf{y}_o$ is the observation vector. Sometimes a climatological state is also incorporated into the analysis instead of, or in combination with, $\mathbf{x}_f$. $\mathbf{K}$ is a matrix of weights converting observed-model differences to model variable space incorporating the errors in each and $\mathbf{H}$, the forward model, converts model variables to observation variables where the error analysis is most easily performed. This equation assumes that observed quantities and model variables are linearly related. More general expressions for the analysis are possible, but for illustrative purposes the expression above is sufficient. Interested readers can see Lorenc, (1986), for the general case.

What distinguishes one sequential method from another is the way in which the weighting matrix $\mathbf{K}$ is determined. The simplest useful technique is "nudging" in which an analysis is effectively performed on every timestep and the last term of (3.1) appears as a forcing to the model equations which nudges the model towards the observations. The next simplest is "successive correction", which is usually less intrusive on the model dynamics. Analysis is performed intermittently and a weighting function $\mathbf{K}$ that decays monotonically with the distance in space and/or time between model variable and observation position is used. Several observations influence the analyzed value depending on the spatial and temporal decay-scales of the weighting function. The analysis step is performed infrequently, either keeping the same weighting function through the iterations, or modifying its decay-scale or "radius of influence", to get a better representation of the smaller scales in the analysis. Lorenc, (1992), has developed successive correction schemes that converge to the optimal interpolation analysis, discussed below.

In Optimal Interpolation (OI), the weighting matrix $\mathbf{K}$ is determined from a minimum variance estimation procedure. The method initially assumes that the error covariances of the forecast or background field, and the observations, are known. Years of experience in numerical weather forecasting have allowed meteorologists to obtain a fair representation of this covariance matrix. In the ocean, length-scales associated with observed ocean variability introduced in Gaussian-type correlation functions are frequently used to model it. The OI analysis is given by equation (3.1) with

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \tag{3.2}$$

where $\mathbf{B}$ and $\mathbf{R}$ are the forecast (often called the Background) and observation error covariance matrices respectively and the superscripts T and -1 denote matrix transpose and inverse - see Gelb (1974) or Daley (1991) for the derivation. The filtering and interpolating characteristics of this equation are nicely discussed by Hollingsworth, (1989). Computable expressions for the analysis error covariance can be deduced, allowing calculation of "error bars".

In the same way that successive correction methods can be viewed as approximations to OI, OI in turn can be viewed as an approximation to the Kalman filter. In the Kalman filter, there is an extra step which involves an evolution equation for the forecast error covariance matrix, besides the forecast and analysis steps, with the latter including the calculation of the analysis error covariance. The "perfect model" assumption usually made in variational model-trajectory methods as discussed below, can be relaxed. A time evolving forecast error covariance is very useful, considering that in both sequential and model-trajectory methods, a problem arises of how to weight previous, present, and future information to obtain the best analysis. The Kalman filter (or its extensions such as the extended Kalman filter or the Kalman smoother, see Gelb 1974) automatically provides those weights via the evolving forecast error covariance. However, the computational demands of the Kalman filter are so great that its application to large and even medium-scale numerical models is impractical.

The search for better (and computationally feasible) approximations to the Kalman filter that can simulate better the time evolution of ocean forecast errors, e.g. Fukumori and Malanotte-Rizzoli (1995), is a current research area of great interest. To reduce the cost of predicting error covariances, low resolution versions of the forecast models may be used to propagate the error, followed by interpolation to the finer grid at assimilation times. It may also be possible to calculate the propagation of error covariance for one variable and then compute the rest of the errors by a balance assumption, such as geostrophy, Dee (1991). Other recent ideas involve the use of iterative eigenvalue solvers (e.g. Lanczos) to compute either a few singular vectors of the tangent linear model and use those to evolve the forecast error, or computing an approximate forecast error covariance using a reduced set of its eigenvector/eigenvalue pairs, also computed iteratively. Both techniques, require an adjoint model for these calculations. Another method, based on the computation of an approximate Cholesky decomposition of the analysis error covariance, allows a feasible computation of the forecast error covariance evolution without the need of an adjoint model (see Cohn and Toddling 1995). An interesting ensemble approach to calculating the error covariance has been developed by Evensen and van Leeuwen (1995)

3.2. VARIATIONAL METHODS FOR TRAJECTORY ANALYSIS

The variational method, first proposed in principle by Sasaki (1970), was given a large practical boost by the work of Lewis and Derber (1985), Le Dimet and Talagrand (1986) in meteorology and Thacker and Long (1987) in oceanography. At the heart of the variational approach is the minimization of some number defined as a 'cost function', or 'objective function', J, measuring the difference between the modelled state, or time trajectory of the system, and observations distributed in space and time. The usual problem is to find the solution to the model equations which most closely fits the observations over some time interval, by minimizing the cost function with respect to the free parameters. A basic form of this might be the quadratic:

$$J = \sum (\mathbf{H}(\mathbf{x}) - \mathbf{y_o})^{\mathbf{T}} \mathbf{R}^{-1} (\mathbf{H}(\mathbf{x}) - \mathbf{y_o}) \tag{3.3}$$

where the summation is over a finite time interval. The symbol $\mathbf{y}_o$ here represents the vector of observations at various locations and times, and $\mathbf{H}(\mathbf{x})$ is its model equivalent, with the vector $\mathbf{x}$ representing the model state variables also at different times. The measurements do not need to correspond explicitly to model variables. All that is required is that a model analogue of the observation can be constructed or diagnosed, and this is illustrated by the use of $\mathbf{H}(\mathbf{x})$ in equation (3.3), which is more general than $\mathbf{H}$ in equation (3.1), since it allows a nonlinear relation between state variables and observations.

What are the free parameters? In most applications they correspond to the model initial conditions at the beginning of the assimilation interval. These will completely define a model trajectory in space and time. In other applications they could be poorly known dynamical parameters in the equations themselves such as, for example, the values of turbulent mixing coefficients, or some measure of the forcing, if it is also considered to be uncertain. Efficient minimization routines such as conjugate gradient or quasi-Newton require gradient information and for a problem with only a few free parameters, these gradients can be obtained by directly perturbing the parameters one at a time. However, for meteorological or oceanographic applications, where the number of free parameters is so large, typically from several tens of thousands to several million, a special approach is needed. Because the elements of the state variable vector $\mathbf{x}$ in (3.3) are not independently free parameters, but must satisfy the model constraint equations, finding derivatives of J with respect to the free parameters is not completely straightforward. A constrained minimization, however, can be posed as an unconstrained problem by introducing Lagrange multipliers λ. A constrained minimum of J is then reduced to finding a stationary point of the Lagrangian L where,

$$L = J + \Sigma\lambda(equations \;\; of \;\; motion), \tag{3.4}$$

where the sum on the second term is used to indicate that there are as many Lagrange multipliers as there are dynamic constraint equations and integrals over space and time as well. The advance by Le Dimet and Talagrand was to show that the unconstrained gradient of L with respect to all the free parameters could be obtained from one single integration of the adjoint of the equations of motion and that it was not necessary to vary each of the free parameters, one at a time.

A practical way to solve (3.4) for real problems was provided by Thacker and Long (1987). They proposed that from the start one write down the Lagrangian, not in its analytic form, but in its numerical (for example finite difference) form, with a Lagrange multiplier for every application of the finite difference equations. If one now differentiates the analogue of (3.4) with respect to $\mathbf{x}$, one obtains equations for the Lagrange multipliers in a finite difference form which is consistent with the original equations of motion.

Scarcity of ocean observations was one of the key issues that attracted oceanographers to the 4DVAR method since with it, data from different times are used to determine a single field (the initial conditions for the period). However, these data may still be insufficient to completely constrain the analysis. The assimilation period could be extended until there are at least the same number of observations as model independent variables (or control variables), but then the data at the later times might bear no relation to the initial state (for example if diffusive or chaotic processes are present). Equivalently the time period for which linear evolution of the Lagrange multipliers is valid will be limited. Furthermore, the observations may not be independent or the error bars in the observations could be so large that the estimation would be useless. A remedy for these problems is to bias the analysis towards a previously guessed field (e.g. a previous forecast) or to impose smoothing constraints to the solution. Both can be enforced by adding penalty terms to the cost function (3.4), so that the new Lagrange function becomes

$$L' = L + penalty \; terms. \tag{3.5}$$

In meteorology, it is common practice to use the previous forecast as bogus data, so the penalty term takes the form

$$J_B = (\mathbf{x}_f - \mathbf{x})^T \mathbf{B}^{-1} (\mathbf{x}_f - \mathbf{x}).$$

It turns out that adding this term to equation (3.3) and performing the minimization at a single time yields the OI analysis (equation 3.2) as solution, thus showing OI can be formulated as a minimization problem (Lorenc, 1986, See also Courtier, this volume, for an excellent classification of different assimilation techniques and their equivalences.).

