7

Applications to time-dependent fluid problems

So-called "inverse methods," a terminology sometimes restricted to steady-state situations, and "state estimation," are really one and the same. Either label can defensibly be applied to both time-dependent or steady physical problems, with the choice among the differing methods being mainly dictated by problem size and convenience. Thus, in principle, observations and their uncertainties could be combined, e.g., with a governing discretized partial differential equation such as the familiar advection/diffusion equation,

$$\frac{\partial C}{\partial t} + \mathbf{v} \cdot \boldsymbol{\nabla} \mathbf{C} - \nabla \left(\mathbf{K} \boldsymbol{\nabla} C \right) = m_C, \tag{7.1}$$

with known $\mathbf{v}$, $\mathbf{K}$, plus boundary/initial condition information, some estimate of the accuracy of the difference equation, the entire set put into the canonical form, and the collection solved for C, and any other imperfectly known elements. The principle is clear, but for large spatial domains and long time intervals, the practicalities of calculating, storing, and using the enormous resulting matrices drive one, even for completely linear problems, toward numerical algorithms that require less storage or computer time, or do not burden one with the possibility of a complete resolution analysis, or all of these things.

The methods described in this book for solving time-dependent least-squares, or inverse problems, or state estimation problems, have focussed on so-called sequential methods (Kalman filter plus the RTS smoother algorithm), and so-called whole-domain or iterative methods involving Lagrange multipliers (adjoint method). These approaches are not exhaustive; in particular, Monte Carlo methods (genetic algorithms, simulated annealing, etc.) hold considerable promise. If we restrict ourselves to the methods we have treated, however, it can be recognized that they are widely discussed and used in the general context of "control theory" with both pure mathematical and highly practical engineering limits. What distinguishes large-scale fluid problems from those encountered in most of the control literature is primarily

the enormous dimensionality of the fluid problem (the "curse of dimensionality," to use R. Bellman's colorful phrase). That one can, in principle, find minima or stationary points of a convex objective function, is hardly in doubt (enough computer power overcomes many numerical difficulties). At the time the first edition of this book was written, the practical application of the ideas in Chapters 3 and 4 to time-dependent geophysical flow problems was primarily one of conjecture: a few investigators, with the computers available then, set out to explore the problems that would arise. Their results suggested that the methods could be used on problems of practical concern. But few results, other than symbolic ones, were available. In the intervening decade, however, computer power continued its inexorable growth, better numerical methods were developed, models improved, and vast new numbers of observations appeared, to the extent that we can now focus on describing actual, scientifically useful results, rather than speculating on the possibility of their existence.

Important global fluid inverse problems, however, nonetheless still outstrip the largest existing, or even foreseeable, computers. Approximations to the estimation algorithms thus remain extremely important, and will be so indefinitely. After examining some of the actual state estimates made, we will describe some of the more important methods being employed to reduce computational load. So-called identical twin results are not discussed here. Numerical experiments with artificial, often noise-free or with artificially simplified noise, data are a necessary step toward testing any estimation method. But as described in much of the published literature, they become an end in themselves, reducing to uninteresting tests of least-squares, a method whose efficacy is hardly in doubt.

7.1 Time-dependent tracers

A two-dimensional version of Eq. (6.1) was used by Kelly (1989) to determine the flow field, and is a regional, comparatively simple example. She expanded satellite measurements of surface temperature T $(\mathbf{r}, t)$ in a set of basis functions (sines and cosines) and minimized the data misfit in time and space, solving for the horizontal flow $\mathbf{v}$. An example of her results is shown in Fig. 7.1 for a region off the coast of California. The solution nullspace is important, and Kelly (1989) discusses the reliability of the estimates.

Observed transient tracers and the extent to which they could be used for determining the flow field were described by Mémery and Wunsch (1990). A very general formalism was described by Haine and Hall (2002), but the method is so demanding of data that the application appears primarily to diagnostics in general circulation models – where 100% "data" coverage is available.

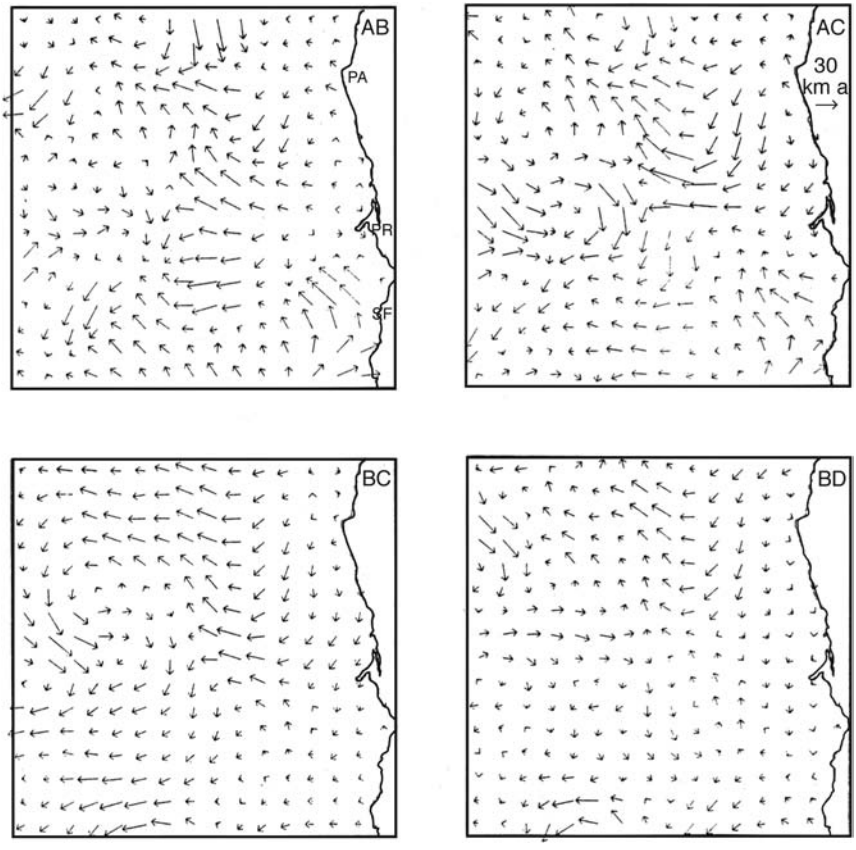

Figure 7.1 Flow fields inferred from the difference of sea surface temperatures at four times, subject to Eq. (7.1). The region depicted is off San Francisco, California. (Source: Kelly, 1989)

