

COURSE 11

**PRACTICAL ASPECTS  
OF OCEAN ACOUSTIC TOMOGRAPHY**

B. CORNUELLE

*Mail code A-030  
Scripps Institution of Oceanography  
La Jolla, CA, 92093, USA*

*Y. Desaubies, A. Tarantola and J. Zinn-Justin, eds.  
Les Houches, Session L, 1988  
Tomographie Océanographique et Géophysique /  
Oceanographic and Geophysical Tomography  
© Elsevier Science Publishers B.V., 1990*

## Contents

1. Introduction	443
2. Resolution in a vertical slice	444
2.1. Introduction	444
2.2. Loop harmonics	445
2.3. Inverses	451
3. Moving ship tomography	452
3.1. Introduction	452
3.2. The projection-slice theorem	453
3.3. Examples	454
4. Time dependence	457
4.1. Introduction	457
4.2. General problem	458
4.3. The Kalman filter	460
4.4. Examples	461
References	463

## 1. Introduction

These lectures cover three main topics:

- 1 Resolution of ocean acoustic tomography in the vertical slice geometry
- 2 Ocean tomography using moving ships
- 3 Tomography in a time-dependent (ocean) medium

The discussion of the first two topics ignores any changes in the medium during the data collection. The first section reviews an analysis of the sensitivity of acoustic travel times to different scales of ocean variation, and illustrates it with simulated examples. It follows Cornuelle and Howe (1988) and re-uses several of the figures from that paper. The fundamental result is that although travel times from ocean acoustic rays are to some extent averages of the sound speed and current fields through which they pass, they are sensitive to ocean variations with scales equal to the ray double loop length, and to higher (shorter) harmonics, extending well into the internal wave band. This is consistent with the observed travel time effects of internal wave displacements, which have about 5% of the amplitude of the travel time perturbations expected from ocean eddy motions. Inverse simulations show that the sensitivity can be used to make reconstructions of the short-scale features, subject to the limitations of finite signal to noise ratio.

The second section describes moving ship tomography, in which one or more ships carry sources and/or receivers to achieve a high density of acoustic paths with relatively few sources (compared to moored instruments). The technique is explored using array design simulations in a simplified ocean. The simulation results are presented and compared using calculations of expected error from the least-squares inverse. Maps of expected error in both physical and spectral space were extensively used during the optimization, since both forms can provide insight. One of the unproven principles of the optimization was to smooth out the irregularities in the error maps in both spaces, striving for uniform resolution. Most of the results in this section are taken from Cornuelle, Munk, and Worcester (1988). The many paths available from a moving ship experiment can give very high resolution of the ocean in the simulations, although the good performance depends significantly on high-quality positioning (order 10 meters) for the

acoustic arrays, which is on the edge of what is currently achievable.

The final section covers the extension of the inverse method to time - dependent ocean structure (and data). The need to integrate successive days or weeks of tomographic data in a dynamically consistent fashion requires some form of time-dependent inverse. The discussion assumes that some model for the evolution of the medium is available, so the later data can be related to the earlier data using the dynamics specified in the model. If the dynamics are non-linear, then the total inverse problem is non-linear, even if the individual observations are linearly dependent on the model. Some simulations are presented for the case of linear dynamics, using tomographic data in a Kalman filter, which is an extension of standard least-squares into the time domain when a linearized predictive dynamic model is added to the inverse parameterization. The simulations suggest that, in the absence of strong advection, the effects of dynamics are minimal for a single moving ship ocean survey with one-way transmissions. Least square methods are sensitive to the residual differences between the predicted data and the observations, and so are good for model validation, which is an important goal of time-dependent inverses in ocean applications. The inverse based on the Kalman filter has the advantage that it produces diagnostics which correspond to standard least- squares measures of goodness-of-fit. The prediction errors and the residuals at each step of the fit provide a check on the performance of the dynamical forecast, and on the consistency of the entire process.

## 2. Resolution in a vertical slice

### 2.1. Introduction

An acoustic ray path in a range-independent ocean sound speed field is periodic over a characteristic distance called the “double-loop range”. The sensitivity of the ray travel time to an ocean feature depends on the projection of the ocean feature on the ray path. Features which match the ray periodicity strongly affect travel time compared to other components, so sampling by multipaths with different double-loop ranges has a simple interpretation in spectral terms. Most ocean sound speed profiles allow maximum double-loop ranges of less than 100 km, so ray travel times respond strongly to ocean features with wavelengths shorter than 100 km but are relatively insensitive to wavelengths between 100 km and the source-receiver range, although the spatial mean between source and receiver is well-measured.

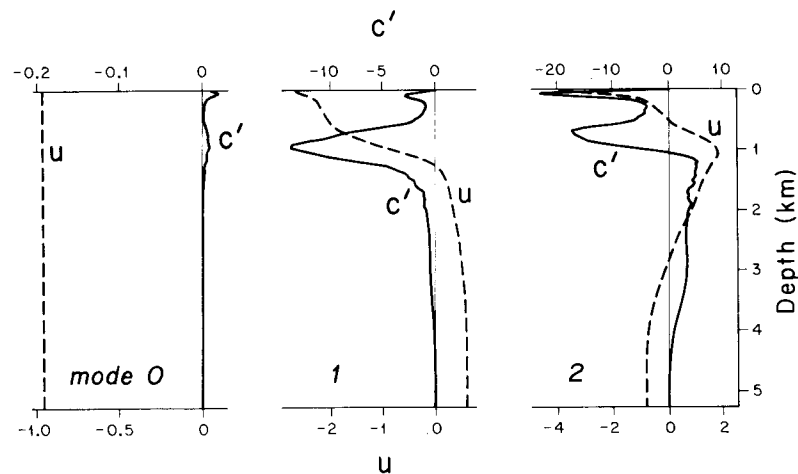


Fig. 1. Dynamical modes for sound speed and currents (from Howe, Worcester, and Spindel, 1987).

### 2.2. Loop harmonics

For transmissions between a single pair of moorings, the horizontal coordinate is  $r$ , the range. We assume that the reference state,  $c_0(z)$  is horizontally uniform. We parameterize the perturbation sound speed  $c'(r, z)$  using two sets of spectral functions: sinusoids in the horizontal and a limited set of special functions ( $F_i(z)$ ) in the vertical. These vertical structure functions (or “modes”) can be: (1) calculated analytically as solutions of the quasi-geostrophic vertical structure equation; (2) obtained statistically by computing Empirical Orthogonal Functions (EOFs) from a dataset; or (3) they may be picked arbitrarily, as in layers or sinusoids.

