

Ocean acoustic tomography: Integral data and ocean models

Bruce D. Cornuelle and Peter F. Worcester

Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093, U.S.A.

Abstract

Tomographic data differ from most other oceanographic data because their sampling and information content are localized better in spectral space than in physical space. Approximate data assimilation methods optimized for localized physical space measurements can lose much of the non-local tomographic information, degrading the potential performance of the tomographic data in fixing the model state. In addition, methods which require inverting the tomographic data to a physical space grid before insertion into the model as point measurements incur special problems because of the non-local nature of the errors. We give a simple 1-D example to illustrate the problems that can arise. Fortunately, methods that directly assimilate integral measurements and preserve all of the information in the tomographic data, such as the Kalman filter, are becoming more practical as computer speed grows.

1. INTRODUCTION

Ocean acoustic tomography is a remote sensing technique that exploits the transparency of the ocean to low frequency sound and the sensitivity of acoustic propagation to the ocean sound speed (temperature) and current fields (Munk and Wunsch, 1979). The travel times of acoustic pulses or the group delays of acoustic normal modes are interpreted to provide information about the intervening ocean. Tomographic data differ from more familiar data types, such as moored temperature and current measurements or hydrographic data, in that the data are integrals of the temperature and current fields along the acoustic paths, rather than values at a single location.

Sound speed perturbations are at least an order of magnitude greater than current speeds, dominating travel time perturbations. Measurements of travel times between multiple points in the ocean can therefore be used to estimate the ocean sound speed field. The effects of sound speed perturbations and currents can be separated using acoustic transceivers, rather than separate sources and receivers. The difference in travel times of acoustic pulses traveling in opposite directions is to first order proportional only to the integral of current along the path. The sum of the travel times is proportional only to the integral of the sound speed field. Reciprocal transmissions can therefore be used to measure both sound speed and current.

Linear inverse methods are commonly used to estimate the sound speed and current fields from tomographic data. These methods are well-developed for data collected more

or less simultaneously, but much less effort has been devoted to the methods needed to combine data collected at different times. Howe *et al.* (1987) applied a Kalman filter to combine data collected at different times from a single source-receiver pair, with the assumption that the ocean tends to climatology with a 10-day time constant (i.e., assuming that the perturbation state vector decays to zero with a 10-day time constant). Spiesberger and Metzger (1991) employed a generalized objective mapping approach in which a temporal covariance function was specified, in addition to the usual spatial covariance function, to analyze travel-time time series from a different experiment.

Neither of these approaches takes advantage of our knowledge of ocean dynamics to constrain how the temperature and current fields evolve. Chiu and Desaubies (1987) included dynamics in the analysis of data from the 1981 Tomography Demonstration Experiment (Ocean Tomography Group, 1982; Cornuelle *et al.*, 1985) in perhaps the simplest possible way, assuming that the ocean perturbations were made up of linear Rossby waves and using the data to estimate time-independent amplitudes and phases of selected Rossby wave components in a global (non-linear) least squares fit. Cornuelle (1990) used a Kalman filter with a linear ocean model to combine data obtained at different times in a simulated moving ship tomography experiment. More generally, however, fully non-linear ocean general circulation models (GCMs) are required to describe realistically the evolution of the ocean. Schröter and Wunsch (1986) described one approach to forcing a steady ocean GCM to consistency with a comparatively sparse set of tomographic measurements, but did not consider the time-dependent problem. Munk *et al.* (1995) and Wunsch (1990) discuss the general problem of combining tomographic data with time-dependent ocean models to obtain the best possible estimate of the state of the ocean. Sheinbaum (1989) performed simulations in which tomographic measurements were combined with a simple, time-dependent, one-layer ocean model. Fukumori and Malanotte-Rizzoli (1995) constructed an approximate Kalman filter for a nonlinear, primitive equation model of the Gulf Stream, and examined the assimilation of various pseudomeasurements, including tomographic observations. The same fundamental methods used for other oceanographic data are in principal also applicable to tomographic data.

There are practical difficulties due to the large number of state variables in ocean GCMs, however. The computer time and memory required to implement exact assimilation methods have led to the development of various approximate methods for assimilating point measurements into ocean GCMs. Unfortunately, approximate methods suitable for point measurements often are not suitable for tomographic data, due to the quite different ways in which point sensors and tomographic measurements sample the ocean. Sequential methods that do not retain the full uncertainty covariance, for example, lose significant information from step to step when used with tomographic data. In this paper, it is not our goal to review the inverse methods required to combine tomographic measurements with ocean models, for which we refer the reader to Munk *et al.* (1995). Rather, we review the sampling properties of tomographic measurements, and present a cautionary note as to what can go wrong when approximate methods developed for point measurements are blindly applied to tomographic data.

2. SAMPLING PROPERTIES OF ACOUSTIC RAYS

Most tomographic applications outside the earth sciences are characterized using spectral methods, such as the projection-slice theorem central to the reconstruction in medical tomography. The irregularity of the ray paths in geophysical and ocean tomography destroys the simplest spectral relationships, but the non-local character of the data remains.

2.1. Vertical slice: Range-independent

A measurement of acoustic ray travel time is sensitive to the integral of sound speed and tangential current along the entire ray path, although the weighting is not uniform in range, because ray travel time is most sensitive to propagation speed perturbations at the upper and lower ray turning points. This depth dependence can be seen most easily in the range-independent case (Munk and Wunsch, 1982; Cornuelle *et al.*, 1993). To first order, ray travel time perturbations are weighted averages of the sound-speed perturbations integrated along the unperturbed ray paths, Γ_i ,

$$\Delta T_i = - \int_{\Gamma_i} \frac{ds}{C_0^2(x,z)} \Delta C(x,z)$$

where ocean currents have been neglected. For the range-independent case, in which $C_0(z)$ and $\Delta C(z)$ are independent of x , this can be converted to an integral over z

$$\Delta T_i = - \int_{\Gamma_i} \frac{dz}{C_0^2 \sqrt{1-(C_0/\hat{C})^2}} \Delta C(z)$$

using Snell's law, $C(z)/\cos \theta = \hat{C}$. \hat{C} is the sound speed at the ray turning point and θ is the angle relative to horizontal. The function

$$\frac{1}{C_0^2(z) \sqrt{1-(C_0(z)/\hat{C})^2}}$$

gives the weighting with which the $\Delta C(z)$ contribute to ΔT_i . The (integrable) singularity at the ray turning point depths, \hat{z} , where $C_0(\hat{z}) = \hat{C}$, clearly shows that ray travel times are most sensitive to the ocean at ray turning point depths. There is an ambiguity in that singularities occur at both the upper and lower turning points.