Making the model fit the data without deviating too much from the first guess field or the penalty could ensure that the final result is smooth in some sense. The imposition of these constraints is based on prior knowledge or prejudice, and adding them as penalty terms to the data-misfit cost function, is equivalent to using 'bogus' data which help to single out a unique solution.

Another possibility with the variational technique is to relax the perfect model assumption made by the 4DVAR method. One can add the model equations as another quadratic term in the Lagrangian (3.4) instead of imposing the model equations as constraints using the Lagrange multipliers, i.e. the dynamics becomes a "weak constraint" instead of a "strong constraint" (Sasaki, 1970). In so doing, the minimization problem is no longer a constrained problem, and all variables at all positions and times become independent. A solution is sought which gives a compromise between observations and model equations depending on the relative weights given to the data misfit and model misfit terms in the cost function. Due to the huge size of the problem, the method has only been applied to small problems. Applications and strategies of solution are discussed by Bennett (1992).

The problem of calculating error bars for the solution is computationally intensive since it requires the computation of the second derivative matrix of the cost function with respect to the control or independent variables, also known as the Hessian matrix. If the cost function is based on statistical assumptions (i.e. the weights are determined by data and prior information errors), the Hessian is the inverse of the analysis error covariance, and therefore contains information about the precision of the solution. Eigenvectors of the Hessian associated with its largest (smallest) eigenvalues are indicative of the best (worst) determined linear combination of parameters, i.e. the Hessian also contains resolution information. However, if the size of the problem is n the Hessian is order $n \times n$. A finite difference approximation to this matrix requires n integrations of the adjoint model, in practice, an impossible computation for large-scale ocean models for which $n \approx 10^6$ to 10^7. In the applications discussed below, examples are given of approximate methods for error and resolution analysis, that can, in principle, be applied to large-scale models.

The most staightforward application of the 4DVAR adjoint technique requires 3 integrations per iteration. First the equations are integrated over the time period chosen. Then the adjoint equations are integrated backwards in time to $t = 0$, forced with the difference between the observations and the forward model values. This gives the values of the gradient of the cost function with respect to all the free parameters, but does not indicate how to adjust the parameters in a descent algorithm, so a third integration is normally needed, using the forward equations in order to gauge a step length. The free parameters are then updated and the process repeated. Whether the variational method is of practical value or not, depends on how rapidly convergence can be achieved and this is hard to gauge a priori. For some applications, convergence has been quite acceptable, for others not.

In the following, some examples are given of 4DVAR applications to illustrate a range of uses spanning most major data types available to oceanographers. However, simpler techniques such as optimal interpolation or its nominal equivalent, 3D VAR, have also proven useful and will be discussed below. We will start with the tropical problem in chapter 4, since that is the most relevant to interests here at ECMWF. Some material is cross-cutting and could appear in more than one section.

4. TROPICAL ANALYSES AND CLIMATE FORECASTING

4.1. Applications of 4DVAR to tropical oceanography

The first attempts to use a 4DVAR scheme in oceanography were made by Thacker and Long (1987) who applied the method to a simple model of the equatorial Pacific formulated in terms of parabolic cylinder functions so as to have only a modest number of degrees of freedom. They used the model to assimilate various distributions of synthetic data from identical twin experiments and to examine the relative importance of altimeter and XBT data in constraining the circulation.

Subsequent studies were carried out by Sheinbaum and Anderson (1990a,b). Their main thrust was directed to using real XBT data from the tropical Pacific ocean, but the model was simplified to a single layer for the upper ocean. Despite the simplicity of the model, which had only a few tens of thousands of degrees of freedom, the results are instructive and applicable in a wider context. Two experiments can be compared, both starting in January: the first is a control in which no data are assimilated while the second is a 6-month active assimilation run. In the former, there are large discrepancies between the model and the data, especially in the east Pacific where the thermocline is consistently 50 m deeper than observed. In the assimilation run, the model fits the data much more closely in January, than in June. Although the assimilation attempted to fit the data over the whole 6 months, this was not possible as there were no initial conditions which permitted a good fit through the whole trajectory because of systematic model or forcing error. Various penalty terms were added to the cost function either to penalise departures from the first guess or to add smoothing terms by penalising gradients or curvature of the anaysis. However, the variational method also has filtering properties: large scale features in the data tend to be extracted in the first few iterations with later iterations tending to extract progressively smaller-scales. Thus, it might not be desirable to iterate too much. A similar conclusion is reached by Derber and Rosati (1989) for a 3DVAR scheme.

Although a model can to some extent adjust to the data it may not retain the data for long after assimilation ceases. An inadequate model could rapidly loose a memory of the assimilated data. To analyse this question a data-retention experiment was performed. Data were assimilated for 6 months (January to June) and then the model run forward for the following 6 months from July to December with no further data assimilation. The fit with the data during this later time was used for assessment. The model retains information in most regions with one notable exception, namely the eastern equatorial Pacific, where the impact of data is quickly lost. The rapid loss of information is a manifestation of an inconsistency between the model and the data. The primary balance in the equatorial region is between the wind stress and the pressure gradient. If the wind stress is too light or the ocean boundary layer parametrisations inaccurate, then it will not support a large enough thermal gradient along the equator. Assimilation of data in the east will cause cooling, increasing the thermal gradient and lifting the thermocline making the model shallower in the east, so correcting for the model forcing deficiency. When assimilation is turned off, however, the thermocline is unsupported and since adjustment in the tropics is fast, the thermocline quickly relaxes, typically in a month. Similar results were found earlier by Moore and Anderson (1989).

Weaver extended the work of Sheinbaum and used a multilayer linear model to assimilate data over periods of order 6 months. The motivation was to see how the altimetric measurements of the top surface could be propagated into the deep ocean. Using altimeter data is not straightforward. One source of difficulty is projecting the surface data vertically. In Weaver and Anderson (1997), there is no explicit projection: the model is left to handle the downward transfer of surface data and this may not be done in an optimal way. Over time scales of months the altimetric data can be used to identify various equatorial waves in the model and subsurface equatorial flows were captured successfully just from surface data. Several degrees off the equator where the waves propagate more slowly the model is not able to easily identify different waves and the propagation of energy to the subsurface is not well done even over timescales up to a year. Here there may be some merit in using statistical relationships between the surface topography and subsurface changes to speed up the penetration of the signal. This is discussed further in the next chapter.

For a number of years there has been an operational ocean analysis run at the National Centre for Environmental Prediction (NCEP) (Ji et al 1995) which uses a multilevel PE model and a 3DVAR assimilation scheme based on the work of Derber and Rosati (1989). The main data sources are subsurface temperature, mainly in the depth range to 500m, from the Atlas buoy array of Fig. (1a), and XBT data, collected from the array of Fig. (1b), augmented with surface temperature data from ships and satellites. The assimilation uses a semi-continuous insertion scheme. Data within a 30 day window are used with the weights decreasing to zero 15 days before and after analysis time. The correction is applied every time-step. The method of assimilating data is to solve a three dimensional variational problem. All data are formally used in the analysis of every grid point but data far from the analysis point have zero weight. The cost function to be minimized is:

$$J = \frac{1}{2}(\mathbf{T} - \mathbf{T}_o)^T \mathbf{R}^{-1} (\mathbf{T} - \mathbf{T}_o) + \frac{1}{2}\mathbf{T}^T \mathbf{B}^{-1} \mathbf{T} \tag{4.1}$$

where $\mathbf{T}$ is the correction to the first guess and $\mathbf{T}_o$ is the difference between observations and the first guess. The first term is a measure of the difference between the analysed field and the observation, while the second is a penalty term, penalising departures from the first guess. The observational covariance matrix is taken, as is common practice, to be diagonal, implying different observation errors are uncorrelated. The true structure of the matrix $\mathbf{R}^{-1}$ is not known: Ji *et al* (1995) crudely approximate this to be isotropic with an e-folding scale of 400 km. Physically the correlation scale is definitely not isotropic in the equatorial region, but it is possible that the errors in modelling this may be more isotropic than the fields themselves. There is no clear concensus on this, however, but most users favour a non-isotropic functional form.

