Please refer to this paper as:

Nolet, G., Imaging the deep earth: technical possibilities and theoretical limitations, in: Proc. XXIIth Assembly ESC, Barcelona 1990 (ed. A. Roca), 107-115, 1991.

PROCEEDINGS XXII ASSEMBLY E.S.C. (BARCELONA, 1990), ED. A. ROCA, D. MAYER-ROSA, IUGG, 1990.



# IMAGING THE DEEP EARTH: TECHNICAL POSSIBILITIES AND THEORETICAL LIMITATIONS

G. Nolet

Department of Theoretical Geophysics
University of Utrecht, The Netherlands

## Introduction

Without seismological data, we would know very little about the structure of the Earth's interior. In fact, the basic principles of interpretation that we apply nowadays where already present in Oldham's famous 1906 paper. In this paper Oldham infers the existence of the core as a low velocity body in the interior of the Earth from a few scattered seismic arrivals. It was essentially the analysis of the wave paths, which he approximated by straight lines, that led him to his discovery. The interpretation methods have been perfected since then, and especially in the past 15 years we have seen the rise of a new technique of seismological interpretation which was named 'seismic tomography'. In this paper, which briefly summarizes the ESC symposium lecture, I wish to explore the practical and theoretical limitations of the method of seismic tomography, and shall also briefly speculate on methods to overcome the practical limitations. Let me first anticipate the major conclusions:

Seismic tomography can only be pushed to the limit of its possibilities if we are willing to start major observational programs:

- we must organize digital recording in a dense grid of stations on the continents,
- we must provide an efficient data storage and distribution network which ensures that important digital data will not get lost
- a major effort must be undertaken to measure delay times of seismic waves in the oceans.

The European Seismological Commission has a clear task in getting these efforts going in Europe.

## The precision of seismic tomography

In body wave tomography, one generally linearizes the formal connection between the predicted travel time delay **d** and the model of velocity- or slowness variations **x**:

$$\mathbf{Ax=d}$$

The matrix is not square, since in general we have more data then unknowns. This leads to two problems:

First, the linear equations may be inconsistent because of errors. We make errors in reading the arrival times of waves, but there are also errors in the matrix elements, since we used ray theory and first order perturbations to formulate the linear system. Instead of computing the solution of this system, we must be satisfied by computing the best fit - in some sense - to these inaccurate equations.

Second, the problem is underdetermined. In other words, there is more than one model, that gives the same best fit to the equations. In fact there are infinitely many such models. We use some form of damping to give us a unique solution and reduce the influence of errors. In practice however, damping is not fully adequate, and our solution is often very much influenced by the data errors.

The early tomographic methods used an exact damped least squares technique. However, the numerical scheme used in these early studies posed a limitation to the size of the problem. The pioneering tomographic studies of Aki et al. (1977) and by Dziewonski et al. (1977) assumed that the least squares system can be solved exactly, which practically requires us to store a complete matrix in memory. This can be done for some 1000 unknowns or cells, but becomes difficult for larger models, even with present day computers. 1000 cells is not much in 3 dimensions, and at present we use iterative matrix solvers to do the inversion. Such algorithms have been the subject of much research in numerical analysis, and iterative techniques like the ones by Gauss and Seidel predate their application in seismology by many years.

There were essentially two ways to go. In the biological and medical sciences the method of Kaczmarc which dated back to 1939 was reinvented and given a new name: ART, for Algebraic Reconstruction Technique. ART however, not only proved slow but is also badly conditioned for the typical situation in seismology: unlike the well-controlled situation of medical X-ray tomography, delays are highly imprecise and seismic rays are distributed in a very irregular way and this influences the solution dramatically.