7.2 Global ocean states by Lagrange multiplier methods

As has been discussed in Chapter 4, the computational advantage of the Lagrange multiplier methods, as compared to the sequential ones, is the absence of a requirement for producing the error covariances for the state and control vectors as part of the computation. Because the Lagrange multiplier method finds the stationary value of the objective function, J, iteratively, over the entire time-span of observations, no average of two independent state estimates is required. Without the covariances, one can greatly reduce both the computational load and storage requirements (the storage requirement for the RTS smoother is formidable in large problems). Of course, as pointed out earlier, a solution without error covariances is likely to be

of limited utility; nonetheless, we can postpone confronting the problem of finding them.

As an example of what is now possible, again focus on the ocean, and, in particular, the problem of determining the three-dimensional, time-varying circulation of the ocean over various durations in the period 1993–2002 (Stammer *et al.*, 2002, 2003, 2004). The model, thought to realistically describe the oceanic flow, is based upon the Navier–Stokes equations for a thin spherical shell using a fully realistic continental geometry and topography. For those interested, the governing system of underlying numerical equations are discussed in textbooks (e.g., Haidvogel and Beckmann, 1999; Griffies, 2004). It suffices here to define the state vector on a model grid (i, j, k) at time $t = n\Delta t$. The state vector consists of three components of velocity, $(u(i, j, k, t), v(i, j, k, t), w(i, j, k, t))$, a hydrostatic pressure $(p\,(i, j, t))$, temperature $(T\,(i, j, k, t))$, and salinity $(S\,(i, j, k, t))$ (some of the variables are defined at positions displaced by $1/2$ of the grid spacing from the others, a distinction we ignore for notational tidiness). A numerical approximation is written (many thousands of lines of Fortran code) to time-step this model approximation, which we abbreviate in the standard form,

$$\mathbf{x}(t) = \mathcal{L}(\mathbf{x}(t-1), \mathbf{Bq}(t-1), \mathbf{\Gamma u}(t-1), \mathbf{p}, t-1, \mathbf{r}), \tag{7.2}$$

with t discrete, and taking $\Delta t = 1$. Here, as in Chapter 4, $\mathbf{q}\,(t)$ are the known boundary conditions, sources/sinks (if any), and the $\mathbf{u}\,(t)$ are the unknown boundary conditions, sources/sinks (controls). The $\mathbf{u}\,(t)$ generally represent adjustments to the boundary conditions (taken up momentarily), as well as, e.g., corrections to the estimated initial conditions $\tilde{\mathbf{x}}(0)$, and model errors. $\mathbf{r}$ is a three-dimensional position vector $(i\Delta x, j\Delta y, k\Delta z)$. Δx, Δy, Δz are the spatial grid increments (not to be confused with the state or observation vectors, $\mathbf{x}$, $\mathbf{y}$). The grid increments do not need to be constant in the domain, but again for notational simplicity we will ignore that detail. $\mathbf{p}$ is a vector of internal model parameters, also regarded as unknown, and which could be included in $\mathbf{u}$. These are adjustments to, e.g., empirical friction or diffusion coefficients (eddy-coefficients).

In conventional ocean (forward) modeling, the system is driven by atmospheric winds, and estimated exchanges of heat (enthalpy) and fresh water between the ocean and atmosphere. These appear in the model as numerical approximations, for example, the wind stress might be imposed as a boundary condition,

$$A_v \left[\frac{\Delta u(i, j, k=0, t)}{\Delta z}, \frac{\Delta v(i, j, k=0, t)}{\Delta z} \right] \Bigg|_{z=0} = [\tau_x(i, j, t), \tau_y(i, j, t)].$$

Here $\boldsymbol{\tau} = (\tau_x, \tau_y)$ is the vector wind stress evaluated nominally at the sea surface $(z = k\Delta z = 0)$, and A_v is an empirical coefficient. $\boldsymbol{\tau}$ and the related fluxes of heat

and moisture are taken as given, perfectly. Some initial conditions $\tilde{\mathbf{x}}(0)$ (perhaps, but not necessarily, a state of rest) are used, and Eq. (7.2) is time-stepped forward subject to the time-varying meteorological fields. In the most realistic simulations, wind and air – sea exchanges are taken from meteorological analyses on global grids and available, typically, two or four times per day and then linearly interpolated to the model time-step. As formulated, the resulting problem is well-posed in the sense that enough information is available to carry out the time-stepping of Eq. (7.2). The system can be unstable – either because of numerical approximations or, more interestingly, because the fluid flow is physically unstable. Many such simulations exist in the published oceanographic literature (and in the even larger analogous meteorological literature).

From the oceanographer's point of view, the simulation is only the beginning: one anticipates, and easily confirms, that the resulting model state does not agree (sometimes well, sometimes terribly), with observations of the actual ocean. For someone trying to understand the ocean, the problem then becomes one of (a) understanding why the model state disagrees with what was observed, and (b) improving the model state so that, within error estimates, the model comes to consistency with the observations and one can begin to understand the physics. In the spirit of this book, the latter would then be regarded as the best estimate one could make of the time-evolving ocean state. With such a best estimate one can, as in the previous chapter, calculate property fluxes (heat, carbon, etc.) and analyze the model in detail to understand why the system behaves as it does.