The minimization of (4.1) is found using a conjugate gradient algorithm, which is relatively expensive, but only a limited number of iterations is used, typically three. To further save computing time, the scheme is not implemented each model timestep. Rather an intermittent procedure is used, whereby corrections are calculated for three timesteps and the corrections then held fixed for the next 9. The minimisation is applied to the full 3D structure all at once, but, as there is only weak coupling between the levels, they are effectively analysed independently. Only the temperature field is directly adjusted as only thermal data are assimilated. There are too few salinity measurements or current measurements to make assimilation worthwhile. However the continuous correction of the thermal field allows the velocity field to be adjusted dynamically by the model. The salinity field is not tied to the thermal field and in this approach the salinity field could drift if not restrained somehow, for example by a relaxation to the climatological values.

Several tests were employed by Derber and Rosati (1989) to assess the operation of the scheme, but the one of most significance is the data-retention experiment, similar in outline to those described above using simpler models and with similar results.

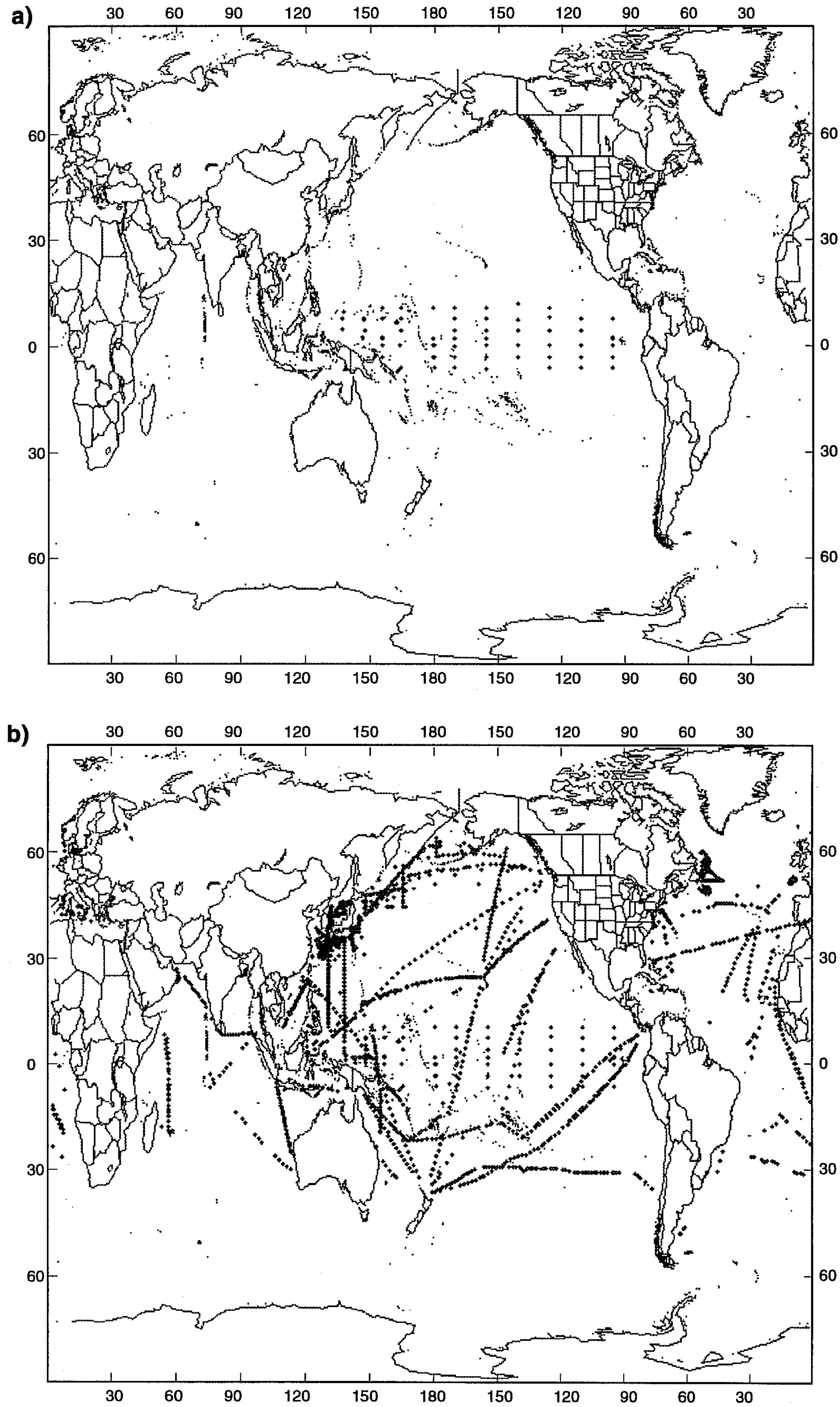

Fig. 1 Upper panel shows the location of TAO moorings, which are capable of measuring surface wind velocity, humidity, surface and sub-surface temperatures. From McPhaden 1994. Lower panel shows a typical distribution of data collected in one month (July 1995). TAO data are included, but outside the equatorial Pacific most data are from XBTS.

Three integrations are compared; a control in which data are not assimilated, an active assimilation run of 1 year duration, and a retention run in which data are assimilated for 6 months and then the model run forward for a further 6 months with no assimilation. In the latter case, errors in the equatorial region grow rapidly, relative to the continued assimilation case and quickly become comparable to those in the no assimilation run. In other words, the model quickly rejects the data near the equator but retains it for longer in the extra-tropics. The rapid rejection of data in the tropics is presumably due to the generation of Kelvin waves as indicated in Moore and Anderson (1987).

Ji *et al* (1995), show that assimilated data from the current TOGA network together with satellite SST data are still not enough to offset errors in the forcing fields such as wind stress and heat flux. The difference in the mean depth of, say, the 20°C isotherm for the two year period 1991-92 in the tropical Pacific, obtained from two analyses which assimilate the same data but are forced with different wind products ought to be small if assimilating data could correct for forcing errors, but in fact relatively large differences between the two analyses occur, particularly in the eastern Pacific around 10° North.

The results from the Ji *et al* analyses are published monthly in the Climate Analysis Bulletin. On the basis of these analyses and other information, preliminary warnings are given of impending El Niño conditions, (see Anderson 1995 or Palmer and Anderson 1994 for aspects of seasonal prediction, an important topic not covered in this article).

The above assimilation schemes all concentrate on assimilating thermal data. This is because temperature is by far the easiest quantity to measure. But what is being lost by, for example, not having current information? At first sight this seems serious because geostrophic adjustment theory suggests that while 'mass' (thermal) data control the adjustment process at large scales in the extratropics, velocity data control the adjustment at small scales or as the coriolis term gets small near the equator. However this analysis does not take account of equatorial or Rossby wave dynamics. Anderson and Moore (1989) showed that mass field data are even more important close to the equator. In an inviscid ocean, the equatorial Kelvin wave is equipartitioned in terms of kinetic and potential energy, implying that to correctly initialise this wave both velocity and mass field data are important. As soon as one goes just a few degrees off the equator, potential energy dominates kinetic, implying a greater role for mass field data. This result was confirmed by Moore *et al* (1987) in a comprehensive ocean general circulation model. The primary balance along the equator at low frequencies is between the pressure gradient, largely determined by the thermal field, and the wind stress, again arguing for an important role for thermal data. Thus, fortunately, the field we can best measure, temperature, is the most important.

Rosati *et al* (1995) also use the 3D Derber-Rosati variational assimilation scheme, but this time in a global ocean model, in contrast to Ji *et al* who apply it to the tropical Pacific, although otherwise the models are similar. As their 3DVAR is expensive, one would like to know that the extra effort was worthwhile. So a further experiment was performed in which the SST field was assimilated by nudging the model SST towards the observations. In both this and the 3DVAR, the model was forced by 'observed' winds. The nudging scheme is very easy to implement and involves no additional computational overhead to running the model. The hope is that by forcing the surface to be correct, this information will be propagated downward into the subsurface ocean by the model itself. Did it work? The test of the quality of an assimilation scheme can be obtained by assessing how good are the forecasts made using a coupled model initialised from the ocean analyses. The 3DVAR analyses lead to much higher correlations between the predicted and observed SSTs, the measure of skill used in this study, apparently vindicating the effort put into the 3DVAR scheme, but see later.