If true dynamical modes or EOFs are known, relatively few are needed to describe the vertical structure in a limited region. Many more of the arbitrary functions must generally be used to reproduce the ocean accurately, but the accuracy of the parameterization is less dependent on location. Because there are few good estimates of the actual ocean vertical covariance functions, and the simulations take place in limited areas, we choose the dynamically calculated modes to limit the number of model parameters. Figure 1 shows the set of dynamical modes for sound speed which parameterized the simulated ocean. The modes were calculated for the western North Atlantic Ocean; mode 0 is the barotropic (vertically uniform velocity) mode, which has negligible sound speed signature.

The same oceanographic data was used to calculate the sound speed profile in fig. 2., which shows sound speed in meters/sec versus depth.

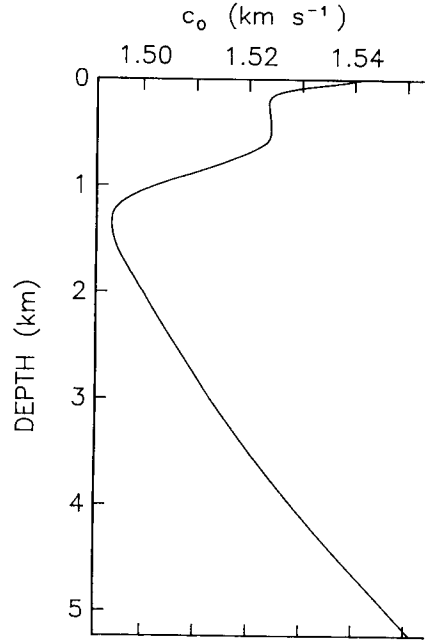


Fig. 2. Sound speed profile typical of the Sargasso Sea.

This analysis concentrates explicitly on sound speed perturbations, but if reciprocal rays were available, then the same results apply to the current sensitivity of travel time differences between oppositely directed ray paths. We parameterize the ocean perturbations as sinusoids in the horizontal and dynamical modes in the vertical so the sound speed anomaly (or, for the difference of reciprocal rays, current) at any point,  $(r, z)$  is

$$c'(r, z) = \sum_{i=1}^N \sum_{j=1}^M \left[ a_{ij} \cos\left(\frac{2\pi jr}{R}\right) + b_{ij} \sin\left(\frac{2\pi jr}{R}\right) \right] F_j(z) \quad (2.1)$$

where  $R$  is the range to the receiver (600 km in the simulations). This horizontal basis has fundamental period  $R$ , to simplify some of the algebra to come, but in principle it should have a fundamental larger than the extent of the data to avoid forcing periodicity. We write the ray depth versus range as  $\zeta_k(r)$ , so the travel time perturbation is

$$T'_k = - \int_0^R \frac{c'(r, \zeta_k(r)) \left[ 1 + \left( \frac{d\zeta_k}{dr} \right)^2 \right]^{1/2}}{c_0^2(\zeta_k(r))} dr + E_k \quad (2.2)$$

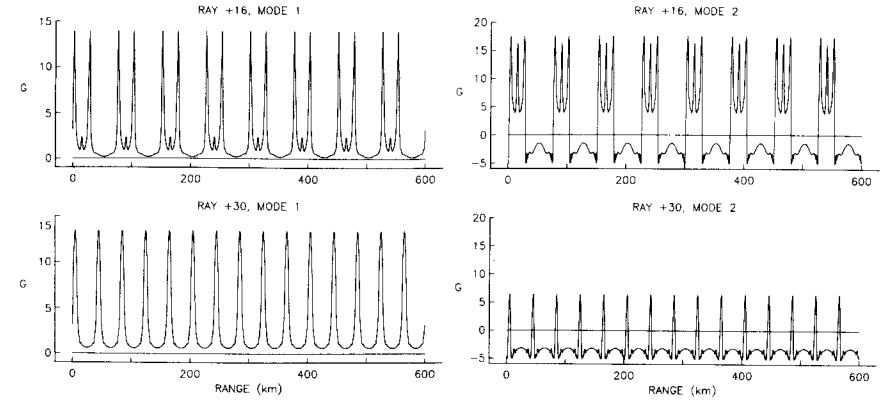


Fig. 3. Examples of ray weighting functions, for two rays and two modes, as indicated. (Units are arbitrary)

Substituting the expansion (1) into the forward problem (2) gives

$$T'_k = - \int_0^R \sum_{i=1}^N \sum_{j=1}^M \left[ a_{ij} \cos\left(\frac{2\pi jr}{R}\right) + b_{ij} \sin\left(\frac{2\pi jr}{R}\right) \right] G_{ik}(r) dr + E_k \quad (2.3)$$

Where  $G_{ik}(r)$  is given by

$$G_{ik}(r) = \frac{F_i(\zeta_k(r)) \left[ 1 + \left( \frac{d\zeta_k}{dr} \right)^2 \right]^{1/2}}{c_0^2(\zeta_k(r))} \quad (2.4)$$

This quantity, called the "ray weighting function" is the projection of the ray on a given vertical mode. Figure 3 shows ray weighting functions for two rays traced in the sound speed profile of fig. 2. The first (steep) ray has 16 turning points between source and receiver, making large excursions away from the sound channel axis, while the second (axial) has 30, remaining close to the axis.