As an example, Cornuelle *et al.* (1993) computed the travel time perturbations due to a sound speed perturbation with an amplitude of -1 m/s at 100 m depth, linearly decreasing to zero at 90 and 110 m. The travel time anomaly, computed by subtracting the travel time in the unperturbed ocean from the travel time in the perturbed ocean, is sharply peaked for rays that turn between 90 and 110 m (Fig. 1). The anomaly is zero for rays with upper turning depths below 110 m because they do not sample the perturbed region. Rays that have upper turning depths above 90 m have non-zero anomalies, because they traverse the perturbed region, but the anomalies are relatively small because the ray weighting function falls off rapidly with distance from the turning point.

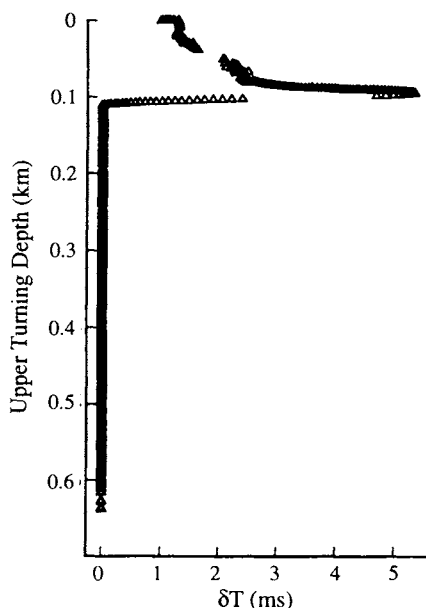


Figure 1. Travel time anomalies computed for a sound speed perturbation with an amplitude of -1 m/s at 100-m depth linearly decreasing to zero at 90 and 110 m. The unperturbed profile is range-independent. (Reproduced from Cornuelle *et al.*, 1993.)

2.2. Vertical slice: Range-dependent

A ray trapped in the sound channel, with upper and lower turning points at regular intervals, samples the ocean periodically in space, so that its travel time is sensitive to some spatial frequencies, but unaffected by others. Chiu *et al.* (1987) found from simulations, and Howe *et al.* (1987) showed experimentally, that some range-dependent information on the sound speed field in a vertical slice can be extracted from data obtained with a single acoustic transceiver pair. Cornuelle and Howe (1987) subsequently showed that features which match the ray periodicity (i.e., have the same wavelength) generate large signatures in travel time. The ray paths are somewhat distorted sinusoids in midlatitudes, and so contain higher harmonics. Scales short compared to a double loop length that match these harmonics also generate significant signatures in travel time. This result can be qualitatively understood in the spatial domain (Worcester *et al.*, 1991). Scales short compared to the double loop lengths of the rays affect only the few rays that pass through the feature, leaving the other rays unaffected. This differing sensitivity of the rays gives the fundamental information required to invert for scales short compared to double loop

lengths. Features which have wavelengths substantially longer than the ray double loop length (about 50 km in midlatitudes) affect all rays similarly and are indistinguishable from a change in the mean over the entire range between the source and receiver. Travel times are therefore sensitive to ocean features with wavelengths comparable to and shorter than a ray double loop length, as well as to the range-average.

This behavior is most easily understood in the wave number domain, by choosing the model parameters used to represent the ocean sound speed perturbation field to be the complex Fourier coefficients in a spectral expansion. For any specific geometry, the sensitivity of the travel time inverse to various wave numbers can be quantified by plotting the diagonal of the resolution matrix. The resolution matrix gives a particular solution to any linear inverse problem as a weighted average of the true solution (e.g., Aki and Richards, 1980). For the specific case of two moorings separated by 600 km, with a source and five widely separated receivers on each mooring, the diagonal of the resolution matrix for spectral model parameters clearly shows the sensitivity of tomographic measurements to the mean and to harmonics with scales comparable to and shorter than a ray double loop length (Fig. 2). There are obvious spectral gaps for wave numbers between the mean and first harmonics of the ray paths, and again between the first and second harmonics. Travel times are not sensitive to wave numbers in these regions. The harmonics extend over bands of wave numbers because the eigenrays connecting the source and receiver have a range of double loop lengths.

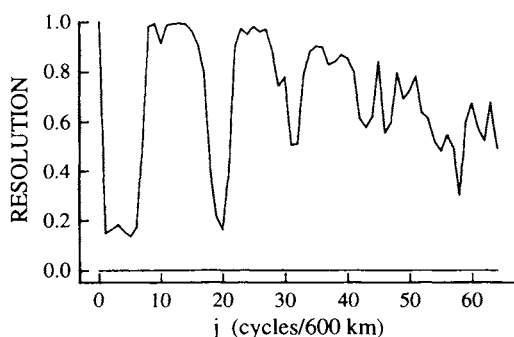


Figure 2. "Transfer function" (the diagonal elements of the resolution matrix) for a 156-ray data set with an expected variance spectrum (for the lowest baroclinic mode) that is constant for wave numbers $j = 1$ to 3 and decreases as j^{-2} for $j > 3$ (wavelengths smaller than 200 km). (Reproduced from Cornuelle and Howe, 1987.)

2.3. Horizontal slice

As was the case for the vertical slice, the key to understanding the horizontal sampling properties of acoustic travel times is to consider the wave number domain, rather than physical space (Cornuelle *et al.*, 1989). For simplicity, consider a two-dimensional ocean consisting of a horizontal slice. Rays are then straight lines connecting sources and receivers (neglecting horizontal refraction, which is usually small due to the small horizontal gradients in the ocean). The sound speed perturbation field $\Delta C(x,y)$ can be expanded in truncated Fourier series in x and y , giving the spectral representation:

$$\Delta C(x,y) = \sum_k \sum_l P_{kl} \exp \frac{2\pi i}{L} (kx + ly), \quad k, l = 0, \pm 1, \dots, \pm N$$

where L is the size of the domain and k, l are the wave numbers in x and y , respectively. With this representation of the ocean, the inverse problem is to determine the complex Fourier coefficients P_{kl} from the travel time data. Consider a scenario in which two ships start in the left and right bottom corners of a 1 Mm square and steam northward in parallel, transmitting from west to east every 71 km for a total of 15 transmissions (Fig. 3a). The inversion of the 15 travel times leads to an estimate which consists entirely of east-west contours. All of the ray paths give zonal averages, with no information on the longitudinal dependence of the sound speed field. Similarly, transmissions between east-to-west moving ships give meridional averages, and the resulting estimate consists entirely of north-south contours (Fig. 3b).