At ECMWF, a full end to end prediction facility is being developed, consisting of a general circulation atmosphere coupled to a global ocean model. The ocean initial conditions are obtained in one of two ways. In the first, the ocean is forced by a chosen wind field, say the stresses, heat flux and fresh water flux from the ERA reanalyses for the period 1979 to Feb 1994, and from the operational analyses from Feb 94 to current time. During this integration of the ocean model, the SST is relaxed strongly to the observed values obtained from the Reynold's OI SST analyses which is almost equivalent to assimilating Reynolds SST analyses. No data are assimilated in the ocean subsurface. In the second, subsurface thermal data also are assimilated. The coverage from the TAO array is good: daily-averaged measurements are obtained from each mooring once a day, at 10 depths ranging from the near surface to 500m. In addition XBT measurements of T(z) from ships are used. The temporal and spacial coverage is less homogeneous than for TAO but covers a larger geographical area. There are too few salinity and current measurements to make an assimilation of these quantities possible. Quality control is performed by comparing against climatology and buddy checking by performing an analysis for all data except the datum being checked. However, despite these checks, it appears that data can still be wrongly used. If the model or forcing has a serious error at some time, the model first guess will diverge from the data, the data will be rejected and the model will remain wrong. When to use and when not to use data is a difficult issue. At present, the number of sub-surface ocean measurements is sufficiently small that data can be checked manually, a luxury not open to meteorologists. Lorenc has considered at length the problem of how to screen data. It seems a technical, tedious matter, but in fact can be very important. The data you most want are those which differ most from the model, provided of course they are correct. These are the data most likely to be rejected in a very abrupt, nonlinear way.

In the integrations to which I will refer below, data are condensed in to superobs. The observation error is taken proportional to the vertical stratification. The background error is taken proportional to the model stratification. An OI analysis is performed every 10 days blocking the data into 10 day windows. The increment is not applied discretely in one hit immediately following an OI, but rather is spread smoothly over the next 10 days. This prevents shocks and allows the model to adjust its own velocity field, as the OI is univariate and does not analyse velocity. The scheme is also single level i.e. each level is analysed independently of every other level. This should probably be improved at some stage. Further details of the scheme can be found in Smith *et al* 1991 and Smith (1995). The fields are analysed in boxes containing no more than 50 measurements. This condition will also be relaxed and larger boxes used. However, care is taken in overlapping the boxes and producing consistent analyses from one box to the next and preliminary tests suggest that analysing several small boxes with less than 50 observations gives analyses similar to using more observations in a box.

The test of the quality of an analysis is the quality of a forecast made from that analysis using a coupled atmosphere ocean model. To assess the impact of data assimilation, forecasts of 6 month duration have been made every season for the last 16 years, back to 1980. To assess the quality of forecasts over this period, the rms error and the correlation between the observed temperature and that predicted are calculated for specially sensitive areas. Specific examples for the NINO3 region in the central equatorial Pacific (±5°, 150W to 90W) are shown in fig 2 for two different wind fields, ERA and FSU. In both cases the data assimilation forecasts are better than the no data assimilation forecasts using the same wind forcing, for the first few months but it is less clear that this holds for forecast lead times beyond a few months. The forecasts using FSU winds seem to be better than forecasts made using ERA winds, indicating that data assimilation is not really able to compensate for differences/errors in the wind forcing.

This ensemble is based on 60 forecasts (4 per year for 15 years), which while it is much larger than the set used by Rosati, is still probably too small. This involves over one hundred years of coupled integration, a lot of computing time, but even so the sample size is still rather small and the significance of the differences between the various integrations may not be terribly significant. The results are in stark contrast to those of Rosati who found very large differences between the forecasts started from analyses with sub-surface data assimilation and those without. This is partly because his forecasts without data assimilation were poor, so allowing a considerable improvement when using assimilation. In the case of fig 2 the forecasts without data assimilation were not too bad, so the impact of data assimilation was reduced. However, one should not conclude the forecasts were as good as they could be. There were significant errors whether data assimilation was used or not, suggestive of model or forcing error. Data assimilation is being used to correct for model or forcing error which should not be the case. It would be better to correct for systematic error by improving the model and/or forcing. This is

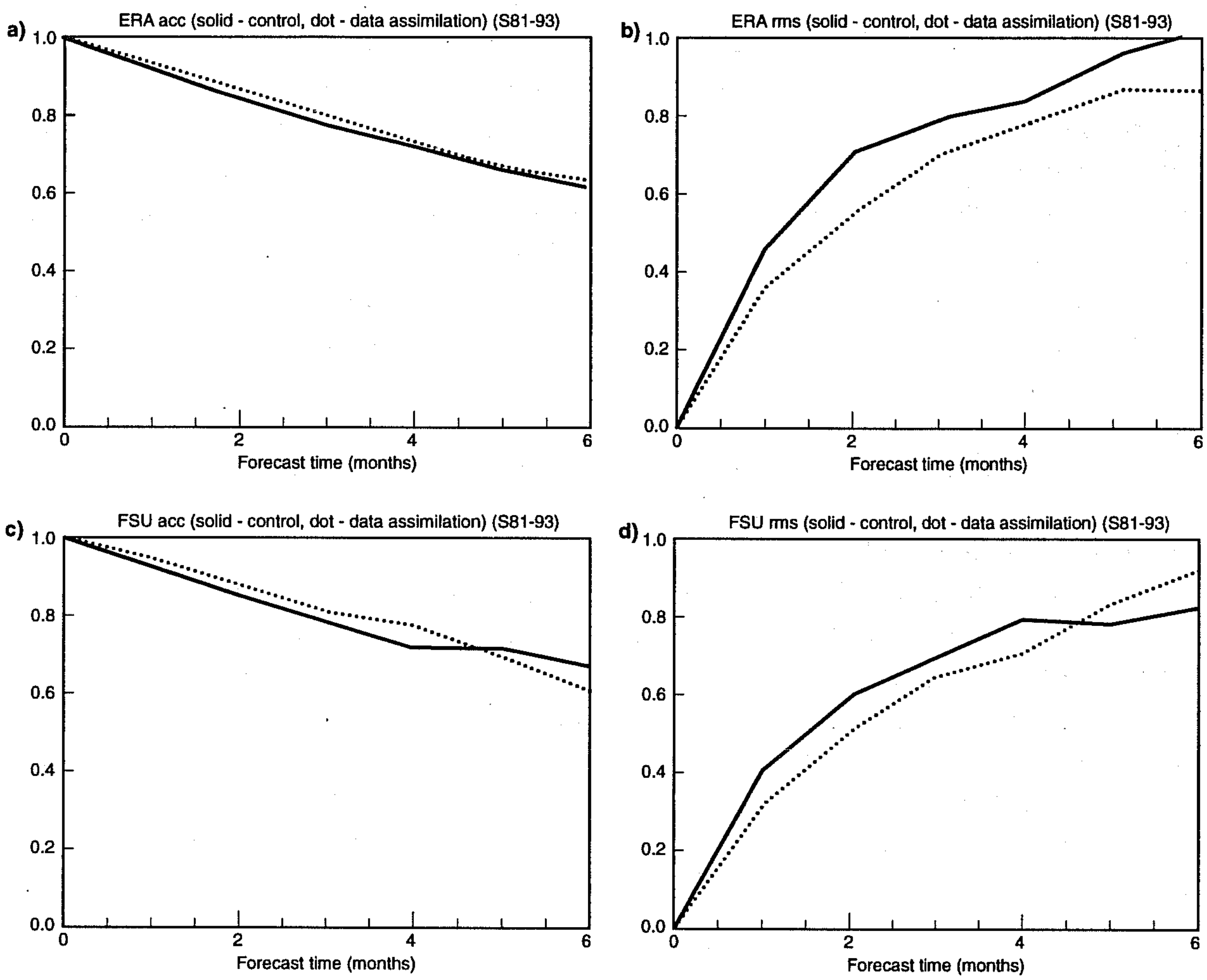

Fig. 2 (a) Correlation between the predicted value of Nino3 SST and that observed as a function of lead time, for two experiments. The solid curve corresponds to forecasts started from ocean initial conditions prepared using ERA winds, the dotted curve corresponds to those started using ERA winds but in addition assimilating subsurface ocean thermal data. In both cases the model was relaxed strongly to observed SST in the preparation of ocean initial conditions. (b) The rms error between the predicted SST and observed for the same two experiments as described in (a). Data are averaged over 52 forecasts made every 3 months over the 13 year period 1981-1993 inc.. (c) As (a) but using FSU winds. (d) As (b) but using FSU winds. The forecasts made using data assimilation have higher correlations and lower rms errors than those made without for the first few months. However, given the variability in forecasts arising from chaotic processes, it is unclear at this stage what measure of significance to ascribe to these differences.

easier said than done, however, and considerable effort is required to make even modest gains. Nonetheless, the message should be clear that data assimilation is not a panacea for model or forcing field ills.