A ray weighting function describing ray travel time variances due to internal waves was introduced for calculations of sensitivity of travel time

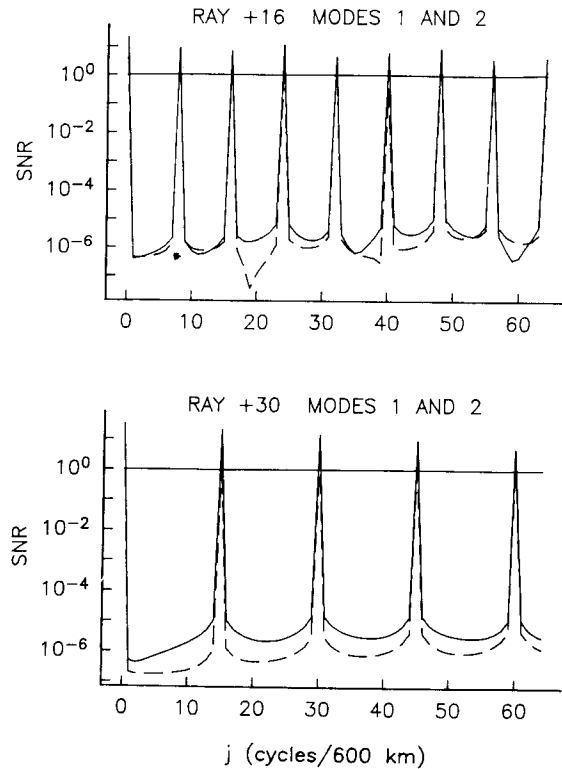


Fig. 4. Rms sensitivity spectrum for two particular rays.

variances to internal wave spectra by [Flatté and Stoughton, 1986]. The weighting function shows where mode  $i$  affects the travel time for ray  $k$ . The sensitivities of the travel times to the model parameters  $a_{ij}$  and  $b_{ij}$  are the cosine and sine transforms, respectively, of  $G_{ik}(r)$ . The fundamental periodicity of  $G$  is controlled by the ray loop length, and the combination of nonlinear vertical mode profile  $F_i(z)$  and non-sinusoidal ray path  $\zeta_k$  contribute to a rich spectrum of harmonics. (fig. 4).

Assume that the source and receiver are both on the sound channel axis and that the path of ray  $k$  is a sine in range around the sound channel axis,  $z_0$ , so depth as a function of range is given by:

$$\zeta_k(r) = \beta_k \sin\left(\frac{\pi k r}{R}\right) + z_0 \quad (2.5)$$

$k$  is the number of turning points for the ray, and  $\beta_k$  is distance of the upper and lower turning points from the axis. There are two possible rays

with  $k$  turning points, one with positive launch angle (above the axis), and the other with negative launch angle, but the second ray is redundant in this idealized case because the paths are exactly symmetric around the sound speed minimum. (In general, the sound speed minimum is not symmetric, and neither are the rays, so downward-going rays contribute useful information.)

Expand  $F_i$  as a polynomial in  $(z - z_0)$ ,

$$F_i(z) = \sum_{n=0}^N \alpha_n (z - z_0)^n \quad (2.6)$$

The variations of the denominator and arclength terms in eq. (4) are small compared to the variation in  $F_i(\zeta_k(r))$ ; neglecting them, the weighting function becomes

$$G_{ik}(z) = \sum_{n=0}^N \alpha_n \beta_k^n \left[ \sin\left(\frac{\pi k r}{R}\right) \right]^n \quad (2.7)$$

The dependence of travel time on the model parameters for the vertical mode  $i$  and wavenumber  $j$  is

$$\left(\frac{\partial T'_k}{\partial a_{ij}}\right)_n = - \sum_{n=0}^N \alpha_n \beta_k^n \int_0^R \cos\left(\frac{2\pi j r}{R}\right) \left[ \sin\left(\frac{\pi k r}{R}\right) \right]^n dr \quad (2.8)$$

and

$$\left(\frac{\partial T'_k}{\partial b_{ij}}\right)_n = - \sum_{n=0}^N \alpha_n \beta_k^n \int_0^R \sin\left(\frac{2\pi j r}{R}\right) \left[ \sin\left(\frac{\pi k r}{R}\right) \right]^n dr \quad (2.9)$$

If  $F_i$  is depth-independent, then  $N = 0$ , and only the cosine term with  $j = 0$  (the range average) affects the travel time. In other words, if mode  $i$  is depth independent, then the data ignore range-dependent structure of the mode, regardless of ray shape. Axial rays generally have the largest numbers of turning points ( $k$ ), with small excursions,  $\beta$ , so the dependence on  $\beta_k$  suggests a decline in sensitivity to range dependence as  $k$  increases for a given source-receiver pair. For the  $n = 1$  term with  $k$  even, the travel time contribution is zero for all model parameters except for the sine term  $b_{ij}$  (with  $j = k/2$ ). Sensitivity to cosine terms is obtained by adding a source and receiver at  $\beta_k$  above or below the axis. For  $n = 1$  and  $k$  odd,

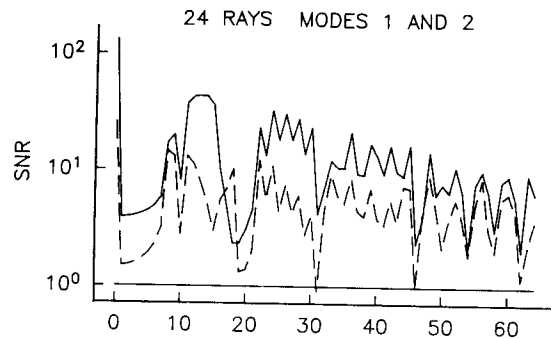


Fig. 5. Rms sensitivity spectrum for a 24-ray data set.

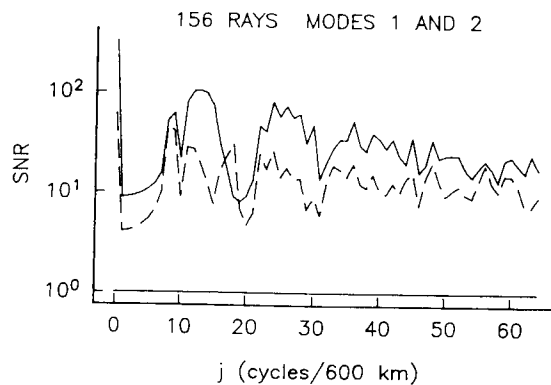


Fig. 6. Rms sensitivity spectrum for a 156-ray data set.

the periodicity of the  $\sin(\pi kr/R)$  term does not match the integration, and does not cancel, giving sensitivity to a range of cosine model parameters.

If  $n = 2$ , then there are contributions from  $a_{ij}$  for  $j = 0$  (the spatial mean term), and  $j = k$ , because the  $\sin^2(\pi kr/R)$  term is equivalent to  $(1 - \cos(2\pi kr/R))/2$ . This can be continued to all orders of  $n$ , giving sensitivity to increasingly high harmonics of range (and to the spatial mean). The limits of this high wavenumber information depend on the sharpness of  $G_{ik}(r)$ .