The specific example we discuss is based upon Stammer *et al.* (2002, 2003, 2004), but displaying later estimates, believed to be improved over those previously published. State estimation is open-ended, as models can always be improved, and error covariances made more accurate. The specific examples are based upon a numerical representation of Eq. (7.2) written in a finite difference form on a grid of 2° in both latitude and longitude, and with 23 layers in the vertical. The total number of grid points (the number of equations in Eq. 7.2) was thus approximately 360 000. With three components of velocity, pressure, temperature, and salinity at each grid point making up the state vector, the dimension of $\mathbf{x}(t)$ is approximately 5.3 million. The model was time-stepped at intervals $\Delta t = 1$ hour. The total state vector in the estimates of Stammer *et al.* (2002, 2003, 2004), $[\mathbf{x}(0)^{\mathrm{T}}, \mathbf{x}(1)^{\mathrm{T}}, \ldots, \mathbf{x}(t_f)^{\mathrm{T}}]^{\mathrm{T}}$, had dimensions over a decade of about 10^{11} elements. An error covariance matrix for such a result would be a forbidding object to compute, store, or understand. (In the interim, state estimates of even larger dimension have been carried out – see Köhl *et al.*, 2006; Wunsch and Heimbach, 2006.)

A great variety of oceanic observations were employed. A proper discussion of these data would take us far afield into oceanography. In highly summary form, the data consisted of, among others:

1. Sea surface elevation observations from altimetry (approximately one/second at a succession of points) over the entire period. These were employed as two separate data sets, one defined by the time mean at each point over the time interval, and the second being the deviation from that mean. (The reason for the separation concerns the error covariances, taken up below.)
2. Observations of sea surface temperature, composited as monthly means, over the entire ocean.
3. Time average climatologies of temperature and salinity at each grid point, $i,\ j, k$. That is, the computational-duration temperature, etc., is constrained in its average to observations, as in

$$\sum_j E_{ij} \left[\frac{1}{t_f + 1} \sum_{t=0}^{t_f} x_j (t) \right] + n_i = y_i, \tag{7.3}$$

   where E_{ij} selects the variables and grid points being averaged.
4. "Synoptic" temperatures and salinities from shipboard instantaneous observations.
5. Various constraints from observed mass transports of major currents.
6. Model-trend suppression by requiring temperatures and salinities at the beginning, $t = 0$, and at the end, $t = t_f$, to be within an acceptable range (where the meaning of "acceptable" requires study of observed large-scale oceanic temperature changes).

The known controls, $\mathbf{q}(t)$, were also based upon observations, and separating them from state vector observations is an arbitrary choice. They included:

1. twice-daily estimates of gridded vector wind stress, $\boldsymbol{\tau}$;
2. daily estimates of gridded air–sea enthalpy (heat) exchange, H_E;
3. daily estimates of air–sea moisture exchange, H_f.

The control vector $\mathbf{u}\,(t)$ included:

1. adjustments to the vector wind, $\Delta\boldsymbol{\tau}$;
2. adjustments to the enthalpy flux, ΔH_E;
3. adjustments to the moisture flux, ΔH_f.

The control matrix, $\boldsymbol{\Gamma}$, vanished everywhere except at the sea surface, where it was unity at each boundary grid point. That is, each of the controls $\Delta\boldsymbol{\tau}$, etc., could vary independently at each surface grid point. Adjustments to the initial conditions, $\tilde{\mathbf{x}}(0)$, can be regarded as part of the controls; here we retain them as part of the state vector.

A count of the observations produces approximately 6.3×10^6 equations of the form $\mathbf{E}\,(t)\,\mathbf{x}\,(t) + \mathbf{n}\,(t) = \mathbf{y}\,(t)$. These were used to form an objective function, J, as in Eq. (4.41). The control terms, $\mathbf{u}\,(t)\,\mathbf{Q}\,(t)^{-1}\,\mathbf{u}\,(t)$, provide another 8.1×10^7 elements of J.

Weights must be specified for every term of the objective function (that is, for every term of Eq. (4.41), including expressions such as Eq. (7.3)). Unsurprisingly, obtaining realistic values is a major undertaking. In practice, all such matrices

were diagonal, with one exception, with information about the covariance of the observational errors (including that of the control terms) being unavailable.

J should include additional control-like terms representing the model error. (Errors in the wind-field and other boundary conditions can be labeled arbitrarily as being part of the model or data or controls; here we generally regard them as control variables.) But no conventional fluid-flow model has ever been produced with a statement of the quantitative skill expected in reproducing a quantity of interest: for example, the skill with which the temperature at point $i, j, k, n\Delta t$ would be determined if the boundary and initial conditions were perfect. Deviations from perfection would arise from limited spatial and temporal resolution, mis-parameterized process, missing physics, incorrect bottom topography, etc. Such errors are unlikely to be Gaussian, and are in general not known. But they ought to be included in J, J'. In practice, what was done was to permit the model to misfit the data in some regions by much more than the formal data error would allow. Such regions include, for example, the equator, and the areas of the intense boundary currents, where model resolution is expected, a priori, to be inadequate to reproduce what is observed. One is redefining, empirically, the data error to include structures the model cannot reproduce. This situation is not very satisfactory, and a major goal has to be to find ways of formally specifying model error. For the time being, one must regard the solution as being one of least-squares – curve-fitting, rather than as a true minimum variance solution.

The very large number of terms in J, and of the state vector dimension, suggest the practical difficulties facing anyone attempting rigorous state estimation with global-scale fluids. Fortunately, modern computing power is rising to the challenge.

These large dimensions, however, dictate the use of Lagrange multipliers in order to avoid the need to compute even approximate error covariances. Formally then, the model (7.2) was appended to J as in Eq. (4.97) to form a new objective function, J'. The dimension of the Lagrange multipliers $\boldsymbol{\mu}(t)$ is equal to the number of model equations over the whole spatial domain, or 2×10^6 at each time-step or 2×10^{10} over the whole time domain. To find the stationary point of J' (at $\tilde{\mathbf{x}}(t), \tilde{\mathbf{u}}(t), \tilde{\mu}(t)$) the iterative linearized search process described on p. 241, was used. Note, in particular, that the $(\partial \mathbf{L}/\mathbf{x}(t))^{\mathrm{T}}$ matrix used in the equivalent normal equations corresponding to the stationary point ((4.168) is not available. It was instead computed implicitly ("on the fly") as described in Chapter 4, using an automatic differentiation tool. The general procedure employed to determine the best-estimate state followed the iterative scheme outlined there, iterating through a series of trial solutions, $\tilde{\mathbf{x}}^{(i)}(t)$, $\tilde{\boldsymbol{\mu}}^{(i)}(t)$, $\tilde{\mathbf{u}}^{(i)}(t)$, until convergence was obtained (more precisely, until the misfit measured in J was deemed acceptable.) In particular, $\tilde{\boldsymbol{\mu}}^{(i)}(t)$ was used to define the directions of the line search to minimize J. Hundreds of iterations were required before J reached an acceptable level, and its individual