An OI scheme is also being used by Carton and Guise (1996), initially for the tropical Atlantic, but now for all the tropical oceans. In their scheme, data are assimilated intermittently, typically every month, whereas in Ji *et al* the data are inserted continuously. The advantage of the former is computational efficiency and a clearer assessment of the analysis error structure. Proponents of continuous assimilation would argue that more use is made of the data and the model is not shocked to the same degree since each increment is small, and the velocity can adjust more easily to the mass field. In the OI, a larger adjustment is made intermittently and a velocity correction can be calculated based on the analysis correction for T using geostrophy except near the equator. However, in most cases no geostrophic balancing is done which means that the model must generate its own velocity adjustment. This may be done in such a way as to erroneously excite gravity waves. Controlling gravity wave excitation by normal mode initialisation has been addressed in part by Moore(1990). Clearly, a better approach is to use 4D variational assimilation in which the best trajectory is found over an extended time period. Then the model shock problem essentially goes away. No comprehensive scheme currently exist, however.

5. USING ALTIMETER DATA

Several attempts have been made to use satellite altimeter data in ocean circulation models. The first problem to be overcome is the question of eliminating the geoid error from altimeter observations. One way is to use altimeter differences between repeated satellite passes. The geoid is completely eliminated but the signal from the mean ocean circulation is lost. In effect only difference information is used. Another way is to add the difference between the altimeter and its time mean to the model time mean. Again no direct correction to the mean state of the ocean would result. The second problem is that it is not immediately clear how to use surface data in order to extract information about the sub-surface.

5.1. Recovering sub-surface flows

Early attempts to assimilate surface height data with models involving more than two vertical levels produced mixed results. Berry and Marshall (1989) could not recover deeper flows but Holland and Malanotte-Rizzoli (1989) could. Both studies used 3-layer models in twin experiment format where artificial altimeter data from a previous model run were assimilated. Berry and Marshall re-initialize their model with the correct surface flow, determined by a surface streamfunction, but leaving deeper flows unaltered at assimilation time. In contrast Holland and Malanotte-Rizzoli (1989) used a 'nudging' method which adds to the upper layer vorticity equation an additional term, proportional to the difference between the model vorticity and the observed. When surface vorticity is altered, there is an immediate effect on the deeper layer currents, because force balances within the flow (hydrostatic and geostrophic) are built into the QG model equations. This is part of the reason for the greater success of the Holland and Malanotte-Rizzoli (1989) experiment.

Haines (1991) modified the assimilation method of Berry and Marshall and was able to recover the flow in the lower layers with comparable success to nudging. Instead of vorticity he considered 'potential vorticity' which is a conserved property of water parcels in adiabatic, frictionless motion. By specifying surface flow by observation, while the deeper potential vorticities retain their previous model values, the deeper layer currents may be found by solving an elliptic equation allowing the model to be re-started after assimilation with a new 3D flow. Later he constructed a 3-layer primitive equation (PE) model based on the shallow water equations and compared the merits of direct model restart with nudging assimilation. Nudging with a rapid adjustment time of typically 1 day gives comparable results to the restart method although the nudging procedure is much easier to implement than the restart as no elliptic inversion is required.

5.2. General circulation model assimilations

Although instructive, the models discussed above are much simpler than realistic ocean general circulation models (OGCMs) currently in use for simulation of ocean basins. Assimilation using full OGCMs has only recently been addressed, initially mainly by feasibility experiments using simulated data although applications using real data from Geosat, Topex/Poseidon and ERS are now in hand. The experience gained with simple layer models has greatly helped in developing altimeter data projection techniques for OGCMs but new methods have also been proposed.

The main problem is how to induce changes in the density field from the altimeter data. This problem was not present in the layer models simply because they do not have temperature or salt as dynamical variables. Cooper and Haines, (1996), have extended the restart method for use with such models. The basic principle is similar to the one discussed above: To conserve potential vorticity on as many water parcels (on isopycnal surfaces) as possible. The philosophy can now be stated in a rather succinct manner: altimeter data should largely not change the water properties in the model, but only cause a spatial rearrangement of the water. Changes to the Lagrangian properties of water parcels should then be produced only between altimeter assimilation times by model physics.

To modify the model using the altimeter data, hydrostatic and (optionally) geostrophic force balance conditions are imposed on the field changes:

$$\Delta p = \Delta p_s + \int_z^0 g\Delta\rho dz$$

$$\Delta \mathbf{u}_g(z) = \frac{1}{\rho_0 f}(\mathbf{k}\times\nabla\Delta p_s + \mathbf{k}\times\nabla\int_z^0 g\Delta\rho dz); \tag{5.1}$$

where Δp_s, the change in surface pressure, is given by the altimeter data, and $\Delta\mathbf{u}_g$ is the associated geostrophic updating current vector as a function of depth, and $\Delta\rho$ is the change in water density. The trick is to decide on an appropriate $\Delta\rho$. One way is to find $\Delta\rho(z)$ in terms of a vertical displacement of each water column, $\Delta h(x,y)$, which is a naturally adiabatic procedure. For large scale flows this also conserves potential vorticity which is then only a function of the latitude and the stratification. Water can be added or removed from the top or bottom levels of each water column although this exchange can be made to balance globally. To determine $\Delta h(x,y)$, an extra assumption is made that at some depth the pressure does not change at assimilation time. Taking z_0 as that level (CH use the bottom), the distance Δh is determined from the condition:

$$\Delta p(z_0) = 0 = \Delta p_s + \int_{z_0}^0 g\Delta\rho dz \tag{5.2}$$

which if the surface and deep densities were uniform, leads to

$$\Delta h = \frac{\Delta p_s}{g[\rho_s - \rho(z_0)]} \tag{5.3}$$

To balance the observed change in surface pressure, light surface water is exchanged for heavier deep water locally to maintain condition 5.2.

To test the scheme, a twin experiment using an eddy-resolving version of the Bryan-Cox ocean model was performed. The fields at the end of model year 30 are used as the initial model conditions and those at the end of year 29 are used to provide simulated altimeter data (over the whole domain). The errors within the thermocline (top 1km) decrease immediately while errors below decrease by 60 percent after repeated assimilation every 9 days for a year.

The use of statistics to project ocean surface data downwards is the main alternative to the dynamically adjusted restart and nudging schemes discussed above. De Mey and Robinson (1987) were among the first to investigate this, using vertical pressure correlations and inferring density from hydrostatic and geostrophic balances. Mellor and Ezer, (1991), and Ezer and Mellor (1994), derive empirical correlations between surface height anomaly and density, temperature and salinity anomalies, from their numerical model as a means of calculating $\Delta\rho$ in Eq. (5.2). The correlations are not usually perfect, although in the experiments done so far they tend to be high over large areas of the ocean and down to 1000 m depth. The main drawback of such a method is that the statistics derived from the model are not necessarily the same as those from the real ocean which may bias the whole procedure when used, for example, with altimeter data.

Pinardi *et al* (1994) have developed a global ocean assimilation system for altimeter and hydrographic data based on a combination of the Mellor and Ezer projection method for the altimeter data, with actual measured water properties inserted into a model with a continuous updating technique. They calculate the temperature correlation coefficients for constructing the synthetic profiles from 24 monthly mean ensemble averages of a model run in which hydrographic temperature data are assimilated. The strongest correlation between temperature and sea surface height anomalies is found around 50 m depth over most of the world oceans. Correlations are lower at the surface due to rapid air-sea heat exchange, and decay relatively rapidly for depths below 50 m.

Initially identical twin assimilation experiments were carried out to examine the usefulness of the synthetic temperature data derived from the sea surface height. Now real TOPEX/POSEIDON data are used with time dependent correlation functions computed every month from the model output (Pinardi *et al* 1996). Preliminary results show the impact of TOPEX data over XBT data only in areas devoid of XBT data.

Recent work by Oschlies and Willebrand (1996) uses vertical current correlations but derives density (and T, S) using a local water conservation method. They show that vertical current correlations provide good results in a twin experiment framework. They go on to apply their method with 1 year of GEOSAT data over the N. Atlantic using a high resolution Bryan and Cox ocean model. The model kinetic energy improves greatly showing that the model is able to accept the more intense eddy field indicated by the altimeter data. No forecasts were made but the thermocline structure within an eddy observed with hydrographic data during the period is reproduced well.

Other techniques have been developed to study larger scale ocean variations than those of eddies. Stammer and Wunsch (1996), for example, look at the time mean flow and seasonal variations in the Pacific basin with a Greens function assimilation method and TOPEX altimeter data. The method assumes that misfits between model and data behave in a linear fashion and can be modelled by the tangent linear model discussed in section (3). Greens functions for the system are constructed numerically and used to assimilate the TOPEX data to improve the model flow in the Pacific over a 1 year period.