There is only weak travel time response to ocean perturbations with scales between the mean term ( $j = 0$ ) and the ray with the fewest turning points ( $j = k_{\min}/2$ ) (longest double loop length). The sharpness of the ray weighting functions (fig. 3) determines the sensitivity to very high wavenumbers, (far) above  $j = k_{\max}/2$ . The sharper and more irregular the vertical structure function, the more high-wavenumber features can affect the data. Mode 2 provides more high wavenumber sensitivity than mode 1, while only the horizontal average of the amplitude of the barotropic

(uniform) mode affects a ray travel time. Rays which are reflected from the surface or bottom should have a richer set of harmonics, and so should contribute more strongly to sensitivity to small scales.

Figure 5 shows the combined rms sensitivity spectrum for all 24 rays between a single source and receiver, separated by 600 km. The figure shows the sum of the squares of the harmonic components of each ray weighting function. The harmonics of the fundamentals (double loop lengths) begin to fill in some of the gaps at higher wavenumber.

A denser, 156 ray dataset, achieved by adding sources and receivers at other depths on both moorings gives a broader sensitivity spectrum (fig. 6), although the gap at low wavenumber remains.

### 2.3. Inverses

Inverse simulations using this enhanced dataset were made assuming a travel time noise of 1 ms and either a white spectrum for the ocean features, or a  $k^{-2}$  (red) spectrum (fig. 7).

The figure shows the resolution of each component of the model, a value of 1 means perfect determination, while a value of 0 means no information. Resolution falls off at high wavenumber in the simulation using a  $k^{-2}$  spectrum because of the white 1 ms measurement noise assumed.

When long time series of ray travel times are available from moored arrays, the ray sensitivity spectra present an opportunity to examine the wavenumber-frequency spectrum of the ocean. If the travel times for individual rays can be tracked throughout the dataset, then the variations can be band-passed to assign variance levels for each ray to each frequency band. The differing sensitivities (transfer functions) of each ray to the wavenumber spectrum of the ocean perturbations mean that the variances are weighted integrals of the (one-dimensional) wavenumber power spectrum in that band, so inverse methods can estimate the spectrum. Point measurements weight all wavenumber components equally, so constrain the mean power level across all frequency bands. If the spectrum is homogeneous and isotropic, then many rays from a horizontal array could be combined to give a better estimate of the spectrum.

This same analysis could be applied to geophysical rays if the spectrum of the inhomogeneities is assumed to be spherically symmetric. In that case, rays with similar ranges and turning depths (ray parameters) can be grouped together, and the variances around the mean travel time for each ray parameter give a different weighted integral of the power spectrum of the inhomogeneities in the earth. If enough rays with significantly different shapes were available, an estimate of the power spectrum of the

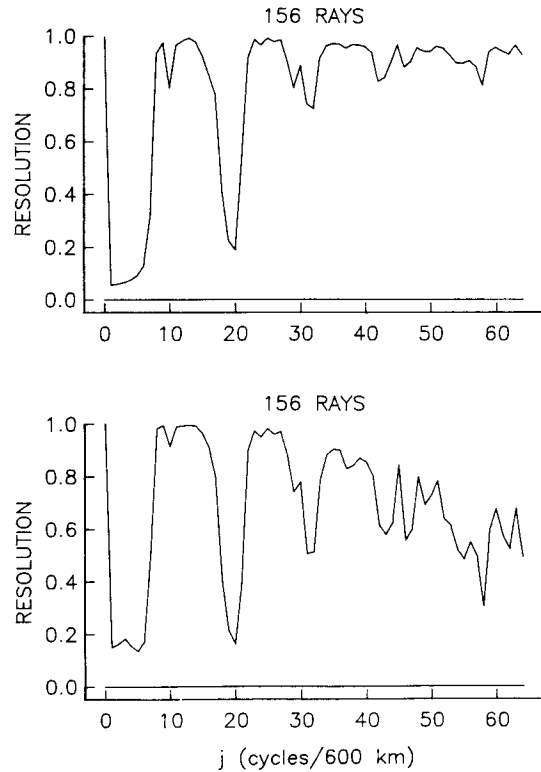


Fig. 7. Resolution as a function of wavenumber : For a white spectrum (top) and for a red spectrum (bottom).

inhomogeneities as a function of depth might be possible.

### 3. Moving ship tomography

#### 3.1. Introduction

At present, acoustic sources are too scarce to let experiments achieve the economies of scale that Munk and Wunsch (1979) hoped for. While the amount of data from a tomographic array can grow as fast as the square of the number of moorings, tomographic arrays perform poorly when only a few sources are available. Munk and Wunsch (1982) suggested an alternative : transmitting between ships steaming on parallel courses to produce scans of the ocean as in medical tomography. Unfortunately, it is difficult to envision enough ship scans to build a uniform high-resolution ocean map

over a large region. Instead, the most promising approach seems to be a compromise between moored arrays and moving ships, in which several transceivers are moored in an array and a ship steams around the array, “eavesdropping” on the transmissions with a dipped receiver, and thus obtaining a potentially large dataset. In order to understand the trade-offs in this type of experiment, many design simulations were undertaken by Cornuelle, Munk, and Worcester (1989) using several different configurations. The results show promise for very high resolution ocean snapshots, supplementing the integrated observations from the moored array.

#### 3.2. The projection-slice theorem

If two ships proceed in parallel down the sides of a box, transmitting continuously, the ray paths are flat, and the ocean has no vertical structure, then the travel time is a function of distance along the ship track, and is the Radon transform of the two-dimensional sound speed field inside the box. For simplicity, assume the ship track is East-West ( $x$ ), and the box is  $X$  by  $Y$  in extent, so the ray paths are North-South, and the travel time  $T(x)$  along the track is related to the sound speed  $c(x, y)$  by

$$T(x) = \int_0^Y c(x, y) dy \quad (3.1)$$

The sound speed can be parameterized by harmonics of the domain ( $X$  and  $Y$ ),

$$c(x, y) = \sum_{k=1}^N \sum_{l=1}^M a_{kl} \cos \left( 2\pi \left[ \frac{kx}{X} + \frac{ly}{Y} \right] \right) + b_{kl} \sin \left( 2\pi \left[ \frac{kx}{X} + \frac{ly}{Y} \right] \right) \quad (3.2)$$

The line integral in the  $y$  direction averages out all but the average component in  $y$ , so substituting eq. (3.2) into eq. (3.1) produces

$$T(x) = \sum_{k=1}^N a_{k0} \cos \left( 2\pi \left[ \frac{kx}{X} \right] \right) + b_{k0} \sin \left( 2\pi \left[ \frac{kx}{X} \right] \right) \quad (3.3)$$

The Fourier transform of  $T(x)$  yields the amplitudes of the components of the unknown field for which  $l = 0$ . If the scan is repeated at another angle, then  $x$  and  $y$  can be re-defined, so that eq. (3.3) holds in the new coordinates; the Fourier transform of a scan in one direction gives the values

of the transform of the unknown field along the corresponding direction in spectral ( $k, l$ ) space. This result is called the projection-slice theorem, and was used as the reconstruction method in early medical imaging. A single scan gives information along a single line in spectral space, so many scans are required before the field is reproduced well.