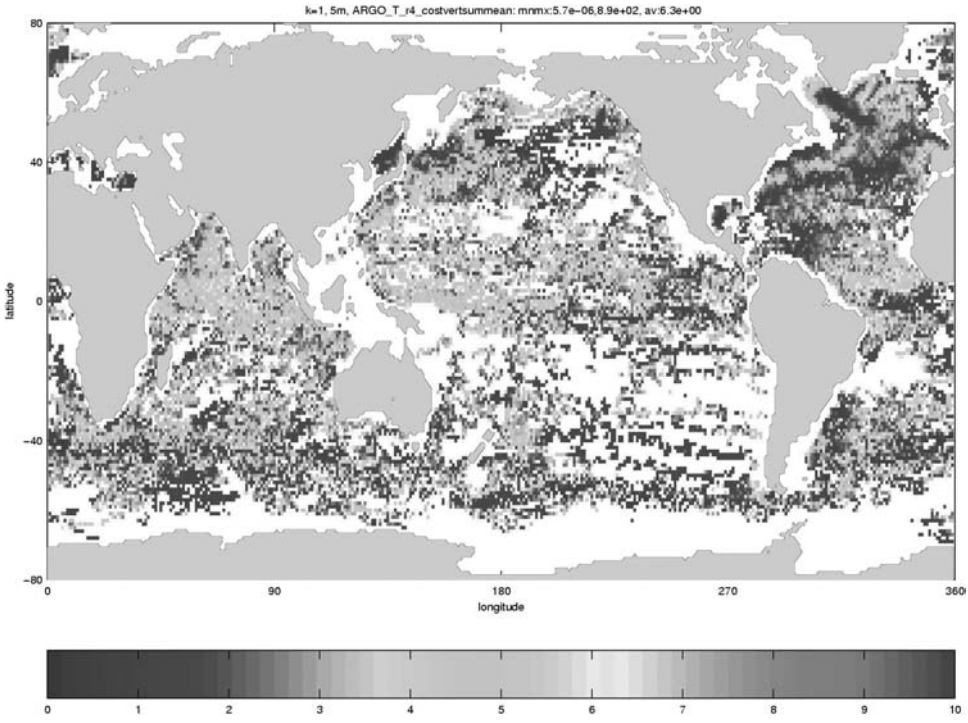

Figure 7.2 Misfit to so-called ARGO float temperature profiles. These instruments produce a vertical profile of temperature in the ocean above about 2000 m at predetermined intervals (of order 10 days). The misfits shown here are weighted by the estimated errors in both model and data, and then averaged over the entire water column. (See color figs.)

terms were sufficiently close to white noise that the solution appeared to make sense. Non-linear optimization has to be regarded as in part an art – as descent algorithms can stall for many numerical reasons or otherwise fail.

Only a few representative results of these computations are displayed here. Much of the information about the results, and the focus of the scientific effort, once the model has been deemed acceptable, lies with the model-data misfits. These misfits reflect the estimated errors in the data and in the ability of the model to reproduce the data – if the latter were perfect. An example is shown in Fig. 7.2 for vertical profiles of temperature over a period of about eight years ending in 2004. The figure is not untypical, showing near-global coverage (most of these data, however, exist only towards the end of the 13-year period of estimation), which is a generally acceptable order of magnitude, but also patterns of larger misfit that ultimately have to be understood. This calculation is ongoing, and some of the misfits are likely to disappear as the optimization proceeds.

Figures 7.3 and 7.4 display the time average of part of the control vector, $\langle \tilde{\mathbf{u}}(t) \rangle$, as determined by the optimization at one stage (Stammer *et al.*, 2002). A

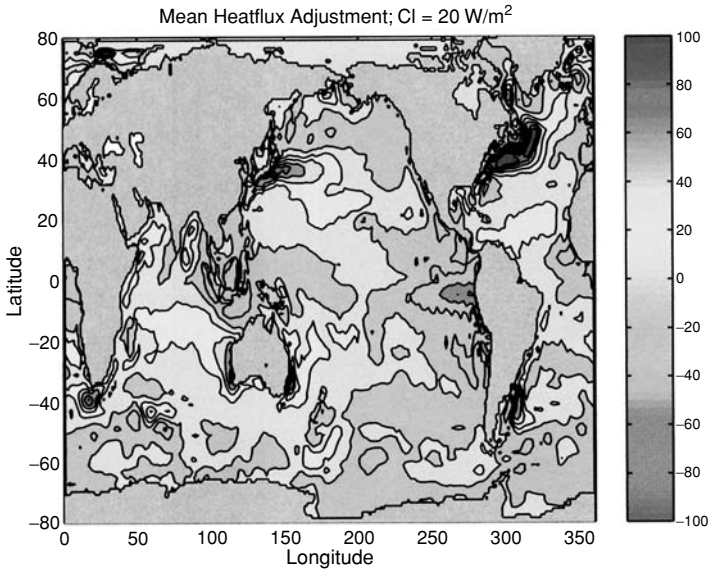

Figure 7.3 Mean changes in (a) net surface heat exchange (W/m^2), determined from six years of meteorological estimates and an ocean model. (See color figs.) (Source: Stammer *et al*., 2002)

complicated spatial pattern emerges – one that is rationalizable in terms of known problems with the meteorological analyses that produced the "known" controls $\mathbf{q}(t)$ to which these are adjustments. None of the values displayed violates general understanding of the accuracy of the meteorological fields. On the other hand, the comments made above about model errors have to be kept in mind, and the very large values of $\langle \tilde{\mathbf{u}}(t) \rangle$ inferred in some of those regions should be viewed with suspicion.