5.3. The importance of altimetic data

Over much of the world's oceans, the altimeter is the only instrument capable of providing synoptic coverage. In such regions the altimeter is likely to be very useful, although the best way of extracting the signal is yet to be found. There are regions where, largely as a result of TOGA, the coverage by in situ measurements is quite good, the tropical Pacific being an example. Within 8 degrees of the equator, observations are made continuously at 10 depths from the surface to 500m at the locations shown in fig 1 and daily averages of the temperature transmitted in near real time to the GTS. This area is also moderately well-sampled by ships of opportunity using XBT probes giving higher resolution thermal profiles down to comparable depths. One might expect then that the large scale patterns would be resolved well by the in situ measurements. Smaller features such as the tropical instability waves (TIWs) in the equatorial east Pacific would be less well resolved by in situ measurements. The role of the altimeter in this region might be to fill in the details of the TIWs and to extend the tropical signal outside of the TAO array area but not to contribute greatly to the larger scale within the 8S to 8N region.

Apparently contradictory results were found by Carton *et al* (1996) when they compared the results of assimilating altimeter, XBT and TAO data together (their control run) with three assimilation runs in which one of the observing systems was withheld in turn. They state that to resolve major features of the seasonal cycle, it is necessary to have either the altimeter or the XBT network and that TAO data are not crucial in resolving the seasonal cycle even within 8S to 8N. Studies of this type are inevitable in a cost conscious world and are in a sense useful but they can be difficult to interpret, and it would be contrary to my judgement to conclude that TAO is only a minor player. My interpretation of the results is more that with holes in the TAO array in the west Pacific, there is scope in combining the XBT and TAO, and that neither observing system is totally redundant though there might be some overlap between the two arrays. However as is pointed out by the work of Hollingsworth (1989) and others, some redundancy is a good thing: without it one can not validate different observing systems.

Some of the results of Carton *et al* showing a major contribution from the altimeter stem from the use of different observing windows: ± 5 days for the altimeter, ± 30 days for the TAO data even though the TAO arrays reports daily. Much information is lost by using a window as long as 60 days. In such a time, major features in the tropical region, such as Kelvin waves, can travel long distances (almost the whole Pacific) and are in fact very well resolved by the TAO array. A further caveat concerns the metric used to assesss the importance of the different observing arrays. Variability relative to the control analysis is the one used. Now the altimeter over-corrects thermocline variations relative to the XBT data because of the statistical relationship used to relate surface height to depth of an isotherm. Thus these two systems are competing and withholding either is likely to exaggerate their importance in terms of variability. Since the purpose of an analysis is to provide initial conditions for a forecast, the best analysis is the one that gives the best forecast. This is the real metric. But it is easier to say than to act upon, for two reasons, and as Carton *et al* point out, they have not yet made this test of performing seasonal forecasts using the different analyses as initial conditions.

Why is it difficult? In principle, the mechanism is clear: you perform a number of predictions from the different initial conditions. Getting statistical significance is, however, hard. If you run a single prediction, many times from just slightly different initial conditions, say starting one day appart for the atmospheric conditions then the forecasts after a few months will differ by up to $\pm 0.75K$, which is comparable with the size of the signal you are trying to predict. For any given start date one probably needs an ensemble of 15 to detect any difference in forecast skill resulting from different initial conditions. This might be reduced if you average over many start dates, but still involves many tens to hundreds of years of coupled integration: a pretty daunting task. Small wonder then Carton *et al* have not yet made this test. A second complication is that data assimilation is compensating for model error i.e. is being used to correct for systematic error, rather than for random error. If there are systematic errors then these should be removed at source, but again easier said than done since systematic error can arise either from errors in the winds and heat fluxes or error in the ocean model.

The scheme used by Carton *et al* to assimilate altimeter data is worthy of mention. They use an OI scheme and use both thermal and altimeter data in the OI, performed on boxes of size 5 degrees by 5 degrees. Although this sounds like a serious limitation, our experience is that it need not be, provided the boxes all overlap sufficiently and one takes care in the blending between boxes. Ideally a single box should cover the whole domain, but the practicalities of finding the weights in OI by solving a linear matrix equation limits the domain size. Computing restrictions a decade ago meant these boxes were quite small. Now one could use perhaps up to 1000 measurements in a box. At ECMWF, we also use a box limited to 50 observations but find that the analyses made by using larger boxes are almost equivalent.

To assimilate altimeter data, they use the difference between the altimeter data between different 10 day tracks, thus eliminating the unknown geoid. It is really the surface height tendency which is being used, rather than the surface height. This is then related statistically to subsurface temperature changes, derived from the TAO array. It is therefore a statistical relationship between observed height and observed temperature. (Sometimes this relationship is obtained between model height and model temperature. The true relationship between observed quantities is clearly preferable: unfortunately there are few places in the ocean where there are sufficient data to determine this relationship reliably, the TAO array in the Pacific being one of the few). In their work they use a single relationship, i.e. it does not vary geographically, and this can give rise to the over perturbing of the thermocline mentioned earlier. No changes are made to salinity in this process, in contrast to the approaches of Haines and Oschlies and

Willebrand referred to earlier. The latter work was addressing the deeper extratropical circulation, rather than the upper tropical ocean, but it remains an open issue whether one should perturb only T or try to preserve T S relationships.

Both Fox and Haines (1996) and Forbes (1996) have tried using the water-mass conserving approach in the tropics. Fox and Haines, using a high resolution global model find that the western boundary currents are improved, the eddy structure is improved and equatorial Rossby waves are represented better as a result of assimilating altimeter data, but their experiment is an identical twin, taking perfect data from one integration to another which bypasses the geoid problem. And there is no other source of data. So the results can not be compared to Carton *et al* . By contrast Forbes has used real altimeter data in a high resolution model of the tropical Pacific. He avoided the geoid problem by using the altimeter data as departures from a two-year mean. Before the data can be assimilated into the model, the total height field is needed. This is obtained by adding the altimeter anomalies onto a derived mean field. Two versions were tried. One obtained through forcing the ocean model with FSU winds, the other through data assimilation of XBT and TAO data into a slightly different ocean model. The results indicate that the TP data can pull the model from one mean state to the other as measured by surface height, which is not too surprising. There is some indication that the subsurface field might be improved too, but it is less clear the extent to which this happens or how it is achieved.

Model variability is also improved, particularly with respect to the higher frequency Kelvin and Rossby waves triggered by intraseasonal/westerly wind bursts. Some of this variability is damped in the model by using monthly mean wind stresses. The inconsistency between the wind forcing and the frequency of assimilation of TP data, could be reduced by using higher frequency winds, such as those produced four times daily by the UKMO operational analysis/forecasting system. However, operational winds are not without their problems, which is why Forbes used monthly-mean FSU winds to drive the model. No forecasts using the coupled model have been produced from the analyses with and without TP data. So again it is hard to judge the real impact of the TP data. Forbes does point out that spurious waves are generated sometimes for reasons which are not clear, and which will require further investigation.

One set of experiments has looked at the impact of sea level data on forecasts. Fischer *et al* 1997 have assimilated sea level from the NCEP analysis into an early version of the HOPE ocean GCM. From these ocean initial conditions a number of forecasts have been made using a hybrid coupled model consisting of a GCM ocean coupled to a statistical atmosphere. The method of assimilating sea level data is to project it vertically onto subsurface temperature, similar to the approach of Mellor and Ezer, but projecting onto the first five vertical EOFs rather than directly onto T. Whether this intermediate filtering using vertical EOFs is particularly beneficial is unclear. Interpreting the coupled results is also unclear. There appears to be no statistical difference in the forecast skill if the metric is correlation. The forecasts are not identical, but the average skill over 4 seasons for 7 years (28 forecasts) is similar. If the metric is correlation of temperature tendency rather than temperature, the results again are ambiguous. There is an impression that the short range forecasts are worse but the longer range predictions are better. However, given the chaotic nature of the coupled system, it is unclear if the sample size of 28 forecasts is sufficient. Fischer *et al* have a sense however that there is a qualitative improvment in the predictions, not necessarily reflected in the correlation measure of skill.

Arnault and co-workers have made extensive use of altimeter data from the earlier GEOSAT altimeter, using a variety of techniques. Examples range from using a Kalman filter approach in simple 1 vertical model (Gourdeau *et al* 1992), through a 3D var approach using a 3 mode model (Bourles *et al* 1992) to a 4D variational approach using a GCM and its adjoint (Grenier Arnault and Morlière :private communication). The 3D approach seems to avoid the projection problem onto the modes by adding side constraints to the cost function penalising departures from smoothness, and from the first guess. Whether this is the best way to use altimeter data is open: the authors find improvments in the model dynamic topography as a result of assimilating the data, but it is not clear if this has been projected correctly onto the modes of the model. The 4D assimilation is a very challenging study using a full GCM but with simplified physics for the tangent linear and adjoint in order to make the evaluation of the cost function gradient practical. There are also interesting ideas on how to reduce the large volume of data needed to specify the trajectory. The 4DVAR is an exciting assimilation technique, but no seasonal predictions have been made using analyses generated by this technique. It is clear this is a goal for which one should strive, but it will not be easily achieved. The presence of systematic error both in the analysis stage, and in the form of drift in the coupled model prediction can all too easily wipe out the hard-won gains from improved initial conditions through sophisticated assimilation procedures.