### 3.3. Examples

The projection-slice theorem has dire consequences for practical ocean mapping, illustrated by fig. 8, which shows error maps in physical and spectral space corresponding to ship-to-ship tomography experiments. The top panel is the known ocean perturbation for the simulations, plotted as sound speed (m/s) at a representative depth in a 1000 km by 1000km domain. The left column is a representation of the ray paths between ship tracks, the middle column is the physical-space reconstruction of the true perturbation from the ship scans shown on the left, and the right column shows the percentage error for each of the spectral components used to represent the ocean field. The overall error variance ( $\sigma$ ) is in the plot label, expressed as a fraction of the variance of the perturbation field. In all the simulations in this section, the travel time measurements were assumed to have errors of 10 ms.

In converting to the more realistic ocean case, with a sound channel and a three dimensional sound speed field, the situation gets both better and worse. On the one hand, the sound channel generally produces many ray paths between a given source and receiver, and if each ship used multiple sources and receivers, a dense slice sampling could be built up. On the other hand, the reconstruction must include the vertical structure of the unknown field, which, if it is complicated, may more than absorb the extra information obtained from the multipaths. Assuming that the multipaths dominate over the extra variables, the discussion in the first section showed that the acoustic rays do not simply average sound speed, but perform a sort of Fourier analysis of features smaller than the maximum double loop range. This means that in the three-dimensional version of the ship-to-ship case illustrated in fig. 8 there is information about the high wavenumbers (which have been ignored in the models used in the simulation). The high east-west wavenumbers could then be reconstructed from a north-south pass, and vice versa.

If the size of the experiment is scaled down until the box is a double-loop length across, then a single ship-to-ship scan can in principle completely determine the sound speed, provided enough sources and receivers are deployed underneath each ship. If the transmissions are reciprocal, then the

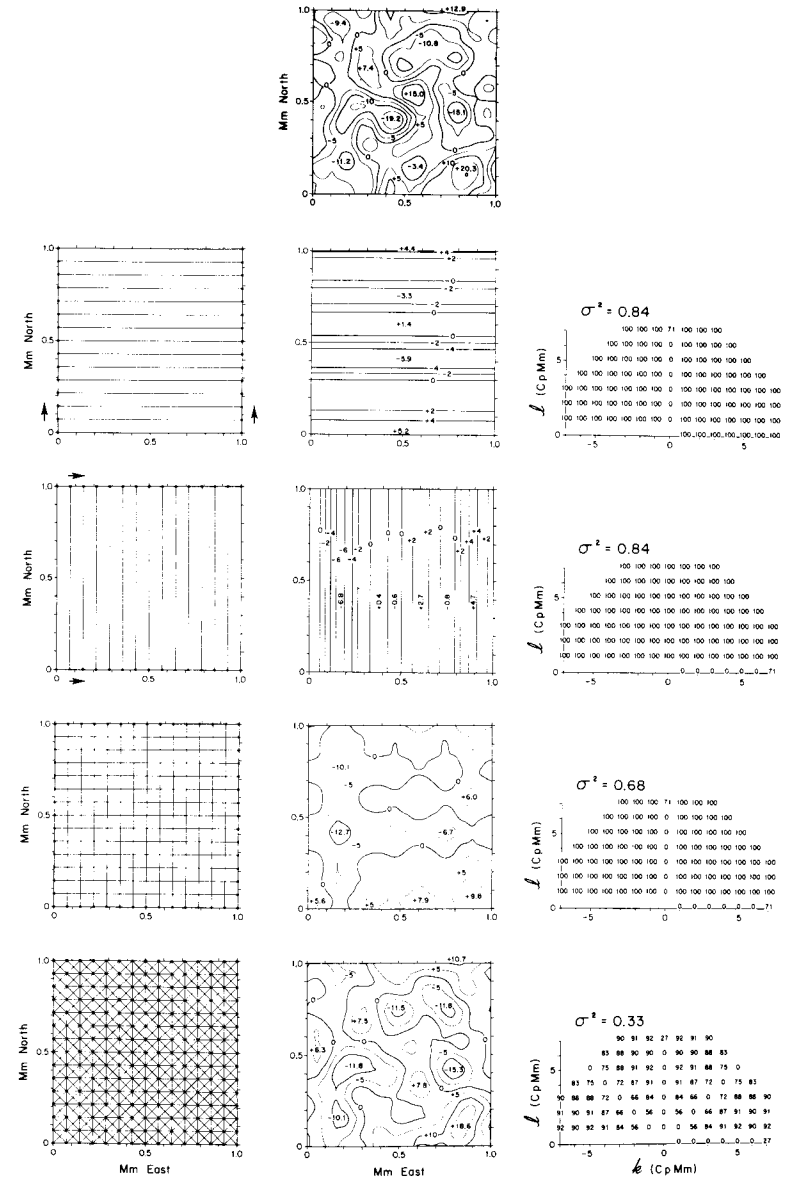


Fig. 8. Illustration of the projection-slice theorem

procedure could in principle determine both sound speed and the component of current tangent to the ray paths. Two scans would be needed to completely determine the velocity field, even in the small experiment.

unless non-divergence was used as a constraint.

Mapping a 1000 by 1000 km box is tedious using only two ships, and mooring deployed within the box improve the reconstructions significantly. figure 9 shows results for a single ship steaming around the outside of a box containing 4 sources. The top figure is a schematic top view of the ray geometry, the left-most plot in the second row is again the true field (contours of sound speed (m/s)), the middle panel in the second row is the reconstruction, and the right-most plot is the difference between the two. The bottom row shows contours of spectral energy level in spectral space (left panel), with dots showing the spectral components used in the ocean representation. The middle panel is the expected error variance of the inverse reconstruction, plotted in spectral space, while the right panel is the standard deviation of the expected error in physical space (also in m/s).