Following early work by Dines and Lytle (1979), Clayton and Comer (1983) and others at Caltech tried different methods that included an averaging scheme and were more sensitive for cells with only a very poor ray coverage. Their method proved to be a variant of a method that was known to mathematicians as SIRT, for Simultaneous Iterative Reconstruction Technique. At the same time, Nolet (1983) showed that the method of linear Conjugate Gradients could be applied to the problem. I used an algorithm developed in the previous year by Paige and Saunders (1982) and which was named LSQR. Both LSQR and SIRT methods are nowadays widely applied in large scale tomography.

The use of iterative techniques does complicate the error analysis of seismic tomography. In any iterative method, the solution is a linear combination of the data: $\mathbf{x}=\mathbf{A}^{-}\mathbf{d}$, where $\mathbf{A}^{-}$ is (hopefully) close to a generalized inverse of $\mathbf{A}$. The error in the final model $\mathbf{x}$ is made up of two parts (e.g. Nolet, 1987):

$$\mathbf{x}-\mathbf{x}_{true}=(\mathbf{A}^{-}\mathbf{A}-\mathbf{I})\mathbf{x}_{true}+\mathbf{A}^{-}\mathbf{e}$$

where $\mathbf{e}$ is the unknown vector of errors in the data $\mathbf{d}$. We see that ther errors propagate as $\mathbf{A}^{-}\mathbf{e}$. This factor can be kept under control by making sure that $\mathbf{A}^{-}$ has no large eigenvalues: these would 'blow up' all components of $\mathbf{e}$ in the direction of the corresponding eigenvectors. Various strategies exist that avoid such unstable situations. Damping is a well-known ad-hoc technique do keep error propagation under control, but much more sophisticated methods are well grounded in a Bayesian statistical framework (Tarantola, 1987). Whatever method is used to reduce the variance of $\mathbf{x}$, the price paid is in the second error term: if we tamper with the eigenvalues of $\mathbf{A}^{-}$, this matrix will less resemble the true generalized inverse for which $||\mathbf{A}^{-}\mathbf{A}-\mathbf{I}||$ is optimally small.

One complication is that many iterative techniques include some invisible weighting of the matrix system which influences the eigenvalues. Another is that systems are so large that we generally have to stop iterating before optimum convergence is reached. Thus, $\mathbf{A}^{-}\mathbf{A}$ is not equal to $\mathbf{I}$ because of lack of resolution *and* because of limited available CPU time. In view of that fact it is most important to use the iterative method with the fastest convergence available. Nolet (1983, 1985), Spakman and Nolet (1987) and Van der Sluis and Van der Vorst (1987) have all shown that LSQR has superior convergence properties. Since it also avoids an arbitrary scaling of the system, it should be accepted as the preferred method for seismic tomography.



How large is $\mathbf{e}$? A pessimistic, but probably quite realistic view is that most of the delay time as provided by the ISC is actually error, and only small part of it is signal. In a recent paper Gudmundsson et al. (1990) try to estimate the statistical properties of the ISC delays by studying average delays in bundles of closely neighbouring rays - we call them summary rays - and by studying asymptotic behaviour of the variance as the cross section of the bundle goes to zero. They find a standard deviation between 1 and 1.4 s for rays bottomming in the Upper Mantle, which decreases to a minimum of 0.5 s at teleseismic distances and increases again to 1 s when the ray reaches the CMB. If events are less carefully selected as in this study - and usually we cannot be too fastidious since we need every bit of information we can get - the S/N ratio will be below 1 over all depth ranges. Gudmundsson interpretes these results in terms of higher heterogeneity in the Upper Mantle and the D" layer.

The resolution and error propagation can - and should - always be checked by running the inversion algorithm on synthetic data as well, and checking the resolving power and the error in the synthetic result (e.g. Spakman and Nolet, 1988). If a well defined hypothesis is tested, the use of synthetic calculations may check on the ability of the tomographic analysis to accept or reject the hypothesis. This was the approach taken by Spakman et al. (1989) for the hypothesis that slabs penetrate into the lower mantle and seems to be a very promising method.