Finally we discuss the impact of altimetric data in combination with acoustic tomography as considered by Menemenlis *et al* (1997) for the Mediterranean. Acoustic tomography is a potentially important observing technique for the ocean. Sound can travel long distances in the ocean and, since the travel time is a function of the ocean density, one has, in principle, a method of remotely observing the deep ocean. An excellent review of acoustic tomography and present strategies for inverting the travel time information to obtain ocean temperature can be found in Knox (1989). Travel time measurements are line integrals of the sound slowness (the reciprocal of sound speed) along ray trajectories. Those data can be inverted to get the temperature field and then the temperatures assimilated into numerical models. This is the approach used by Menemenelis *et al* . The model is a GCM, but a linear reduced resolution analogue model is contructed from it and used to produce the analyses. The method is tested for a limited domain although in principle the application could be extended to the globe if warranted. The altimeter currently provides near global data but not yet acoustic tomography.

The variational approach allows direct use of the travel times since any observed quantity can be assimilated into a model, provided such a quantity can be expressed in terms of model variables (Sheinbaum 1995). Whether it is better to assimilate travel times, or to invert the data first, may depend on how best to quantify the observation error. For example, Stoffelen and Anderson 1997 suggest that to assimilate scatterometer wind data in an atmospheric model, it is better to invert radar backscatter to winds and assimilate these, than to assimilate radar backscatter measurements directly, because of the nonlinear relationship between backscatter and wind and its affect on the error characteristics. On the other hand, in meteorology, the radiances are now assimilated directly into the model, rather than inverting them to temperatures and assimilating these.

The elegance and sophistication of some of the techniques discussed above are a tribute to the improved quality of the newly available oceanic data, particularly TAO, XBT and altimetry, and to the growing confidence of oceanographers in applying these methods using the power of modern computer resources to tackle circulation problems on a global scale. There is still some way to go before ocean observation and forecasting becomes as routine as it is for the atmosphere, but the incentives are great because of the influence the oceans have on climate on timescales of weeks and longer.

6. LOOKING FORWARD

Data assimilation in ocean models had hardly begun a decade ago. Since then, the foundations have been laid with serious attempts to observe, model and forecast the oceans as never before. Assimilation techniques designed to analyse the instantaneous 3-D ocean state as well as to simulate a 4-D trajectory are both useful for operational forecasting. Several different strategies have been developed for analysing the instantaneous 3-D ocean variations. Some are similar to those developed for meteorological forecasting although some strategies are quite new and distinct, for example those in section (5) for connecting surface and subsurface properties using conservation methods. The extension to 4-D trajectory variation is evolving and although it has quite some way to go before an operational scheme is in place, nonetheless many experimental tests using 4-D variational assimilation have taken place to build up experience.

The first operational analyses are presently being carried out for all the tropical oceans. The profound impact of oceans on climate and the potential predictibility in the tropics has encouraged attempts to forecast ENSO events in the Pacific using comprehensive models.

Many of the studies described here are carried out by individuals, often in university environments, using small quantities of data over limited time periods. Such initiatives will remain vital for innovation, but a concerted effort at operational centres, where data assimilation will be carried out regularly, is also required. Only forecast centres can offer the quality controlled environment that the gathering and analysis of such large amounts of new data require. Fortunately, a number of centres are planning to develop such systems for the ocean. Examples are the initiatives at the US National Center for Environmental Prediction, the UK Meteorological Office, the Australian Bureau of Meteorology Research Centre, The Japan Meteorological Agency, and ECMWF.

In this brief review, the aim has been to describe sufficient of the theoretical developments in data assimilation along with applications to oceanic situations to convey the particular successes and problems which have occured over the last decade. Few of the applications have reached maturity. For this field, the main new requirement seems to be improved high resolution winds from meteorological models. In many respects 3-D ocean state analysis can borrow from the well developed method of OI which has been used in meteorology for many years. For 4-D trajectory analysis even meteorologists are only beginning to investigate the possibilities and oceanographers have had to find their own ways of using these methods.

In the field of forecasting it is the coupled ocean-atmosphere models of air-sea interaction in the tropics that hold the greatest promise in the near future and both 3-D and 4-D assimilation will be valuable here. In middle latitudes the new satellite instruments, particularly the altimeter, will offer real-time synoptic data for the first time. Tomographic data, although more experimental, holds the hope of extending synoptic coverage below the ocean surface. In mid-latitudes assimilation will not only help with forecasting but also to improve understanding of how the ocean circulation works and the role played by eddies in controlling heat and water transport and budgets within the major ocean basins.

References

Anderson D L T 1995 The tropical Ocean Global Atmosphere programme *Contemporary Physics* **36** 245-265

Anderson D L T and Willebrand J 1989 *Oceanic circulation models: combining data and dynamics* (Dordrecht: Kluwer Academic Publishers) p 616.

Anderson D L T and Moore A M 1989 Initialization of equatorial waves in ocean models *J. Phys. Oceanogr.***19**, 116-121.

Balmaseda M A, Anderson D L T and Davey M K 1994 ENSO prediction using a dynamical ocean model coupled to statistical atmospheres *Tellus,* **46A** 497-511

Bennett A F 1992 *Inverse methods in Physical Oceanography* (Cambridge: Cambridge University Press) p 346.

Berry P and Marshall J 1989 Ocean modelling studies in support of altimetry *Dyn. Atmos. Oceans* **13** num 3-4, 269-300.

Bourles B., S. Arnault and C. Provost 1992 On altimetric data assimilation experiments in a linear model of the tropical Atlantic ocean. Oceanologica Acta **15** 525-535.

Busalacchi A.J. 1996 Data assimilation in support of tropical ocean circulation studies In Modern approaches to Data Assimilation in Ocean Modeling. Ed P. Malanotte-rizzoli. Elsevier.

Cane M.A. , A. Kaplan,R.N.Miller, B. Tang, E. Hackert, A. Busalacchi. 1996 Mapping tropical Pacific Sea level: data assimilation via a reduced state space Kalman Filter. *J Geophys Res submitted*

Carton J., B.S. Guise, Xianhe Cao, and L Miller 1996 Impact of altimeter, thermistor and expendible bathythermograph data on retrospectic analyses of the tropical Pacific Ocean *J. Geophys. Res.* **101** Vol C6 14147-14160.

Cohn S and Toddling R 1995 Approximate Data Assimilation Schemes for Stable and Unstable Dynamics, submitted to the *J. Met. Soc. Jap.*.

Cooper M and Haines K 1996 Altimetric assimilation with water property conservation, *J. Geophys. Res.* **101**, Vol C1, 1059-1077.

Daley R 1991 *Atmospheric Data Analysis* (Cambridge: Cambridge Univ. Press) p 457.

De Mey P and Robinson A R 1987 Assimilation of altimeter eddy fields in a limited area quasi-geostrophic model *J. Phys. Oceanogr.* **17** 2280-2293.

Dee, D P 1991 Simplification of the Kalman filter for meteorological data assimilation, *Quart. J. Roy. Meteor. Soc.* **117** 365-384.

Derber J C and Rosati A 1989 A global ocean data assimilation system *J. Phys. Oceanogr.* **19** 1333-1347.

Evensen G. and P.J. van Leeuwen (1996) Assimilation of GEOSAT altimeter data for the Agulhas Current using the Ensemble Kalman Filter with a quasigeostrophic model. Mon Weather Rev **124** 85-96.

Ezer T and G.L. Mellor 1994 Continuous assimilation of Geosat altimeter data into a three-dimensional primitive equation Gulf Stream model. *J. Phys. Oceanogr.* **24** 832-847.

Fischer M, Latif M, Flugel M and Ming Ji 1995 On the benefits of sea level assimilation in the tropical Pacific. *Max Planck Inst. fur Meteorologie* Report No 170.

Forbes R M 1996 Initial results from experiments assimilating satellite altimeter sea surface height data into a tropical Pacific ocean model Ocean Applications Technical Note 12 Available from THe Met Office Bracknell.

Fox A and Haines K 1996 Altimeter data assimilation in the OCCAM model Sigma, 21 3-5.

Fu L L, Vázquez J and Perigaud C 1991 Fitting dynamic models to the GEOSAT sea level observations in the tropical Pacific Ocean Part I. A free wave model, *J. Phys. Oceanogr.* **21** 798-809.