The (two-dimensional) ocean is relatively well-mapped (total error variance is a few percent of the initial variance) with only 4 moorings, and adding two extra moorings improves the maps to error levels below one percent, beyond the limits of the simulations. The array design shown in fig. 9 was chosen as a result of optimization experiments in which the configuration and number of central moorings were varied in order to minimize the overall percentage error variance in the 1000 km box, which is the fraction of the expected variance of the field which is not reconstructed, and is noted as  $\sigma^2$  in the figures. Optimization of this measure of goodness seems to coincide with obtaining the smoothest error maps in both physical and spectral space.

These simulation results would apply to a static array (in which all instruments were moored) as well, but the ship borne receivers seem to be the most practical way to obtain this level of resolution. There are many technical problems to overcome in a moving experiment that have been ignored so far. Birdsall and Metzger (personal communication, 1988) have been able to correct for the effects of doppler shift on the acoustic transmissions, and the availability of precise satellite positioning at sea means that less information has to be diverted into solving for the positions of the instruments. The effects of positioning errors were included in the simulations, which showed that as long as the travel time error due to the positioning uncertainty is small compared to the measurement error, it has little effect, and the effects saturate for large uncertainties. For measurement errors in the practical range (10 ms), the positioning error can make a significant difference if it is larger than 10 m.

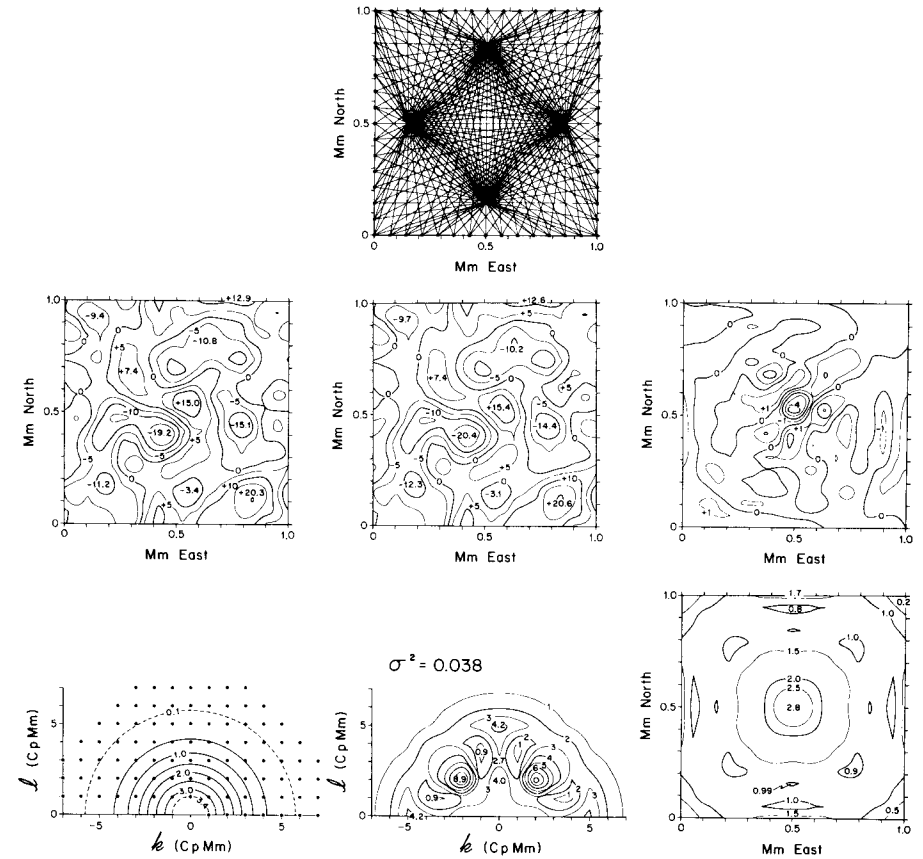


Fig. 9. Mapping a field with moving ship

## 4. Time dependence

### 4.1. Introduction

Most of the inverse reconstruction methods which have been used in ocean acoustic tomography parameterize the ocean with a set of functions. They discretize the three-dimensional ocean into tiny boxes or expand in sinusoids, with the goal of making the parameterization sufficiently dense that it does not change the character of the unknown field. For example, parameterizing an ocean into a few large boxes is not acceptable if the true scales of variation are smaller, (regardless of the amount of data available), although Munk and Wunsch (1979) found it useful as a teaching tool. Discretization simplifies the reconstruction problem, which becomes a finite-dimensional

problem of best fitting the model parameters (the ocean structure) to the data (the travel times).

If the parameterization extends into time, then time-dependence does not qualitatively change the problem. If, for example, ocean perturbations were made up of linear (very weak) planetary waves, and the waves were used as the discretizing basis, giving sinuoidal dependence in  $x$ ,  $y$ , and  $t$ , then the reconstruction can be done by fitting the data to the amplitudes of the selected components (Chiu and Desaubies, 1987). Unfortunately, the expected time evolution of the ocean is generally not available in closed form as part of the discretization, but is constrained by equations of motion which must be integrated by time stepping from initial conditions using boundary conditions, often by a large and complicated numerical code.

#### 4.2. General problem

If the data can be approximated as instantaneous with respect to the evolution of the ocean, then the kernel for the tomographic observations can be adapted to the pointwise discrete time representation. Let  $\mathbf{m}(t)$  be the vector of values that describes the true state of the ocean at time  $t$  (which is assumed to vary in steps). In the current application,  $\mathbf{m}(t)$  would be the set of amplitudes for the spatial basis functions at timestep  $t$ . The observations, organized into the vector  $\mathbf{d}(t)$ , are related to the model parameters, by the time-independent equation:

$$\mathbf{d}(t) = \mathbf{s}(\mathbf{m}(t)) + \mathbf{q}(t) \quad (4.1)$$

$\mathbf{q}(t)$  is observational noise, comprising instrument noise and errors incurred due to the finite parameterization. (When the basis is chosen, all scales which are not included become “noise”. This is natural when, for example, the object is to model eddy processes and to filter out internal waves.)