To interpret these results in wave number space, note that with the field expressed as above,

$$\int_0^L \Delta C(x,y) dx = 0 \quad \text{for } k \neq 0$$

East-west transmissions therefore only give information on the parameters P_{0l} , as can be seen in the expected predicted variance plot in wave number space of Fig. 3a. Similarly, with north-south transmissions only the parameters P_{k0} are determined (Fig. 3b).

Combining east-west and north-south transmissions determines both P_{0l} and P_{k0} . Not surprisingly, this is still inadequate to generate realistic maps because most of the parameters remain unknown (Fig. 3c). More complex geometries with scans at 45° give a distinct improvement by determining the parameters for which $k = l$, but at the cost of excessive ship time (Fig. 3d). In all cases, the division between well-determined and poorly-determined parameters is simplest in spectral space.

To generate accurate maps of the ocean mesoscale field requires ray paths at many different angles to determine all of the wave number components. This requirement must be independently satisfied in all regions with dimensions comparable to the ocean decorrelation scale. Because this is impossible to achieve using two ships in any reasonable time period, a combination of moored and ship-suspended instruments was found to be required to achieve residual sound speed variances of a few per cent.

These results are a direct consequence of the projection-slice theorem (Kak and Slaney, 1988).

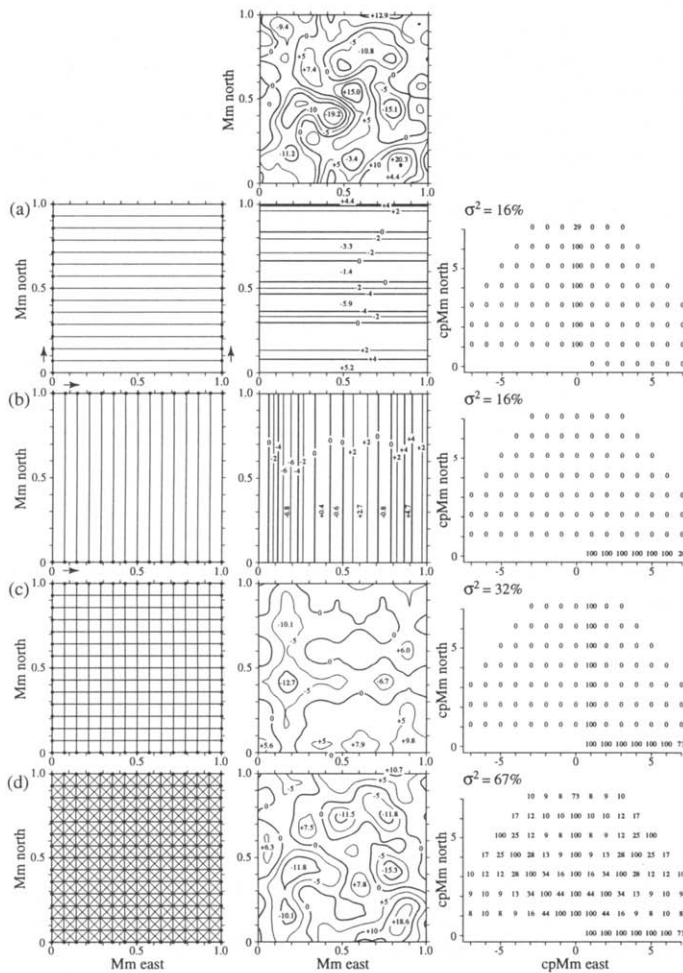


Figure 3. Ship-to-ship tomography. The top center panel is the "true ocean", constructed assuming a horizontally homogeneous and isotropic wave number spectrum with random phases. Energy decreases monotonically with increasing scalar wave number, giving an approximately Gaussian covariance with $1/e$ decay scale of 120 km. (a) W→E transmissions between two northward traveling ships (left panel). Inversion of the travel time perturbations produce east-west contours in ΔC (middle) with only a faint relation to the "true ocean". Expected predicted variances in wave number space (right) are 0% (no skill) except for $(k, l) = (0, 1), (0, 2), \dots, (0, 7)$, which accounts for $\sigma^2 = 16\%$ of the ΔC variance. (b) S→N transmissions between two eastward traveling ships. (c) Combined W→E and S→N transmissions, accounting for 32% of the ΔC variance and giving a slight pattern resemblance to the true ocean. (d) Combined W→E, S→N, SW→NE, and SE→NW transmissions, accounting for 67% of the variance and giving some resemblance to the true ocean. (Adapted from Cornuelle *et al.*, 1989.)

3. INTEGRAL VS. POINT DATA: INFORMATION CONTENT IN A 1-D EXAMPLE

Because of the sampling properties of the tomographic measurements, when travel time data are used to estimate the sound speed or current in a volume of ocean, the uncertainty in the solution is generally local in wave number space, rather than in physical space. This is in contrast to the uncertainty in estimates made from independent point measurements, which tend to have errors localized in physical space. The non-locality can be characterized in least-squares estimation by examining the output model parameter uncertainty ('error') covariance. For model parameters localized in physical space (such as boxes or finite elements), significant off-diagonal terms in the output uncertainty covariance matrix $\hat{\mathbf{P}}$ (see Appendix) show that the uncertainty at one point is related to the uncertainty at another point. These correlations arise in estimates made from tomographic data; the error at one point along a ray path tends to be anti-correlated with the error at other points on the ray path, because the sum of the points is known. (Off-diagonal correlations can arise in satellite altimetry measurements as well, because the orbit may contaminate many measurements along the ground track with approximately the same error.)

It is common practice in objective mapping to display only the diagonal of the physical space uncertainty covariances, which is what is plotted in the 'error map' used in many papers (Bretherton, Davis, and Fandry, 1973). Assimilation methods that insert values at points in the model, such as 'nudging' (e.g., Malanotte-Rizzoli and Holland, 1986), sometimes use the local error bars from the objective map to adjust the strength of the data constraint. We will show below that although neglecting the off-diagonal components of the uncertainty covariance of the estimates is benign for point measurements with local covariances, it can be dangerous when the off-diagonal terms are significant, as when tomographic data are used. Doing the data insertion by some form of sequential optimal interpolation (OI), approximating the Kalman filter, is less dangerous, but the treatment of the model parameter uncertainty covariance between steps can still both destroy information and give misleading results when used with non-local data.