Fukumori I. and Malanotte-Rizzoli P. 1995 An approximate Kalman filter for ocean data assimilation; An example with an idealized Gulf stream. jgr **100**, Vol C4, 6777-6793.

Gelb A (ed.) 1974 *Applied optimal estimation*, (Cambridge Mass.: MIT Press) p 374.

Ghil M and Malanotte-Rizzoli P 1991 Data assimilation in meteorology and oceanography, *Adv. Geophys.* **33** 141-266.

Gourdeau L., S. Arnault, Y Menard and J Merle 1992 Geosat sea-level assimilation in a tropical Atlantic model using Kalman Filter. Oceanologica Acta **15** 567-574.

Greiner E. S. Arnault A Morliere 1997 Twelve monthly experiments of 4D Variational assimilation in the tropical Atlantic during 1987. Progress in Oceanography.

Haines K 1991 A direct method of assimilating sea surface height data into ocean models with adjustments to the deep circulation, *J. Phys. Oceanogr.***21**, 843-868.

Haines, K, Malanotte-Rizzoli P, Young R E and Holland W R, 1993 A comparison of two methods for the assimilation of altimeter data into a shallow-water model, *Dyn. Atmos. Oceans***17**, 89-133.

Holland, W.R. and P. Malanotte-Rizzoli, 1989: Along-track assimilation of altimeter data into an ocean circulation model: space versus time resolution studies, *J. Phys. Oceanogr.***19**, 1507-1534.

Hollingsworth A 1989 The role of real-time four-dimensional data assimilation in the quality control, interpretation and synthesis of climate data, in *Oceanic circulation models, combining data and dynamics*, Anderson D L T and Willebrand J, (eds) (The Netherlands: Kluwer Academic Publ.), 304-339.

Hurlburt H E, Fox D N and Metzger E J 1990 Statistical inference of weakly correlated sub-thermocline fields from satellite altimeter data. *J. Geophys. Res.***95**, Vol C7, 11 375-11 409.

Ji M, Leetma A and Derber J 1995 An Ocean Analysis System for Seasonal to Interannual Climate Studies, *Mon. Wea. Rev.***123**, 460,481.

Kleeman R., A.M. Moore and N.R. Smith 1995 Assimilation of subsurface thermal data into an intermediate tropical coupled atmosphere ocean model Mon Weather Rev, **123**, 3103-3113

Knox R A 1989 Ocean Acoustic Tomography: a primer in *Oceanic circulation models, combining data and dynamics*, Anderson D L T and Willebrand J (eds) (Netherlands: Kluwer Academic Publ.), 141-183.

Latif M and Flugel M 1991 An investigation of short-range climate predictability in the tropical Pacific, *J. Geophys. Res.***96** C2 2661-2673.

Le Dimet F X and Talagrand O 1986 Variational algorithms for analysis and assimilation of meteorological observations, *Tellus* **38A**, 97-110.

Levitus S 1982 Climatological atlas of the world ocean, *National Oceanographic and Atmospheric Administration (NOAA) Tech. Paper* **3**, p 173.

Lewis J M and Derber J C 1985 The use of adjoint equations to solve a variational adjustment problem with advective constraints, *Tellus*, **37A** 309-322.

Long R B and ThackerW C 1989a Data assimilation into a numerical equatorial ocean model (Part I): The model and the assimilation algorithm. *Dyn. Atmos. Oceans***13**, 379-412.

Lorenc, A.C., 1981: A global three-dimensional statistical analysis scheme *Mon. Wea. Rev.***109**, 701-721.

Lorenc A C 1986 Analysis methods for numerical weather prediction, *Quart. J. Roy. Meteor. Soc.***112**,1177-1194.

Lorenc A C Bell R S and McPherson B 1991 The Meteorological Office analysis-correction data assimilation scheme, *Quart. J. Roy. Meteor. Soc.***117**, 59-89.

Lorenc A C 1992 Iterative analysis using covariance functions and filters, *Quart. J. Roy. Meteor. Soc.***118** 569-591.

Malanotte-rizzoli P. Editor Modern approaches to Data Assimilation in Ocean Modeling. 1996 Elsevier.

Mariano A J 1990 Contour Analysis: a new approach for melding geophysical fields *J. Atmos Ocean. Tech.* **7** 285-295.

Mellor L G and Ezer T 1991 A Gulf Stream model and an altimetry assimilation scheme, *J. Geophys. Res.***96** C5 8779-8795.

Menemelis D., Webb T., Wunsch C., Send U., Hill C. 1997 Basin-scale ocean circulation from combined altimetric, tomographic and model data. Nature 385 618-619.

Miller R N 1986 Toward the application of the Kalman filter to regional open ocean modelling. *J. Phys. Oceanogr.***16**, 72-86.

Miller R N 1989 Direct assimilation of altimetric differences using the Kalman filter, *Dyn. Atmos. Oceans***13**, 317-334.

Miller R N and Cane M, 1989 A Kalman filter analysis of sea level height in the tropical Pacific, *J. Phys. Oceanogr.***19**, 773-790.

Moore A M and Anderson D L T 1989 The assimilation of XBT data onto a layer model of the tropical Pacific Ocean, *Dyn. Atmos. Oceans***13**, Vol 3-4, 441-464.

Moore A M Cooper N S and Anderson D L T 1987 Data assimilation in models of the Indian Ocean, *J. Phys. Oceanogr.*, **17**, 1965-1977.

Moore A M 1989 Aspects of geostrophic Adjustment during tropical ocean data assimilation, *J. Phys. Oceanogr.***19**, 435-461.

Moore A M 1990 Linear equatorial wave mode initialisation in a model of the tropical Pacific: An initialisation scheme for tropical ocean models, *J. Phys. Oceanogr.***20**, 423-445.

Moore A M 1991 Data assimilation in a quasigeostrophic open ocean model of the Gulf Stream region using the adjoint method. *J. Phys. Oceanogr.***21** 398-427.

Oschlies A and J Willebrand 1996 Assimilation of Geosat altimeter data into an eddy-resolving primitive equation model of the North Atlantic Ocean *J. Geophys. Res.***101 C6** 14175-14190

Palmer T N and Anderson D L T 1994 The prospects of seasonal forecasting: a review paper *Quart. J. Roy. Meteor. Soc.***120** 755-793.

Pinardi N, Miyakoda K, Rosati A and Gudgel R G 1994 A global ocean assimilation system for hydrographic data and satellite altimeter data, *Oceanology International 1994, The Global Ocean. Conference Proceedings (Brighton)*, Vol 1, Brighton 1994.

Robinson A R, Spall M A, Walstad L J and Leslie L G, 1989 Data assimilation and dynamical interpolation in gulfcast experiments, *Dyn. Atmos. Oceans***13**, 3-4, 269-300, 301-316.

Rosati A., R. Gudgel and K. Miyakoda 1995 Decadal analyses produced from an ocean data assimilation system.Monthly Weather Rev **123** 2206-2228.

Sheinbaum J 1989 Assimilation of oceanographic data in numerical models. *D.Phil. Thesis*, Oxford UK.

Sheinbaum J and Anderson D L T 1990a: Variational assimilation of XBT data, Part I, *J. Phys. Oceanogr.***20** 672-888.

Sheinbaum J and Anderson D L T 1990b: Variational assimilation of XBT data, Part II: Sensitivity studies and use of smoothing constraints, *J. Phys. Oceanogr.***20** 689-704.

Sheinbaum J 1995a Variational assimilation of simulated acoustic tomography data and point observations: A comparative Study, *J. Geophys. Res.***100** Vol C10 20745-20761.

Smith N. R., J.E. Blomley and G. Meyers 1991 A univariate statistical interpolation scheme for subsurface thermal analyses in the tropical oceans.*Prog. in Oceanography*, **28**, 219-256.

Smith N. R. 1995 An improved system for tropical ocean subsurface temperature analyses J. Atmos Ocean Tech **12** 850-870.

Stoffelen A. and D.L.T. Anderson 1997 Ambiguity removal and assimilation of scatterometer data Quart J Royal Met Soc **123** 491-518.

Tarantola A 1987 *Inverse problem theory-methods for data fitting and model parameter estimation*, (The Netherlands: Elsevier).

Thacker W C 1988: Fitting models to data by enforcing spatial and temporal smoothness, *J. Geophys.*

Res. **93**, 655-665.
Thacker W C and Long R B 1988: Fitting dynamics to data, *J. Geophys. Res.* **93**, 1227-1240.
Weaver A.T. and D.L.T. Anderson 1997 Variational data assimilation in numerical models of the ocean. J Physical Oceanography In press.