If a discretized numerical or analytical model exists for the ocean, and we adopt its parameterization (basis functions) to represent the field in the reconstruction as well, then  $\mathbf{m}(t)$  can be expressed as a function of the initial conditions  $\mathbf{m}(t_0)$  and the boundary conditions  $[\mathbf{b}(t_0), \mathbf{b}(t_1), \dots]$  over the integration time. ( $\mathbf{m}(t)$  may also depend on parameters internal to the model, such as viscosity or layer depths.) Concatenating all these vectors together into a single vector,  $\mathbf{h}$ , which is not a function of time, we can write

$$\mathbf{m}(t) = F(\mathbf{h}, t) \quad (4.2)$$

so (4.1) becomes a function of time and initial/boundary conditions:

$$\mathbf{d}(t) = \mathbf{T}(\mathbf{h}, t) + \mathbf{q}(t) \quad (4.3)$$

and the total reconstruction problem is a nonlinear fit for  $\mathbf{h}$  to the concatenated dataset  $[\mathbf{d}(t_0), \mathbf{d}(t_1), \dots]$ . Most techniques attempt to linearize this problem around some model trajectory,  $\mathbf{m}_0(t)$ , the time evolution of the model state vector corresponding to the initial/boundary conditions  $\mathbf{h}_0$ . Denoting  $\mathbf{h}' = \mathbf{h} - \mathbf{h}_0$  and  $\mathbf{d}'(t) = \mathbf{d}(t) - \mathbf{T}(\mathbf{h}_0, t)$ , one can obtain a linearized relation (for sufficiently small  $\mathbf{h}'$ ):

$$\mathbf{D}' = \mathbf{B} \cdot \mathbf{h}' + \mathbf{Q} \quad (4.4)$$

where  $\mathbf{D}'$  is the concatenation of the differences of the observations from those predicted by the reference trajectory and  $\mathbf{Q}$  is the concatenation of the data errors.  $\mathbf{B}$  is the matrix of dependence of data on initial conditions and boundary conditions.

Le Dimet and Talagrand (1985) show that it is possible to compute  $\mathbf{B}$  efficiently by integrating the adjoint of the model equations back in time. If this is not possible or convenient, the dependences must be calculated by brute force, which can require many model runs. In a chaotic system, the trajectory can become highly sensitive to the initial and boundary conditions, in which case some elements of  $\mathbf{B}$  (the Fréchet derivative) do not exist. If the reference trajectory is not sufficiently close to the true trajectory, then the exact least-squares solution may not be desirable, so the complete matrix  $\mathbf{B}$  may not be useful, and some form of descent or conjugate-gradient iteration (Tarantola, 1988) may be preferable. When there is no way to pick a reasonable reference trajectory the fitting problem can be strongly nonlinear, and, barring special structure of the problem, random searches may be necessary, such as simulated annealing or Monte Carlo inversion (Tarantola, 1988).

If the dynamics and forward problem are restricted to be linear, then elementary control theory offers two least-squares solutions to the reconstruction problem, which have been discussed by C. Wunsch in Course 5. The method that will be used to generate examples for this discussion is the Kalman filter (KF) which has been implemented in a linear ocean model with discrete timesteps. The purpose of the examples is to show some effects of ocean physics on the information available from an observing array. The KF is basically linear, although it can be applied to nonlinear problems provided a good reference trajectory is available. The KF uses least-squares formalism, which assumes that all fields are jointly Gaussian random variables with means and covariances known in advance (before the data are taken). The principal advantages of this approach are its relative simplicity and the error estimates which can be constructed along with the reconstructions.

### 4.3. The Kalman filter

The discrete-time Kalman filter can be cast as the extension of weighted least-squares estimation to the case where data are available at successive timesteps with linearized dynamics between steps. The following is not meant as a derivation, but as a justification in terms of the familiar single time inverse. We re-use the notation from above, but assume that the forward problem and the dynamics have been linearized around a reference trajectory and drop the primes from the variables. The observations, (with the predictions from the reference trajectory removed),  $\mathbf{d}(t)$ , are related to the model parameters by a kernel which is discretized as a matrix,  $S(t)$ :

$$\mathbf{d}(t) = S(t) \cdot \mathbf{m}(t) + \mathbf{q}(t) \quad (4.5)$$

Both  $\mathbf{m}(t)$  and  $\mathbf{q}(t)$  are assumed to have mean  $\tilde{\mathbf{m}}(t)$  and zero mean, respectively, and to be normally distributed with covariances  $\tilde{P}(t)$  and  $Q(t)$ , respectively.  $\tilde{P}(t)$  is the expected error in the estimate of the model parameter (perturbation) vector at time  $t$ , so the statistics are relative to  $\tilde{\mathbf{m}}(t)$ . The data errors are assumed to be uncorrelated with the model parameter errors. The matrix  $S(t)$  is time varying because the same data may not be available at every timestep.

If  $\mathbf{m}(t)$  is unrelated between timesteps, then eq. (4.5) is inverted by standard least-squares, so the new estimated state  $\hat{\mathbf{m}}(t)$  is

$$\hat{\mathbf{m}}(t) = \tilde{\mathbf{m}}(t) + \tilde{P}(t)S(t)^T(S\tilde{P}S^T + Q(t))^{-1} \cdot (\mathbf{d}(t) - S \cdot \tilde{\mathbf{m}}(t)) \quad (4.6)$$

if there are fewer data than model parameters, (formally underdetermined) or

$$\begin{aligned} \hat{\mathbf{m}}(t) = & \tilde{\mathbf{m}}(t) \\ & + (S(t)^T Q(t)^{-1} S(t) + \tilde{P}(t)^{-1})^{-1} \cdot S(t)^T Q(t)^{-1} \cdot (\mathbf{d}(t) - S \cdot \tilde{\mathbf{m}}(t)) \end{aligned} \quad (4.7)$$

if there are more data (formally overdetermined).

The error estimates are

$$\hat{P}(t) = \tilde{P}(t) - \tilde{P}(t)S(t)^T \cdot (S\tilde{P}S^T + Q(t))^{-1} \cdot S(t)\tilde{P}(t) \quad (4.8)$$

(formally underdetermined), or

$$\hat{P}(t) = (S(t)^T Q(t)^{-1} S(t) + \tilde{P}(t)^{-1})^{-1} \quad (4.9)$$

(formally overdetermined). Note that the latter requires the inversion of the model parameter error covariance separately, without the regularizing of observational noise.