In order to highlight the contrast between tomographic and point measurements, we have chosen a very simple model problem for pedagogical clarity. We use a 1-D, periodic realm of 20 piecewise-linear finite elements, each with identical widths and randomly chosen temperatures, which are assumed to be exactly convertible to sound speeds. The unknowns (model parameters) are the temperatures at each of the points (and by interpolation, the temperature of the entire interval). The model parameters are assumed to be independent, identical, normally distributed random variables with zero mean and independent, equal (unit) variances. The initial model parameter uncertainty covariance matrix (\mathbf{P}) is therefore diagonal with 1.0 on the diagonal. The 'error map' for this a priori state is a uniform, unit error variance in physical space, which is conveniently the same as the diagonal of the matrix plotted as a function of the location index. If we measure the value of one of these finite elements, say element 5, by sampling it at its center without noise, the output uncertainty variance is now zero for the sampled point (Fig. 4). If we instead measure the integral across all the points (as a representation of a tomographic measurement), the uncertainty variance is instead reduced only slightly at all the points (Fig. 4). This obscures the fact that something very specific has been learned, just as specific as for the point measurement. In fact, the sum of uncertainty variances (the trace of the covariance) has been reduced by the same amount (1 unit) in each case, to 19 from 20.

The full uncertainty covariance matrix gives complete information about the character of the remaining uncertainty. The first and fifth columns of the covariance are plotted in Fig. 5 for both point and averaged measurements. (This is the covariance between the error at points 1 or 5 and the error at all the rest of physical space.) The errors at the points missed by the point measurement show no correlation with the errors at their neighbors, while the fifth point (the site of the measurement) shows no error. In contrast, the average shows identical behavior at all points (Fig. 5). The variance is reduced slightly, as shown in Fig. 4, but there is a uniform negative correlation between the error at any point and the error at all other points, representing the knowledge of the average. That is, if the value at the one of the points is higher than estimated, the values at all the other points will tend to be lower than estimated, in order to keep the average the same.

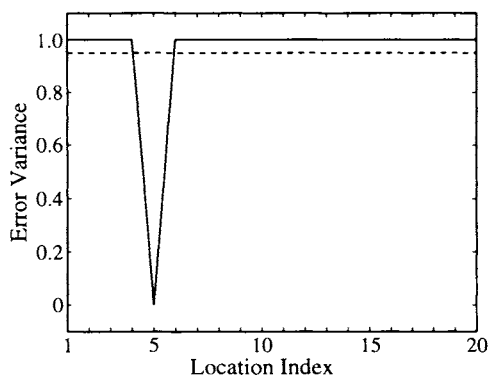


Figure 4. Error variance map (diagonal of the error covariance) from a single, perfect, point measurement at location 5 (solid) or from a single, perfect measurement of the average over 20 points (dashed).

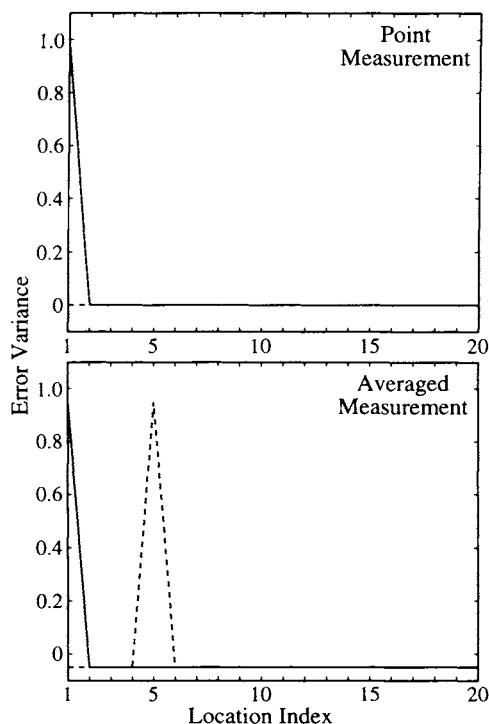


Figure 5. Columns 1 (solid) and 5 (dashed) of the error covariance for the point measurement at location 5 and for the averaged measurement.

The similarity in the amount of information available from either type of measurement is most easily seen by decomposing the output error covariance $\hat{\mathbf{P}}$ into eigenvalues and eigenvectors.

$$\hat{\mathbf{P}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$$

where \mathbf{U} is the matrix with the eigenvectors of $\hat{\mathbf{P}}$ as its columns, and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues. In both cases, there is one zero eigenvalue, representing the component of the model that is known exactly, and 19 eigenvectors with eigenvalue 1. The null space in each case is degenerate, and can be represented by many different sets of basis functions, but the eigenvector that is known is either a spike at the measured point or a uniform level across all points (Fig. 6). In either case, only 19 unknowns are left to be determined.

Representing the estimate generated from the average as 20 point measurements with equal values and independent error bars with values as in Fig. 4, amounts to dropping the off-diagonal terms in the error covariance. Although the diagonalized covariance has the same trace as the exact covariance, and thus has a similar amount of information as measured using the trace, the eigenvectors of the diagonalized covariance are 20 unknown functions, each with variance 19/20 of the original. This is a considerably different state of knowledge than in the original error covariance, because 20 unknowns remain to be determined, and the slight reduction in their uncertainty is not very useful.

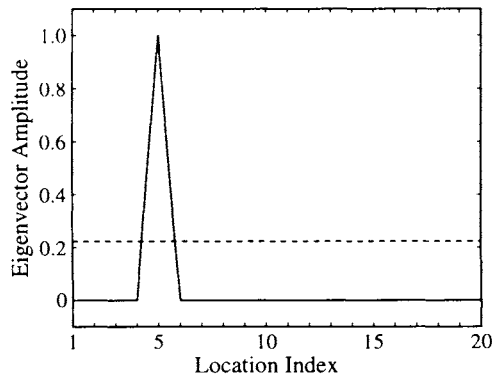


Figure 6. The eigenvector with zero eigenvalue (the single, perfectly known component of the model), for the point measurement at location 5 (solid) and the average measurement (dashed).

4. INTEGRAL VS. POINT DATA: ESTIMATION IN A TIME-DEPENDENT 1-D EXAMPLE

Suppose now that the measurements are repeated, and that we wish to combine all the measurements to obtain an improved estimate of the field, assuming that we have a model for the dynamical evolution of the field. The time-dependent least-squares estimation problem can either be solved sequentially (the Kalman filter), or globally, giving identical results when fully optimal methods are used. In the examples to follow, the estimates have been described as the result of sequential estimation, since that is similar to many approximate schemes.