Some of the flavor of the state vector can be seen in Fig. 7.5, where mean velocities $\langle \tilde{u}(t) \rangle$, $\langle \tilde{v}(t) \rangle$ are displayed for one model depth ($\tilde{u}(t)$ is the scalar velocity component, not the control vector). The reader is reminded that the spatial resolution of this model is 2° of latitude and longitude, and physical arguments exist that it would need to approach $1/16^\circ$ over much of the ocean to be adequately realistic. Dimension remains a major challenge. Figure 7.6 gives another view of the time average flow field, in the vertical dimension, across 26° N in the North Atlantic, from a later calculation with 1° spatial resolution.

The full flavor of model time dependence is best seen in animations of the results that are available on the web. To show something of the great temporal variability in the model, Fig. 7.7 displays the meridional flux of heat across 21° S in the South Atlantic, showing values that can actually reverse (to the south), and vary very rapidly from day-to-day.

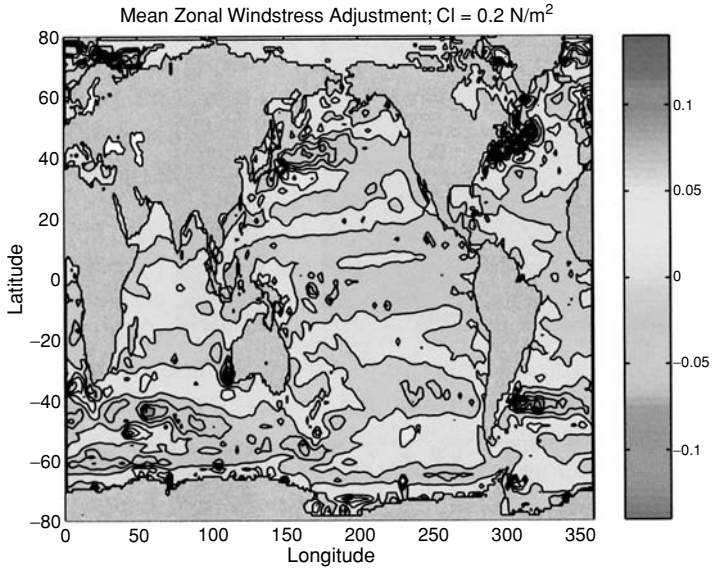

Figure 7.4 Adjusted wind stress (part of the control vector) in N/m^2. (See color figs.) (Source: Stammer *et al.*, 2002)

Because the model calculates the meridional transport of temperature at every grid point at every time-step, it can be integrated zonally and in time to produce the curves displayed in Fig. 7.8. These are compared directly with the values inferred from the static calculation described in Chapter 6. An explicit error bar is available for the latter, as shown, but not for the former. These two types of calculation are clearly two sides of the same coin – one taking explicit account of the temporal variation, the other treating the system as momentarily static.

In the immediate future, one can anticipate exploitation and extension of this type of state estimation, as the approximations made in building numerical models come under increasing scrutiny and question. Numerical models representing any of the important physical partial differential equation systems of physics, chemistry, and biology (e.g., Schrödinger, Navier–Stokes, Maxwell, etc.) are necessarily an approximation. Generally speaking, the modeler makes a whole series of choices about simplified representation of both kinematics and dynamics. Some of these simplifications are so commonplace that they are not normally regarded as parameterizations. In the Navier–Stokes equation example, the geometry of the container is commonly smoothed to a model grid spacing and little further thought given to it. But it is clear that flow fields can be greatly influenced by exactly where the boundary is placed and represented (the classical example is the problem of making the Gulf Stream leave the coastline at Cape Hatteras – if done incorrectly, the Gulf Stream will follow the wrong pathway, with disastrous results for heat and other

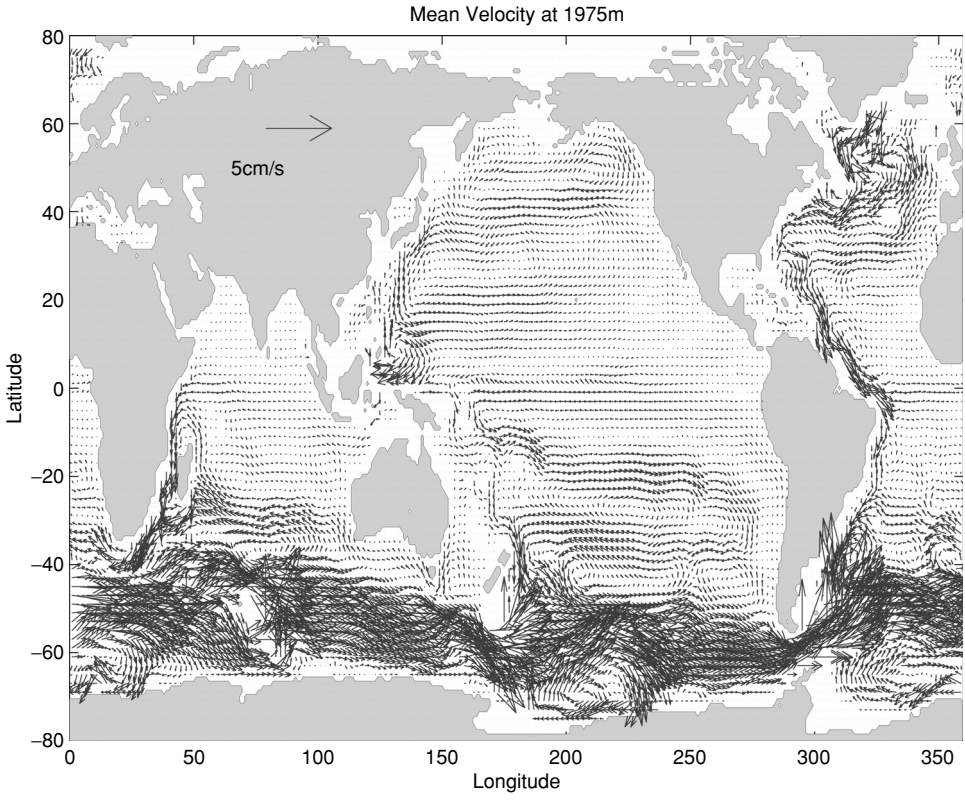

Figure 7.5 Estimated mean velocity field at 1975 m depth, in cm/s, from a six-year estimation period. (Source: Stammer *et al.*, 2002)