Always the best way to handle the influence of errors is to reduce the errors in the observations, rather than use more data to reduce the variance of the mean. I wish to remark that, generally speaking, one needs 100 times as many rays to reduce the average error in a cell by a factor of 10. The same effect can be obtained by measuring the delay time with a precision that is 10 times better. Using digital, broad-band instrumentation such an increase of precision is within reach (e.g. VandeCar and Crosson, 1990). The square root behaviour of standard deviations is well known to us from simple averaging of experimental values, but it can be shown to hold as well in the case of the ill posed tomographic problem with an approximate iterative solution. From this we should conclude that it is more practical and rewarding to improve the precision of the arrival time readings than to wait for more earthquakes to happen.

**The resolution of seismic tomography**

The fact that $\mathbf{A}^{-}\mathbf{A}\neq\mathbf{I}$ is due to the greatly inadequate sampling of the earth by seismic rays. This is the practical limitation that most halts progress in seismic tomography: only a few regions on Earth possess the station density and the seismicity which provides an adequate ray coverage of the interior. Even worse, stations reporting to the ISC are mostly located on land, with a heavy emphasis on Western Europe, very few stations in the southern hemisphere and almost none in oceanic regions.

How do we get more raypath through oceanic regions? We cannot organize the *seismicity*, although we can be fairly sure that our event coverage becomes more and more ideal as we wait long enough to catch more of the intraplate earthquakes. But we can organize the *observation* of these events. We should wish to have as many seismic stations on the ocean bottom as we have on land - but it is obvious we cannot pay for those. And even if we can place them, the cost of monitoring more than just a handful of them would become prohibitive. Does this mean we cannot obtain delay time data with a reasonable density from oceanic areas?

Marine acoustics is concerned with measuring noise in the ocean. Recent developments have brought these measurements at subsonic frequencies, close to where we expect earthquake signals. I undertook a simple search among the tapes of some marine acoustic experiments. Figure 1 shows the strongest of these events, a magnitude 6.3 earthquake on May 6, 1987 in the Andreanoff Islands region in the middle of the Aleutians, recorded by a floating seismograph in the Pacific near California at a distance of 47°. This instrument was operated by Gerald d'Spain of the Marine

Geophysics group at Scripps. Analysis of the power spectrum of this signal as a function of time shows the P wave very strong in the 2-3 Hz band (Figure 2). This is surprisingly high, and encouraging since ocean noise is lower at higher frequencies. However, I failed to find recordings from weaker earthquakes in records from hydrophones suspended from buoys at the surface. Why was the Andreanoff earthquake signal recorded and the others not?

The problem with recording in the oceans is that the water motion by itself is a tremendous noise factor - it accellerates the seismograph, or it brings a hydrophone on a different pressure level and both are recorded as signal. Fortunately, most of the water movement is concentrated in the waves near the surface, and this motion damps away exponentially with depth. However, the earthquake signals are small. Table 1 shows my calculations of the pressure signal associated with an earthquake at 90° and 1 Hz. They can be viewed as minimum levels. The ambient noise away from the surface is a few tenths of a Pascal at 1 Hz, and this is obviously not the prime noise factor. It is the effect of water motion on the instrument and cables that is deteriorating the signal. If the sensor moves only 1 mm up or down, this gives a signal of 10 Pa, in which even a magnitude 6 signal would drown.

Several groups have - on and off - worked at devising mechanical suspensions which damp the motion. This is a technical problem. It was attacked succesfully by Kebe (1981) in Hamburg. A useful signal to noise ratio at seismic frequencies was obtained by Cotaras et al. (1988). It still remains to devise a durable and lightweight system that can float freely in the oceans and that can, with enough electronic intelligence, recognize a seismic signal. Data management is no problem in this case - we may use satellite communication, a technique pioneered by Poupinet (1987). The location of the buoy can be done with sufficient precision by the satellite as well.