A primitive dynamical example that includes advection demonstrates the effect of error propagation in a data assimilation scheme using non-local data. The 20-point domain was retained from the earlier example, but is now assumed to be periodic, with periodicity 20, so that the twenty-first point is the first point. The dynamics were uniform, constant advection of temperature as a passive tracer: every time step, the field shifted one place to the left, so 21 time steps completely rotate the domain back to the starting point. The a priori information state was again independent, unit variances for each point.

It is obvious that 20 sequential, perfect point measurements, one per time step at a single location, would completely determine the field if the dynamics were modeled perfectly. On the other hand, 20 sequential, perfect 20-point averages would only determine the average of the domain, doing no better than a single perfect measurement of the average. If the measurements have error bars, then repeated averaged measurements increase the precision of the estimate of the average, but give no new information about the shorter scale structures, since every measurement is the same.

Instead of 20-point averages, the tomographic data were therefore taken to be 5-point averages, which do not trivially repeat in the periodic domain and which yield some information on smaller scale structures. To avoid singularities due to perfect measurements, both point measurements and averages were assumed to be contaminated by noise. The point measurements were assumed to have an uncertainty variance of 0.1, representing measurement error. Because each tomographic measurement averages 5 points, the uncertainty variance was set at 0.02, one-fifth of the error variance assumed for the point measurements, keeping the signal-to-noise variance ratio (SNR) equal between the averaged and point measurements. (This is a relatively arbitrary choice, but is simplest for comparisons, because a single measurement of either type produces the same decrease in model uncertainty.)

We used a Kalman filter (see Appendix) to combine the measurements, and compared the performance after all the measurements had been used. In this example, the tomographic measurements show significantly larger point error bars than the point measurements, even though the estimate used optimal error propagation (Fig. 7). Columns 1 and 5 of the output error covariance matrix (Fig. 8) show the contrast in structure between the point measurements and the tomography. The eigenvalue spectra (Fig. 9) show that the averaged data have a varying information content in spectral space. The 20-point mean is well determined (the smallest eigenvalue). The eigenfunctions corresponding to the next two smallest eigenvalues resemble sine and cosine functions with one cycle in the 20-point domain (Fig. 10). The eigenfunctions corresponding to the next two smallest eigenvalues similarly resemble sine and cosine functions with two cycles in the 20-point domain (Fig. 10). The 4 vectors in the null space, which each average to zero in any 5-point domain

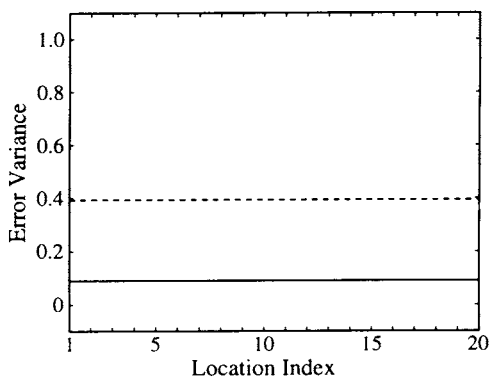


Figure 7. Error variance map (diagonal of output error covariance) with advection after a 20-step dynamical inverse, using the point measurements (solid) and the 5-point averages (dashed).

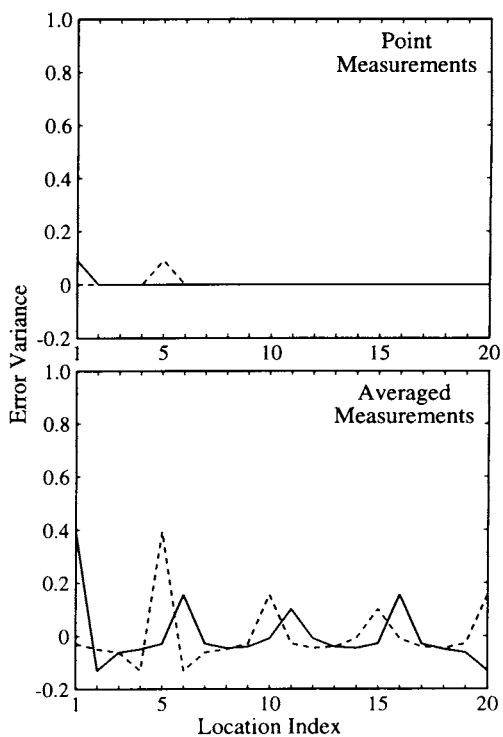


Figure 8. Columns 1 (solid) and 5 (dashed) of the error covariance with advection for the 20 point measurements and for the 20 5-point averages.

and are thus completely undetermined by the tomography, appear to be more complex, but correspond to aliased samples from sine and cosine functions that have either one or two cycles over the 5-point domain (Fig. 10). In spectral terms, the averaged measurements selectively determine the large-scale components better than the short-scale components, while the point measurements are equally sensitive to all scales. For geophysical systems with red signal spectra, this suggests that the averaged measurements may perform better than in this simulation, which assumes a white spectrum.

Even in this simple example, the large-scale components of the field (i.e., the five eigenvectors shown in Fig. 10 associated with the five smallest eigenvalues in Fig. 9) are better determined by the tomographic measurements than by the point measurements. This is not entirely obvious from Fig. 9, because the eigenvectors of the 20 point measurements are 20 delta functions that are localized in physical space, while the eigenvectors of the tomographic measurements are localized in spectral space, and so the eigenvalues (variances) plotted in the figure do not correspond to similar eigenvectors for the two cases.

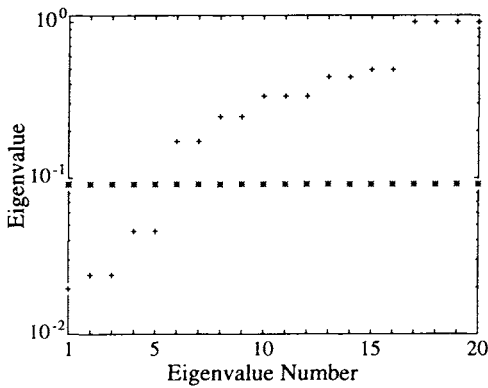


Figure 9. Eigenvalue spectra for the error covariance matrix with advection for the 20 point measurements (*) and the 20 5-point averages (+).