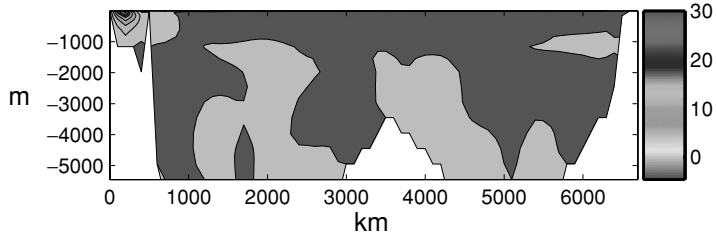

Figure 7.6 Twelve-year mean velocity across the North Atlantic Ocean (cm/s) at 26° N from a 13-year optimization of a 1 degree horizontal resolution (23 vertical layers) general circulation model and a large global data set. Red region is moving southward, remaining regions are all moving northward. Note the absence (recall Chapter 6) of any obvious level-of-no-motion, although there is a region of weak mean meridional flow near 1100 m depth. The northward-flowing Gulf Stream is visible on the extreme western side. (See color figs.)

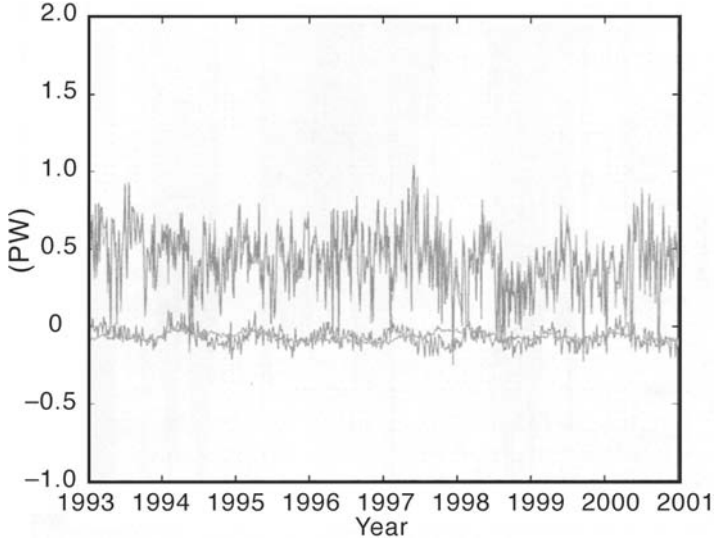

Figure 7.7 Time dependence in components of the meridional flux of heat across 21° S, in this case from eight years of analysis. The different colors correspond to different spatial integrals of the heat budget. (See color figs.) (Source: Stammer *et al*., 2003, Fig. 14)

budgets). The estimation machinery developed here is readily extended to include, e.g., bottom topography in a model (Losch and Heimbach, 2005), and, presumably, such other arbitrary representations as slip/no-slip/hyper-slip boundary conditions that have large elements of ad hoc approximation buried in them.

7.3 Global ocean states by sequential methods

The solution to state estimation problems by filter/smoother methods is highly desirable because one obtains both the estimated state and controls, but also their complete error covariances. In a fluid model, having perhaps 10^7 state elements at each time-step, and a similar order of magnitude of control vector elements, the construction, e.g., in a Kalman filter step, of the error covariance of the state alone requires the equivalent of running the model $10^7 + 1$ times at each time-step. A time-reverse step to, e.g., implement the RTS smoother, involves a similar calculation. This computational load is incurred even if the system is entirely linear. Non-linear systems will usually require many more computations.

With modern computers, dealing with problems with dimensions of thousands or even hundreds of thousands may be feasible. But global scale fluid problems, if tackled even remotely rigorously, are beyond our current capability. Yet the goal of a rigorous (for linear systems) implementation of the sequential algorithms is so

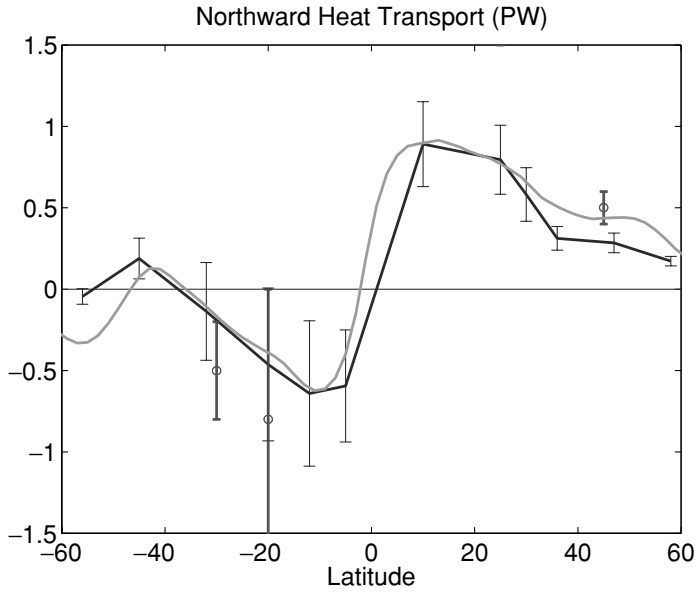

Figure 7.8 The blue curve shows the time and zonally integrated global meridional flux of heat (enthalpy) in the constrained model. The green curve shows the same field, estimated instead by integrating the fluxes through the surface. The two curves differ because heat entering from the surface can be stored rather than necessarily simply transported. Red bars are from the calculation of Ganachaud and Wunsch (2000) described in Chapter 6. Blue bars are the temporal standard deviation of the model heat flux and are an incomplete rendering of the uncertainty of the model result. (See color figs.) (Source: Stammer *et al*., 2003)

attractive that one is led to explore useful approximations to the full filter/smoother algorithms.

Consider the calculations of Fukumori *et al*. (1999), who used the same numerical model as in the Lagrange multiplier method calculation, but employing the reduced state approximation described in Chapter 5, p. 267. The grid was spaced at 2° of longitude and 1° of latitude, with 12 vertical layers, and again with the six elements of velocity, pressure, temperature, and salinity for an approximate state vector dimension at each time-step of 3×10^5 over a total time interval of three years, 1992–5. This state vector defines the fine-scale model. Apart from the data used to define the initial temperature and salinity conditions, the only observations were the time-varying sea surface pressure averaged over 2.5° sub-satellite arcs.