If there is an estimate for the dynamics of the evolution of  $\mathbf{m}(t)$ , let  $A(t)$  represent the linearization of the evolution equation around  $\tilde{\mathbf{m}}(t)$ .  $A(t)$  is a linear transformation which predicts  $\tilde{\mathbf{m}}(t+1)$  (the initial guess for the next step) given  $\tilde{\mathbf{m}} \cdot (t)$  (the final estimate from the current step), with error  $\mathbf{r}(t)$ .

$$\tilde{\mathbf{m}}(t+1) = A(t) \cdot \tilde{\mathbf{m}}(t) + \mathbf{r}(t) \quad (4.10)$$

$\mathbf{r}(t)$  can be due to the dynamical effects of the ignored scales, to violations of the assumptions used to make the evolution equations, or to errors and instabilities in the numerical model code, to name a few. The error  $\mathbf{r}(t)$  is assumed to be normal with covariance  $R(t)$ , to be uncorrelated with the measurement noise, and to be known in advance. (If  $\mathbf{r}(t)$  is due to errors in the dynamics of the model, it may contain significant systematic errors, invalidating the assumption.) The estimation, eq. (4.6) or (4.7), together with the error estimates eqs. (4.8) and (4.9), and the prediction (4.10), form a complete cycle, with the addition of the prediction of the model parameter error covariance matrix,  $\hat{P}(t)$ , by the linearized dynamics:

$$\hat{P}(t+1) = A(t)\hat{P}(t)A(t)^T + R(t) \quad (4.11)$$

In models with large numbers of state variables, the calculation and prediction of the parameter error covariance dominates the expense, and there are many approaches to streamlining the prediction.

### 4.4. Examples

If the dynamics of the medium are perfectly duplicated by the model, then no information is lost between timesteps, and the time dependence will distort the error maps without degrading them. figure 10 shows error maps generated for a moving ship experiment similar to those shown earlier, but including the effects of changes in the medium during the data acquisition. The ship carrying the listening array takes 12 days to steam around the perimeter of the box, during which time the ocean is allowed to evolve by propagation of linear planetary waves and advection by a barotropic, horizontally uniform flow. Because one-way transmissions directly observe only the baroclinic field, which has wave periods longer than 200 days, the dynamics have little effect on the maps if there is no advection.

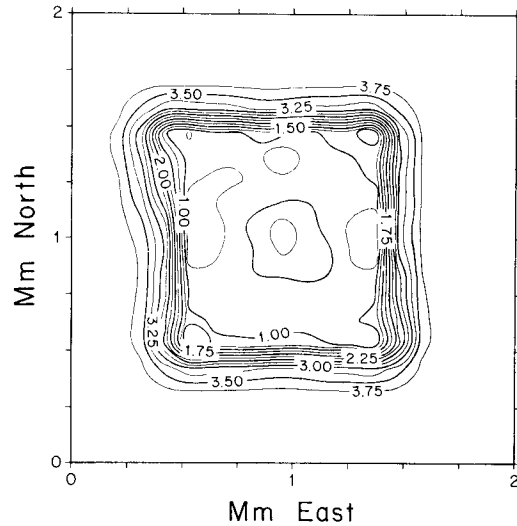


Fig. 10. Error map for a 10 cm/sec westward flow.

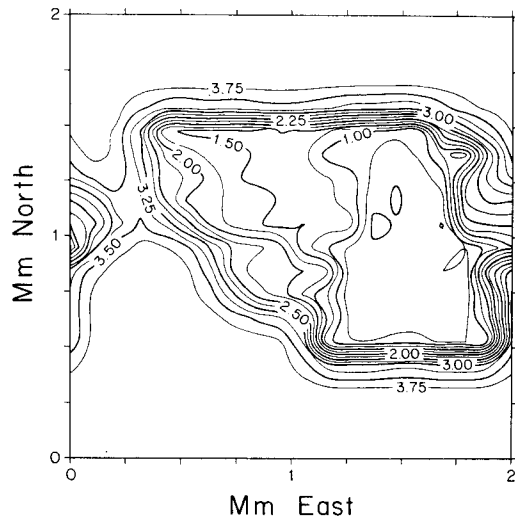


Fig. 11. Error map for a 1 m/sec current flowing eastward.

Figures 10 and 11 show maps of the standard deviation of the expected reconstruction error (m/s) from a moving ship simulation in which the ocean features are advected by a uniform flow during the experiment.

Both simulations use a 2000 by 2000 km domain, in an attempt to avoid the effects of periodicity at the boundaries, although the strong advection

case moves through the boundaries. The dynamics spread the influence of the data over a larger region, without reducing the overall percentage error of the reconstruction.

## References

- Chiu, C.S., 1985. Estimation of Planetary Wave Parameters from the Data of the 1981 Ocean Acoustic Tomography Experiment. PhD Thesis, MIT/WHOI, 255 pp.
- Cornuelle, B., and B. Howe, 1987. High spatial resolution in vertical slice ocean acoustic tomography. *J. Geophys. Res.*, **92**, 11680–11692.
- Cornuelle, B., W. Munk, and P. Worcester, 1989. Ocean acoustic tomography from ships. *J. Geophys. Res.*, **94**, 6232–6250.
- Flatté, S.M. and R.B. Stoughton, 1986. Theory of Acoustic Measurement of Internal Wave Strength as a Function of Depth, Horizontal Position, and Time. *J. Geophys. Res.*, **91**, 7709–7721.
- Howe, B.M., P.F. Worcester and R.C. Spindel, 1987. Ocean Acoustic Tomography : Mesoscale Velocity. *J. Geophys. Res.*, **92**, 3785–3806.
- Le Dimet, F., and O. Talagrand, 1986. Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects. *Tellus* **38A**, 97–110.
- Munk, W.H. and C. Wunsch, 1979. Ocean acoustic tomography : A scheme for large scale monitoring. *Deep Sea Res.*, **26A**, 123–160.
- Munk, W.H. and C. Wunsch, 1982. Observing the oceans in the 1990s. *Phil. Trans. R. Soc. Lond.* **A307**, 439–464.
- Tarantola, A., 1988. Inverse problem theory ; methods for data fitting and model parameter estimation, (Elsevier Science Publishers B.V., Amsterdam).