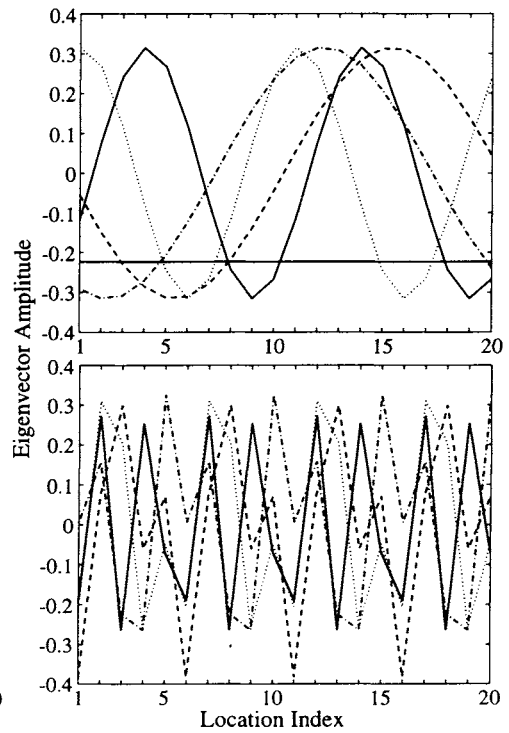


Figure 10. The 5 best-determined eigenvectors (top) and the 4 null space vectors (bottom), for 20 5-point averaged measurements with advection.

Eigenvector 1 for the 20 5-point averages is related to the mean, for example, while eigenvector 1 for the 20 point measurements is a delta function at location 1. The variance of the mean deduced from the 20 point measurements is simply $1/20$ of the 20 identical variances (eigenvalues) of the 20 eigenvectors. Because the eigenvectors are normalized to have unit length, the elements of eigenvector 1 for the 20 5-point averages all have magnitude $1/\sqrt{20}$, as can be seen in Fig. 10. The variance of the mean is then $1/20$ of the variance (eigenvalue) of the eigenvector. The ratios of the eigenvalues in Fig. 9 therefore accurately reflect the ratios of the error variances of the estimates made using tomographic measurements and point measurements, showing that the large-scale components are better determined by the tomographic data. (Another way of looking at this is to note that because this simple example is homogeneous, sines and cosines are also eigenvectors of the point measurement covariance.)

A somewhat different question is how well the tomographic and point measurements resolve the detailed spatial structure of the field, rather than just the large-scale components. The tomographic measurements were seen in Fig. 7 to have significantly larger point error bars than the point measurements for the case with 20 measurements. In that case, however, the tomographic measurements had more null space vectors than the point measurements. To give the tomographic and regular measurements similar numbers of null space vectors, we repeated the simulations using only 16 data in each case. The null space for the point measurements has four elements, representing the four points not measured, while the null space for the tomography remains the same. The trace of the output error covariances (the total uncertainty variance after the inverses) were 9.2 and 5.5 for the averaged and point measurements, respectively. The point measurements thus do better in resolving the detailed spatial structure of the field than the averaged measurements, when the unknown field has a white spectrum and when both types of data have equal SNR. The averaged measurements are most sensitive to the larger scales, as discussed above, and have to be differenced in order to resolve finer scales. The difference of two large numbers is easily contaminated by random noise. If the calculations are repeated giving the tomography data 0.1 of their original variances, (tomographic SNR = 10 time point measurement SNR), the trace of the output error covariance for the tomography is now 5.1, so the greater measurement precision has greatly improved the ability of the tomographic data to resolve the detailed spatial structure. This result has been shown before in a number of places, (e.g., Cornuelle *et al.*, 1985), but rarely in such a simple example. The performance of averaged measurements is equal to that of point measurements (as measured by the trace) as long as the averaged measurements do not overlap. The redundant data generated by overlapping averages reduce the calculated performance, just as repeated point sampling in the same place would.

4.1. Approximate sequential methods with advection

The issue that remains is to compare measurements fed in sequentially without keeping the off-diagonal model error covariance elements. We again use the example of 20 sequential measurements over 20 time steps, but approximating the forecast of the model parameter uncertainty covariance matrix (Appendix, equation A7). This is meant to model sequential optimal interpolation methods, where only a simplified version of the model parameter uncertainty is propagated between steps. If only the diagonal of the covariance is kept, then the total expected error for the tomographic measurements changes only slightly, increasing by about 3% compared to the exact (full covariance) result, while the total error for the point measurements is unchanged. The diagonal-only Kalman filter is still optimal for the point measurements because the covariance is completely local, and simple advection does not produce off-diagonal terms during the evolution of the model. Because the total expected error changes only slightly for the tomography, it is tempting to assume that the loss of off-diagonal terms has only slightly degraded the estimates.

Unfortunately, a look at the eigenvalue spectra from a sequential, diagonal-only estimation with 20 tomographic or point measurements (Fig. 11) shows that for the tomography, the approximate sequential interpolation arrived at a vastly different (and incorrect) state of information than the exact Kalman filter. The spectrum of eigenvalues for the tomography no longer shows the large-scale components as being best determined, and the model state apparently includes information about all components (no zero eigenvalues, so no

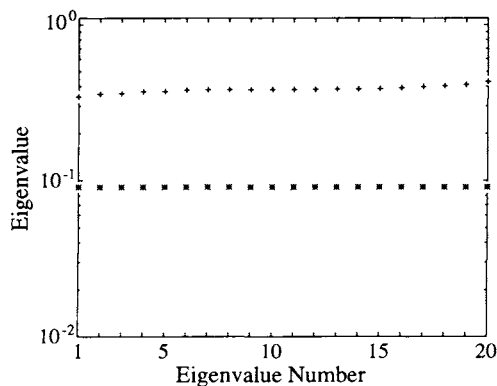


Figure 11. Same as Fig. 9, but for a simulation in which only the diagonal of the covariance matrix was kept.

null space). This is in contrast to the point measurements, whose eigenvalue spectrum is unchanged.

Because of the approximate error propagation, the error covariance is no longer a good figure of merit, and the true performance of the simplified method can best be evaluated by Monte Carlo methods, simulating an ensemble of true fields and looking at the error in the reconstruction. We only wish to point out that the diagonal-only method remains optimal for the point measurements, while becoming severely suboptimal for the tomographic measurements, but in a subtle way that could easily be overlooked. This contrast is heightened by the the trivial dynamics chosen for the simulations. Realistic dynamics, such as quasi-geostrophic flow in three dimensions, generally creates non-local covariances, even from point sampling, so that the sequential optimal interpolation would degrade the point measurements somewhat. On the other hand, for short time scales and normal advection velocities, the point measurement information will remain much more local than tomographic information, and so is more compatible with local approximations. Conversely, if the dynamical model is built in spectral space, so the horizontal basis functions are sines and cosines, then the tomographic data is much more local than point measurements, which are sensitive to all scales.