If such a buoy could be developed, a program to operate an oceanic network with 1000 buoys for 6 months would cost about 4 million ECU, and only a fraction of that if it can be set up as a piggyback experiment within an oceanographic program. The big attraction of such an array would be that the network would have a dynamic configuration: even if an earthquake is repeated at the same location, the ray coverage provided by the data would be different each time.

The biggest problem is gaining information in oceanic regions. This is not to say that we cannot improve our observations on land. On land, many regions have only a few seismic stations reporting to the ISC. We cannot expect governments to install costly seismic stations for purely scientific reasons. Rather, we could take example by the astronomers, who have nurtured amateur astronomy for ages - until this day comets and variations in stellar brightness are often detected by dedicated astronomy amateurs.

The prohibitively large cost of a seismograph has made it more difficult for amateur seismologists to contribute to the field. But this is quickly changing - a short period seismograph with a simple AD converter that connects to a PC should, together with an accurate time code reciever, not cost more than 1000 ECU, about the same price as a medium size telescope such as used by an amateur astronomer. Many high schools in western countries are able to afford such a system. With a little bit of effort, the seismometer could perhaps be made broad band. Of course, this philosophy to sacrifice quality for quantity goes against the grain with many observational seismologists. And it is true, a large number of very accurate instruments will always be needed, especially for source studies. But we should realize that we are facing lateral heterogeneity on a very small scale. We need to fill in the often very large separation of stations, and the only thing that has to be very accurate to measure a delay time is the clock.

Until the station coverage improves, we have to do with the data on hand. There are several ways to get more out of a seismogram than just the delay time of the first arrival. The most logical step is to use later arrivals such as pP and PP. It has been suggested to initiate an International Seismological Observing Period, or ISOP, to improve the precision and the quantity of the readings of later phases on the seismogram. As I stated before, we gain much more by increasing the



precision of delay time readings than by waiting until we have enough overlapping raypaths to reduce the standard error in the average or summary ray, and ISOP can be expected to increase the reliability of tomographic images substantially.

**Theoretical limitations**

This raises the question how far we can possibly go: what is the smallest structure that we can still resolve at depth using delay time tomography if we would live in a seismologist's heaven with no limitaions to the abundance of raypaths crossing the Earth?

The question is hypothetical, since for some time to come the scarcity and the large variance in the data will determine the limits of resolution. But there are some important theoretical limitations that are worth investigating. The answer to the question of maximum resolution is a complicated one, we cannot give a simple number. What we can resolve depends on the sign and magnitude of the velocity anomaly and on the depth. However, for the purpose of this discussion, we shall adopt the Fresnel zone as a reasonable measure of attainable resolution.

Wielandt (1987) has shown that there is an important asymmetry in delay times: fast regions will generally give negative delay times, but slow regions will only give positive delay times if the heterogeneity is large with respect to the ray length, or located close to reciever or station. The reason is that rays seek a minimum time path. If the ray needs only to bend a little bit to diffract around a low velocity body, we probably read the diffracted wave as the first arrival. But when interpreting ISC delay times, we do interprete this wave as the transmitted wave! Thus, there is a disturbing bias in seismic tomography: we are likely to see high velocity regions, but low velocity bodies remain largely invisible.

In order to investigate such effects in some detail for more realistic situations, we have applied a new form of ray tracing, based on graph theory and developed by Moser (1991), to compute the wavefronts of the first arriving waves, irrespective of the type. This enables us to compute the Fresnel zone for teleseismic waves in realistic Earth models (Figure 3,4). Generally, the Fresnel zone is widest at the turning point of the ray. For an epicentral distance of 60°, the Fresnel zone is as wide as 400 km. A dramatic narrowing down occurs near the CMB, where the wave can make no detours in the low velocity core without being punished by long delays.

It is often claimed that the D" layer, just above the CMB, is more heterogeneous than the lower mantle in general. The study by Gudmundsson et al. (1990) quantified this effect. They determined the variance of ISC delay times as a function of depth. The correlation between the variance of their model, and the width of the Fresnel zone of teleseismic rays (Figure 5) does however suggest that their results must be viewed with caution. The lower mantle may very well seem to possess only small velocity variations because the seismic waves lack the resolution to resolve this. Further research is needed to substantiate this. Such research should be aimed at establishing the frequency dependence of delay times. Obviously, such analysis cannot be done with ISC reported times. We may have to wait until a large set of broad band data has been assembled.