The general procedure was to both reduce the state vector dimension and to use the Kalman filter/RTS smoother algorithms in their steady-state asymptotes. They reduced the modeled vertical structure to two degrees of freedom by employing a truncated vertical normal mode representation. Dominant horizontal scales of

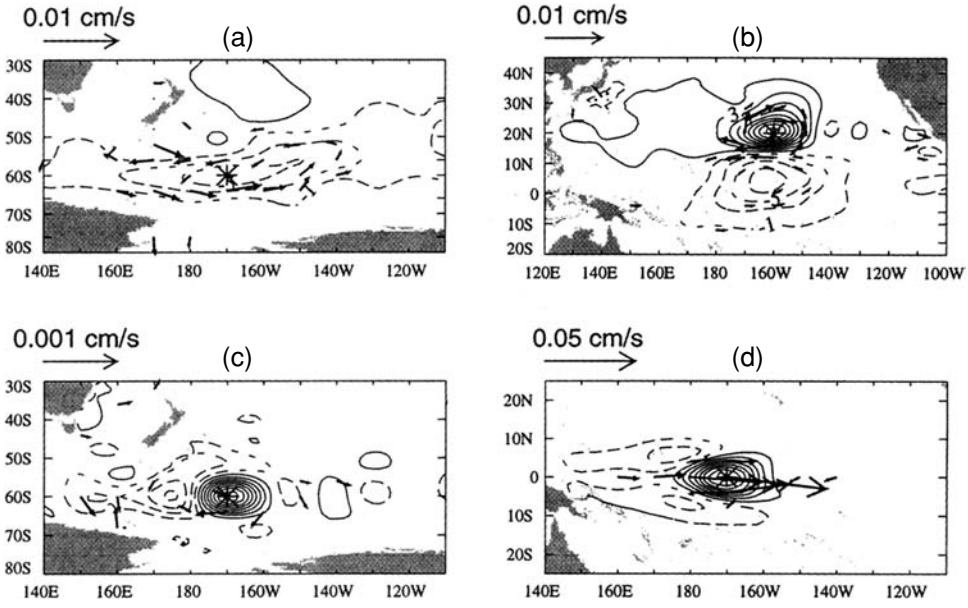

Figure 7.9 Horizontal structure imposed by the approximate Kalman filter corresponding to a 1 cm difference in predicted and observed sea levels at the position of the asterisk: (a) mass transport, (b) stream function, (c) temperature at 1700 m, and (d) temperature at 175 m. (Source: Fukumori *et al*., 1999)

model variability were identified by running the fine-scale model in a forward computation. The resulting fields were reconstructed using the Eckart–Young–Mirsky representation in the singular vectors, and truncating at $K = 300$, accounting for 99% of the correlation structure of the model (not of the variance – the singular value decomposition was applied to the correlation rather than the covariance matrix). A study of the spatial structure of this reduced representation led to a coarse-scale model defined on a grid that was 10° in longitude, 5° in latitude. The coarse-to-fine-scale transformation $\mathbf{D}^+$ was carried out using simple objective (but sub-optimal) mapping using a fixed spatial covariance function dependent only upon grid-point separation (neither orientation nor proximity to boundaries was accounted for) and $\mathbf{D}$ was the pseudo-inverse of $\mathbf{D}^+$ (rather than as we defined them in Chapter 5). Special methods were used to assure that lateral boundary conditions were satisfied after a field was mapped from the fine to the coarse grid.

The Kalman gain, computed from the asymptotic steady-state of $\mathbf{P}(t)$, is shown in Fig. 7.9 for temperature elements of the reduced state vector, and for the stream function at different depths.

The effectiveness of the system used is tested in part by its ability to reproduce observations not included as constraints. An example for this computation is shown

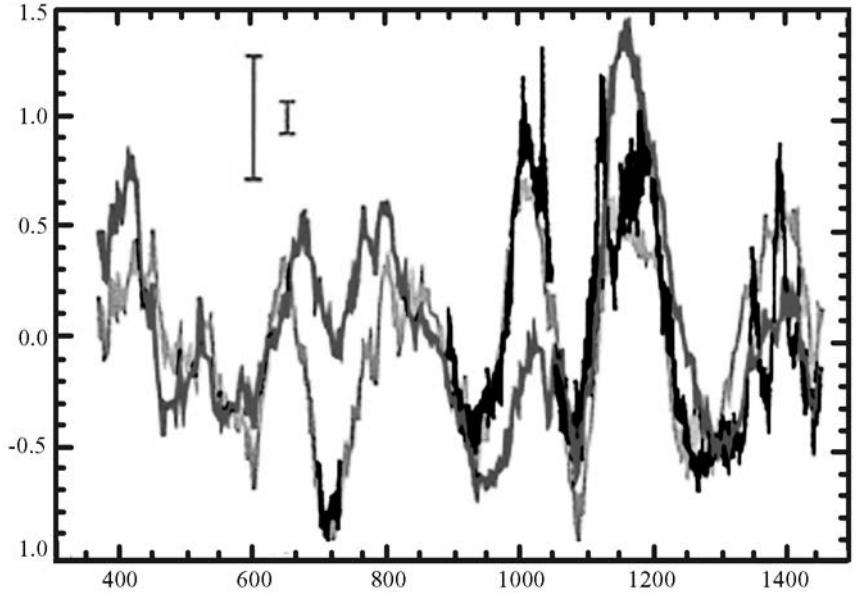

Figure 7.10 Comparison of temperature at 200 m depth at 8° N, 180° E of data (black), unconstrained model (red) and model constrained using altimetric data (blue). See Fukumori *et al.* (1999) for details and further comparisons. An approximate filter/smoother combination was used. (See color figs.) (Source: Fukumori *et al.*, 1999)

in Fig. 7.10 for temperature variations on the equator. Error bars shown represent the estimated value from the asymptotic RTS smoother.