Most modern data assimilation methods do not completely ignore off-diagonal terms in the model parameter uncertainty covariance matrix, however, even for point measurements. It is therefore natural to ask how well other possible approximations to the uncertainty covariance matrix perform. Perhaps the simplest class of approximations are ones in which varying numbers of diagonal bands of off-diagonal elements are retained, while the remaining elements are set to zero. Plotting the eigenvalue spectra as a function of the number of bands retained (Fig. 12) shows that retaining one off-diagonal band, in addition to the diagonal elements, results in the reduction of a single eigenvalue, corresponding to the mean. Little further change in the spectra is evident as additional off-diagonal bands are retained, until 15 off-diagonal bands are included. At that point the spectra begin to

resemble the spectrum obtained when the full matrix is used. For the simple example considered here, retaining additional off-diagonal bands of the uncertainty covariance matrix is therefore not a particularly effective approximation, as nearly the complete uncertainty covariance matrix needs to be retained before the results are similar to those obtained using the full matrix.

The decomposition of the error covariance into eigenvectors suggests a more natural approximation for sequential assimilation, however, in which only the components of the model error covariance with large eigenvalues are propagated by the model. In the case of a single measurement, the savings are small, because 19 out of 20 vectors need to be propagated, but with more complete observations, the savings could be larger.

4.2. Separating the inverse from the assimilation

Even the approximate method used in the previous example kept the inverse as part of the update of the model. Some older assimilation methods invert the measurements and then blend in the results as pseudo-point measurements with error bars. This approach is impossible when using the averaged measurements, because the uncertainty of the output estimate is not local, and so the pointwise error bars cannot express the infinite (but correlated) uncertainty imposed on the solution by the elements of the null space. Even if the data are inverted outside the model, it is necessary to use the model state as the reference; otherwise the inversion procedure will tend to pull the model toward whatever reference state is used. This problem of infinities is avoided in exact sequential optimal estimation and the Kalman filter, because the data are merged into the model directly, inverting for corrections to the current best forecast of the model parameters, and the a priori error bars describe the model's current state of knowledge.

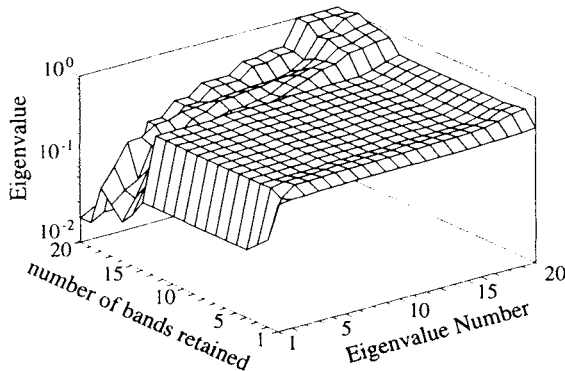


Figure 12. Eigenvalue spectra for the error covariance matrix with advection for the 20 5-point averages, as a function of the number of bands of off-diagonal elements retained in the model parameter uncertainty covariance matrix. The spectrum obtained retaining only one band (i.e., the diagonal) is the same as that in Fig. 11; the spectrum obtained retaining 20 bands (i.e., the full matrix) is the same as that in Fig. 9.

5. DISCUSSION

These simple examples were constructed to emphasize the spectral nature of tomographic measurements, in contrast to the local nature of point measurements. This is closely related to the projection-slice theorem (Kak and Slaney, 1988), but the irregularity of the ray paths in ocean acoustic tomography destroys the simplest spectral relationships, concealing the spectral nature of the sampling.

The example reported in Figs. 9 and 10 shows that the error covariance matrix of the averaged measurements has sines and cosines as eigenvectors, while the error covariance matrix of the point measurements is diagonal with delta functions as one set of eigenvectors. For an unknown field with a white spectrum, and data with equal signal-to-noise ratios, non-overlapping averaged measurements increase our knowledge of the unknown field by the same amount as the same number of point measurements, but the spectral content of that knowledge is very different. Because the averaged measurements determine the lower wave numbers better than the higher wave numbers, they have advantages if the spectrum of the unknown field is red. Determining high wave number information from the averaged measurements is more difficult, unless the measurement errors are sufficiently small to make differencing of the integral measurements practical. The relative utility of tomographic measurements and point measurements thus depends strongly on the goal of the measurement program.

The non-local nature of the averaged measurements also makes it difficult to use approximations to the Kalman filter in dynamical models with local parameterization. Conversely, the averaged measurements *can* be used efficiently by an approximate Kalman filter based on spectral functions.

Acknowledgments. This work was supported by the Office of Naval Research (ONR Contracts N00014-93-1-0461 and N00014-94-1-0573) and by the Strategic Environmental Research and Development Program through the Advanced Research Projects Agency (ARPA Grant MDA972-93-1-0003).

APPENDIX

The form of least-squares estimation used here assumes that the expected value of the model parameter vector has been removed, so $\langle \mathbf{m} \rangle = 0$, and that an initial guess exists for the covariance of the uncertainty around the expected value, $\langle \mathbf{m} \mathbf{m}^T \rangle = \mathbf{P}$. The data are related to the model parameter vector by a linear relation,

$$\mathbf{d} \approx \mathbf{G} \mathbf{m} + \mathbf{n} \quad (\text{A1})$$

where \mathbf{n} is the random noise contaminating the measurements. Any known expected value of the noise is assumed to have been removed, so $\langle \mathbf{n} \rangle = 0$, and the noise is assumed to have covariance $\langle \mathbf{n} \mathbf{n}^T \rangle = \mathbf{N}$ and to be uncorrelated with the model parameters. This relation can be inverted to obtain an estimate of the model parameters,

$$\hat{\mathbf{m}} = \mathbf{P} \mathbf{G}^T (\mathbf{G} \mathbf{P} \mathbf{G}^T + \mathbf{N})^{-1} \mathbf{d} \quad (\text{A2})$$

and the expected uncertainty in this estimate is

$$\hat{\mathbf{P}} = \mathbf{P} - \mathbf{P}\mathbf{G}^T(\mathbf{G}\mathbf{P}\mathbf{G}^T + \mathbf{N})^{-1}\mathbf{G}\mathbf{P} \quad (\text{A3})$$