It is a major task of the seismological community to arrange for the management, storage and distribution of such data. Worldwide, several initiatives are under way to do just that: IRIS in the US, Orfeus in Europe, and Poseidon in Japan. I strongly urge that the data management centers of these consortia are funded at an adequate level. In Europe, where there is no multinational counterpart of the USGS, this not only important for purely scientific endeavours such as discussed in this Symposium lecture - it is also of great significance for the public applications of seismology. Even though digital broad band records give us much more information than has ever been available from analogue recording, the rapid accessibility as well as the long term storage of digital records are no trivial problems.

## References


Aki, K., A. Christoffersson and E.S. Husebeye, Determination of 3D seismic structure of the lithosphere, J. geophys. Res., **82**, 277-296, 1977.

Clayton, R.W. and R.P. Comer, A tomographic analysis of mantle heterogeneities from body wave travel time (abstract), EOS, **64**, 776, 1983.

Cotaras, F.D., I.A. Fraser and H.M. Merklinger, Near-surface ocean ambient noise measurements at very low frequencies, J. Acoust. Soc. Am., **83**, 1345-1359, 1988.

Dines, K. and J. Lytle, Computerized geophysical tomography, Proc. Inst. Elec. Eelectron. Eng., **67**, 1065-1073, 1979.

Dziewonski, A.M., B.H. Hager and R.J. O'Connell, Large-scale heterogeneities in the lower mantle, J. Geophys. Res., **82**, 239-255, 1977.

Dziewonski, A.M., Mapping the lower mantle, J. Geophys. Res., **89**, 5929-5952, 1984.

Gudmundsson, O., J.H. Davies and R.W. Clayton, Stochastic analysis of global travel time data: mantle heterogeneity and random errors in the ISC data, Geophys. J. Int., **102**, 25-44, 1990.

Kebe, H.-W., Self-noise measurements using a moored sonobuoy with a suspended hydrophone, Marine Geophys. Res., **5**, 207-220, 1981.

Moser, T.J., G. Nolet and R. Snieder, Ray bending revisited, subm. to Bull. Seism. Soc. Am., 1991.

Nolet, G., Inversion and resolution of linear tomographic systems (abstract), EOS, **64**, 775-776, 1983.

Nolet, G., Solving or resolving inadequate and noisy tomographic systems, J. Comp. Phys., **61**, 463-482, 1985.

Nolet, G., Seismic wave propagation and seismic tomography, in: *Seismic tomography*, ed. G. Nolet, 1-24, Reidel Publ. Comp., 1987.

Oldham, R.D., The constitution of the interior of the Earth as revealed by earthquakes, Q. J. Geol. Soc., **62**, 456-475, 1906.

Paige, C.C., Saunders, M.A., LSQR: An algorithm for sparse linear equations and sparse least squares. ACM Trans. Math. Softw. **8**, 43-71, 1982.

Poupinet, G., Seismic data collection platforms for satellite transmission, in: *Seismic tomography*, ed. G. Nolet, 239-250, Reidel Publ. Comp., 1987.

Spakman, W., and G. Nolet, Imaging algorithms, accuracy and resolution in delay-time tomography, in: *Mathematical Geophysics*, ed. N.J. Vlaar et al., 155-187, Reidel Publ. Comp., 1988.

Spakman, W., S. Stein, R. van der Hilst and R. Wortel, Resolution experiments for NW Pacific subduction zone tomography, Geophys. Res. Lett., **16**, 1097-1101, 1989.

Tarantola, A., *Inverse Problem Theory*, Elsevier, 1987.