7.4 Miscellaneous approximations and applications

Tomographic applications in the ocean have been discussed in detail by Munk *et al.* (1995). A number of interesting applications of tomographic inversions have been carried out in the years since then (e.g., ATOC Consortium, 1998; Worcester *et al.*, 1999; Worcester and Spindel, 2005). In tomography, very large-scale coverage of the ocean can be obtained almost instantaneously (at the speed of acoustic waves), and hence most calculations have been in the context of an ocean state that is momentarily static.

A very large number of approximate estimates of the ocean using highly reduced forms of Kalman filter has been published. One example is described by Carton *et al.* (2000). In many applications, the need for speed and efficiency outweighs the need for dynamical or statistical rigor. It is not possible to prescribe any method that is uniformly applicable to all problems.

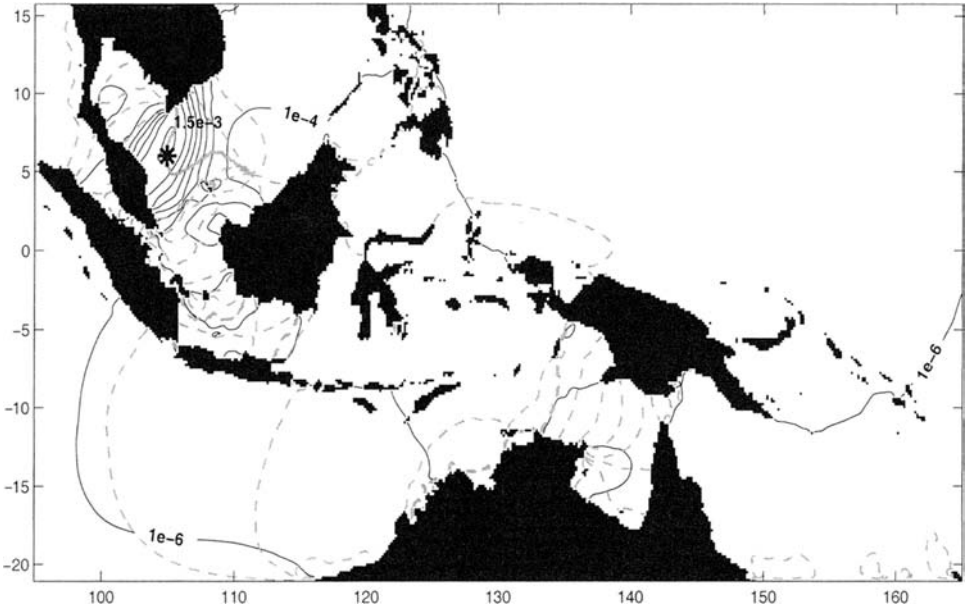

Figure 7.11 Amplitude and phase for a tide modeling representer (Green function), as calculated by Egbert and Erofeeva (2002). Solid contours are the amplitude of influence of a data point at the asterisk. Phases (dashed) correspond to tidal propagation phase changes relative to that point. (Source: Egbert and Erofeeva, 2002)

The mathematical machinery of state estimation has many other applications. Examples are the study of model sensitivity to parameter or other perturbations (e.g., Galanti and Tziperman, 2003; Hill *et al*., 2004; Li and Wunsch, 2004) by using the adjoint solution directly. Farrell and Moore (1993) have used the SVD of the state transition matrix, **A**, to find the most rapidly growing perturbations in the atmosphere. Marchal (2005) applied the RTS smoother to estimate the variations in atmospheric carbon-14 over the last 10 000 years. The list of applications is now too long to be discussed here.

The representer form of inversion (Chapter 5) has had considerable use, particularly in models that calculate tides and employ altimetric sealevel measurements as constraints. See Egbert and Erofeeva (2002) for a representative discussion. One of their representers is depicted in Fig. 7.11. The methodology has been used in several global calculations.

We have seen the close connection between Lagrange multiplier methods and Green functions. Menemenlis *et al*. (2005) apply Green functions to a variety of parameter estimation problems with a global GCM. Fukumori (2001) discusses this connection, as well as published applications to observability and controllability in ocean models.

7.5 Meteorological applications

This subject was touched on in Chapter 5. Numerical weather prediction (NWP) represents the most highly developed, in a numerical engineering sense, of the methods discussed here, and the terminology "data assimilation" originated there. The main focus of this book has, however, been deliberately placed elsewhere because of the highly specialized nature of the weather forecasting problem. Governmental, aviation, military, agricultural, and wide public interests demand predictions over time spans ranging from minutes (in tornado-prone regions) to a few days, and on a continuous basis. To the extent that one is able to make useful forecasts on any of these timescales, the question of whether either the model or the methodology is completely understood scientifically is secondary. That methods such as extended Kalman filters represent highly desirable goals for weather forecasting has long been known (e.g., Ghil *et al*., 1981), but NWP has faced such enormous data sets and complex models, that rigorous implementation of the Kalman filter has remained beyond the realm of practice. The smoothing problem has received scant meteorological attention.

A very large number of approximate methods have been used in NWP, with its own textbooks (Daley, 1991; Kalnay, 2003), and review articles continue to appear (e.g., Lorenc, 1986; Ghil and Malanotte-Rizzoli, 1991; Talagrand, 1997). A major issue for anyone attempting to make use of this important literature is the proliferation of ad hoc terminology in the meteorological application. Thus what this book calls the adjoint or method of Lagrange multipliers is widely known as 4DVAR; objective mapping is commonly 3DVAR; and many of the physical terms in fluid mechanics have their own particular meteorological definition (e.g., "mixing ratio"). Most of the meteorological methods can be recognized as variations on methods described here, albeit the development usually begins with the continuous time/space formulation. Chapter 5 discussed some of the approximations invoked, following discretization, including so-called nudging, where the Kalman gain is replaced by a plausible, but guessed, fixed matrix; state reduction; the use of approximate rather than exact adjoints; and many others that have proven of practical use.

For a discussion of the application of methods very similar to those discussed here to the problems of atmospheric trace gas movement, see Enting (2002). The collection of papers by Kasibhatla *et al*. (2000) applies the methods to biogeochemical modeling.

Because meteorological forecasters have recognized the importance of giving their users an idea of reliability, a very large meteorological effort has gone into so-called ensemble forecasting methods, which were briefly discussed in Chapter 5; Kalnay (2003) is a good starting point in what is now a very large literature.