If dynamics are available to forecast the model parameter vector between time steps, so that

$$\mathbf{m}_{t+1} = \mathbf{A}\mathbf{m}_t + \mathbf{q} \quad (\text{A4})$$

where \mathbf{A} is the transition matrix, and \mathbf{q} is the uncertainty in the forecast due to errors in the dynamics (with zero mean and uncertainty covariance $\mathbf{Q} = \langle \mathbf{q}\mathbf{q}^T \rangle$). The Kalman filter performs a sequential cycle, correcting the starting guess by inverting the differences between the observations and the predicted data,

$$\hat{\mathbf{m}}_t = \mathbf{m}_t + \mathbf{P}_t\mathbf{G}^T(\mathbf{G}\mathbf{P}_t\mathbf{G}^T + \mathbf{N})^{-1}(\mathbf{d}_t - \mathbf{G}\mathbf{m}_t) \quad (\text{A5a})$$

$$\hat{\mathbf{P}}_t = \mathbf{P}_t - \mathbf{P}_t\mathbf{G}^T(\mathbf{G}\mathbf{P}_t\mathbf{G}^T + \mathbf{N})^{-1}\mathbf{G}\mathbf{P}_t \quad (\text{A5b})$$

and forecasting the estimate and covariance to the start of the next step:

$$\mathbf{m}_{t+1} = \mathbf{A}\hat{\mathbf{m}}_t$$

$$\mathbf{P}_{t+1} = \mathbf{A}\hat{\mathbf{P}}_t\mathbf{A}^T + \mathbf{Q} \quad (\text{A7})$$

This cycle then repeats.

REFERENCES

- Aki, K., and P. Richards, 1980. *Quantitative Seismology, Theory and Methods*. 2 Vols. W.H. Freeman and Co.
- Bretherton, F. P., R. E. Davis, and C. B. Fandry, 1976. A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep Sea Res.*, **23**, 559-582.
- Chiu, C.-S., and Y. Desaubies, 1987. A planetary wave analysis using the acoustic and conventional arrays in the 1981 Ocean Tomography Experiment. *J. Phys. Oceanogr.*, **17**, 1270-1287.
- Chiu, C.-S., J. F. Lynch, and O. M. Johannessen, 1987. Tomographic resolution of mesoscale eddies in the marginal ice zone: A preliminary study. *J. Geophys. Res.*, **92**, 6886-6902.
- Cornuelle, B.D., 1990. Practical aspects of ocean acoustic tomography. In: *Oceanographic and geophysical tomography: Proc. 50th Les Houches Ecole d'Ete de Physique Theorique and NATO ASI*, Y. Desaubies, A. Tarantola, and J. Zinn-Justin, eds., Elsevier Science Publishers, 441-463.
- Cornuelle, B.D., and B.M. Howe, 1987. High spatial resolution in vertical slice ocean acoustic tomography. *J. Geophys. Res.*, **92**, 11,680-11,692.
- Cornuelle, B.D., W.H. Munk, and P.F. Worcester, 1989. Ocean acoustic tomography from ships. *J. Geophys. Res.*, **94**, 6232-6250.
- Cornuelle, B.D., P.F. Worcester, J.A. Hildebrand, W.S. Hodgkiss Jr., T.F. Duda, J. Boyd, B.M. Howe, J.A. Mercer and R.C. Spindel, 1993. Ocean acoustic tomography at 1000-km range using wavefronts measured with a large-aperture vertical array. *J. Geophys. Res.*, **98**, 16,365-16,377.

- Cornuelle, B.D., C. Wunsch, D. Behringer, T.G. Birdsall, M.G. Brown, R. Heinmiller, R.A. Knox, K. Metzger, W.H. Munk, J.L. Spiesberger, R.C. Spindel, D.C. Webb and P.F. Worcester, 1985. Tomographic maps of the ocean mesoscale, 1: Pure acoustics. *J. Phys. Oceanogr.*, **15**, 133-152.
- Fukumori, I., and P. Malanotte-Rizzoli, 1995. An approximate Kalman filter for ocean data assimilation: An example with an idealized Gulf Stream model. *J. Geophys. Res.*, **100**, 6777-6793.
- Howe, B.M., P.F. Worcester and R.C. Spindel, 1987. Ocean acoustic tomography: Mesoscale velocity. *J. Geophys. Res.*, **92**, 3785-3805.
- Kak, A.C., and M. Slaney, 1988. *Principles of Computerized Tomographic Imaging*. IEEE Press, New York.
- Malanotte-Rizzoli, P., and W.R. Holland, 1986. Data constraints applied to models of the ocean general circulation, Part I: the steady case. *J. Phys. Oceanogr.*, **16**, 1665-1687.
- Munk, W., P.F. Worcester, and C. Wunsch, 1995. *Ocean Acoustic Tomography*. Cambridge Univ. Press, Cambridge.
- Munk, W., and C. Wunsch, 1979. Ocean acoustic tomography: A scheme for large scale monitoring. *Deep-Sea Res.*, **26**, 123-161.
- Munk, W., and C. Wunsch, 1982. Up/down resolution in ocean acoustic tomography. *Deep-Sea Res.*, **29**, 1415-1436.
- Ocean Tomography Group, 1982. A demonstration of ocean acoustic tomography. *Nature*, **299**, 121-125.
- Schröter, J., and C. Wunsch, 1986. Solution of nonlinear finite difference ocean models by optimization methods with sensitivity and observational strategy analysis. *J. Phys. Oceanogr.*, **16**, 1855-1874.
- Sheinbaum, J., 1989. *Assimilation of Oceanographic Data in Numerical Models*. Ph.D. Thesis, Univ. of Oxford, Oxford, England, 156 pp.
- Spiesberger, J.L., and K. Metzger Jr., 1991. Basin-scale tomography: A new tool for studying weather and climate. *J. Geophys. Res.*, **96**, 4869-4889.
- Worcester, P.F., B.D. Cornuelle, and R.C. Spindel, 1991. A review of ocean acoustic tomography: 1987-1990. *Reviews of Geophysics, Supplement, U.S. National Report to the International Union of Geodesy and Geophysics 1987-1990*, 557-570.
- Wunsch, C., 1990. Using data with models: Ill-posed problems. In: *Oceanographic and geophysical tomography: Proc. 50th Les Houches Ecole d'Ete de Physique Theorique and NATO ASI*, Y. Desaubies, A. Tarantola, and J. Zinn-Justin, eds., Elsevier Science Publishers, 203-248.