Van der Sluis, A. and H.A. Van der Vorst, Numerical solution of large, sparse linear algebraic systems arising from tomographic problems, in: *Seismic tomography*, ed. G. Nolet, 49-84, Reidel Publ. Comp., 1987.

VanDecar, J.C. and R.S. Crosson, Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares, Bull. Seism. Soc. Am., **80**, 150-169, 1990.

Wielandt, E., On the validity of the ray approximation for interpreting delay times, in: *Seismic tomography*, ed. G. Nolet, 85-98, Reidel Publ. Comp., 1987.




**Table 1: P-wave pressures at 100m depth**

| $m_b$ | Seabottom displacement (m) | Water Pressure (Pa) |
|---|---|---|
| 4 | $5\times10^{-8}$ | 0.4 |
| 5 | $2\times10^{-7}$ | 1.5 |
| 6 | $1\times10^{-6}$ | 7.2 |
| 7 | $5\times10^{-6}$ | 36 |
| 8 | $2\times10^{-5}$ | 152 |

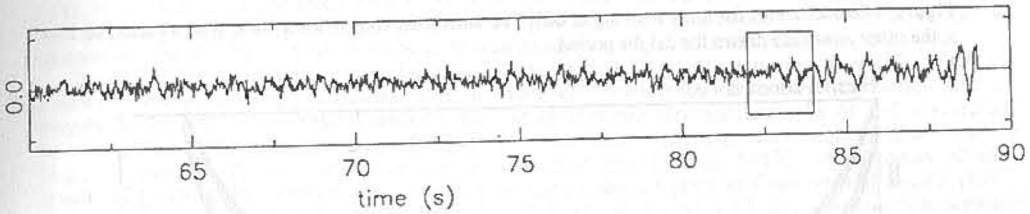


**Figure 1** The record of a floating vertical seismograph shows the arrival of a P wave near 82.5 s. The arrival near 87 s is a pP wave.

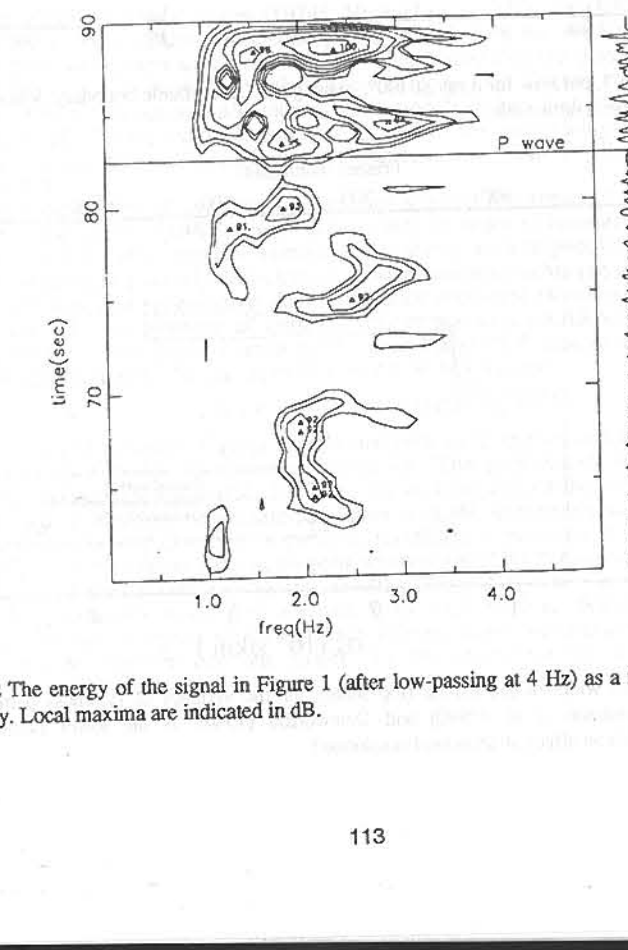


**Figure 2** The energy of the signal in Figure 1 (after low-passing at 4 Hz) as a function of time and frequency. Local maxima are indicated in dB.

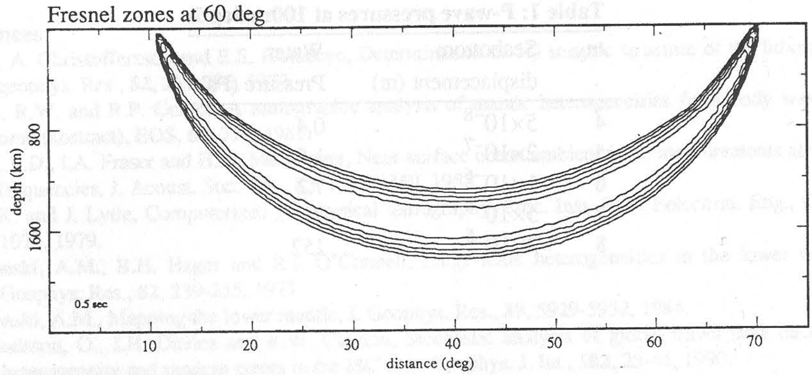


**Figure 3** Fresnel zones for a ray arriving at 60°. The narrowest zone is for a wave with a period of 1 s, the other zones are drawn for 2,3,4 s period.

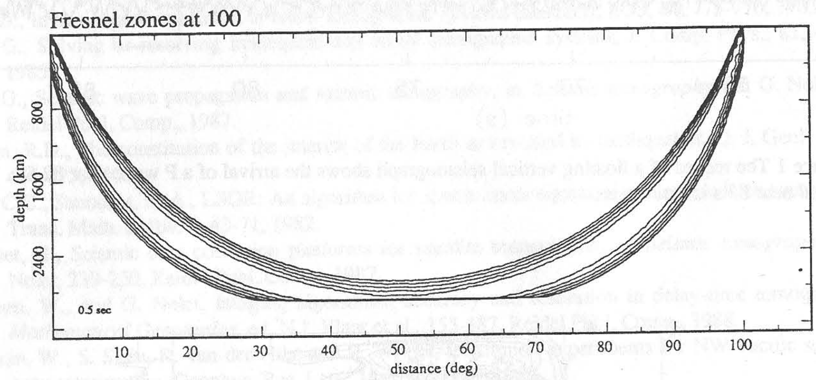


**Figure 4** As Figure 3, but now for a ray at 100°, grazing the Core-Mantle boundary, where a sudden increase in resolution is noticeable.

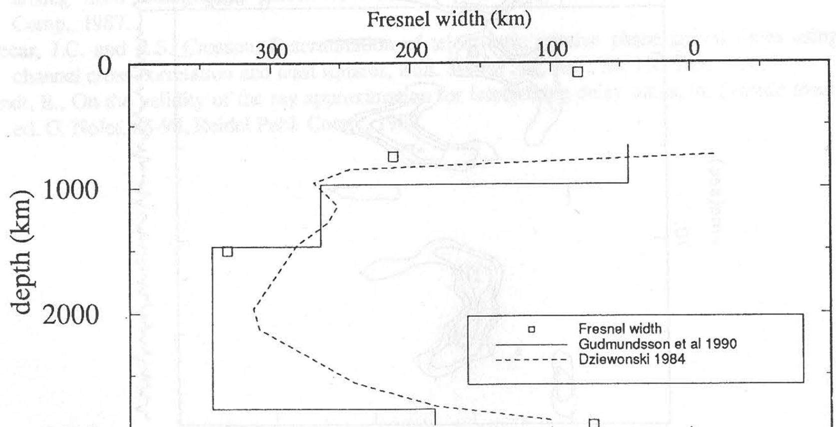


**Figure 5** The Fresnel width in the mantle (top scale), vs. the variance of slowness perturbations inferred by Gudmundsson et al. (1990) and Dziewonski (1984). Is the lower mantle really homogeneous or is this an effect of decreased resolution?