Confidence Regions for Mantle Heterogeneity

By

R. Jay Pulliam*

Philip B. Stark
Department of Statistics
University of California
Berkeley, CA 94720

Technical Report No. 392
June 1993

*Current Address: Dept. of Theoretical Geophysics, Utrecht University, The Netherlands

Department of Statistics
University of California
Berkeley, California 94720

Confidence Regions for Mantle Heterogeneity

R. Jay Pulliam*

Philip B. Stark

Department of Statistics

University of California

Berkeley CA 94720

June 8, 1993

Abstract

Least-squares tomographic models of mantle P and S structure from travel times show large-scale variations correlated with surface tectonic features, as well as coherent structures in the lowermost mantle. The reliability of these global features of velocity models depends on whether the velocity throughout the feature can be estimated well simultaneously: we need to be able to say with confidence that a feature involving many voxels is likely to be real. We find a lower bound on how wide a 95% simultaneous confidence region for mantle P or S velocity must be, as a function of position in the mantle. Suppose (perhaps optimistically) that summary-ray travel-time errors are

*Current Address: Dept. of Theoretical Geophysics, Utrecht University, The Netherlands

independent with mean zero and standard deviation 0.25 s for direct P phases, and 0.5 s for all other phases, regardless of distance. If summary travel-time residuals are fitted by least-squares to a mantle model parametrized by 4872 10° by 10° voxels, the half-width of a 95% confidence envelope for P velocity (using 889,909 summary P rays, 84,046 summary PP rays, 67,228 summary pP rays, 23,024 summary PcP rays, 11,239 summary $PKPab$ rays, 47,227 summary $PKPbc$ rays and 141,843 summary $PKPdf$ rays) ranges from ± 0.025km s^{-1} to ± 99.8km s^{-1}. It is wider than 0.193km s^{-1} in half of the mantle, and wider than 0.593km s^{-1} in a quarter of the mantle, by volume. Under the same assumptions and with the same parametrization, a 95% confidence envelope for S velocity based on 163,354 summary S rays, 13,781 summary SS rays, 7,441 summary ScS rays and 12,494 summary sS rays ranges from ± 0.031km s^{-1} to $\pm\infty$km s^{-1}, and is wider than 0.254km s^{-1} in half of the mantle, and wider than 0.554km s^{-1} in a quarter of the mantle. With this ray set, parametrization, and error model, the least-squares estimate of P velocity is within a 95% confidence envelope around the radially symmetric *iasp91* model for 86.6% of the mantle's volume. The least-squares S estimate is within a 95% confidence envelope around *iasp91* for 88.9% of the mantle's volume. Thus, on a global scale, the mantle's velocity structure is nearly consistent with the iasp91 radially symmetric model. Smaller voxels, more realistic assumptions about the errors, and three-dimensional structure outside the mantle would make the confidence intervals still wider.

1 Introduction

For about fifteen years, seismologists have fitted three-dimensional seismic velocity models of the mantle to travel-time data. (See, e.g. *Sengupta and Toksöz* [1976], *Dziewonski et al.* [1977], *Clayton and Comer* [1983], *Dziewonski* [1984], *Inoue et al.* [1990], *Pulliam* [1991], *Woodward and Masters* [1991ab], *Pulliam et al.* [1993], and *Vasco et al.* [1993].) Constructing these models requires quite large computations. Appraising the uncertainty of the models requires still larger computations which were not possible until quite recently (the results below used a Cray Y-MP C90 with 2 Gb of RAM; *Vasco et al.* [1993] use a Connection Machine 2 to compute a model covariance matrix for damped least-squares mantle tomography). Some of the three-dimensional velocity perturbations in these models correlate with surface tectonic features such as ridges, continents and subduction zones. The evidence for the three-dimensional velocity perturbations is that the features "make sense" tectonically, and that the models fit the travel-time data better than the one-dimensional models from which they derive.

However, since the models introduce hundreds or thousands of new parameters (sometimes in the form of voxels, sometimes as truncated spherical harmonic expansions in the angular variables tensored with polynomial or piecewise polynomial functions of radius), we should expect improved fit to the data, even if the velocities are not constrained very well. If we allowed arbitrary velocity perturbations, within the linearization of the problem it would be possible to fit the data exactly, provided each raypath samples some points not sampled by any other; however, the inferred velocities would be totally unreliable. Furthermore, the improvement in fit achieved by these models is not spectacular: typically less than 30%, depending on how (and whether) summary rays are formed and on the extent of damping.

Here we show that global features of least-squares tomographic mantle models are not statistically reliable. Uncertainties in some individual voxels might be small (we obtain only lower bounds), but if one wants to infer simultaneously the velocity in many voxels, the uncertainty grows quickly, so that global properties that might constrain mantle convection or composition can not be determined using current approaches. (We argue that other approaches might still yield useful inferences.)

We lower-bound the uncertainty by using a technique introduced by *Stark and Hengartner* [1993] to show that the uncertainty is large in a simplified version of the problem, constructed in such a way that the true uncertainty is larger. In the idealization,

- Summary-ray travel times have statistically independent errors with zero mean and standard deviation 0.25 s for direct P phases, and 0.5 s for S, core, and reflected phases, regardless of distance.
- ISC source locations are accurate.
- Ray paths computed from the *iasp91* 1-dimensional velocity model [*Kennett and Engdahl*, 1991] are accurate.
- Linearization errors are negligible.
- The "true" mantle P and S velocities are constant in a set of 4872 voxels, roughly 10° by 10° in angular extent, with depth boundaries given in Table 2.
- Seismic velocity in the rest of the Earth is radially symmetric and follows the *iasp91* model.

These assumptions are all optimistic: the apparent uncertainty in the model problem is smaller than the true uncertainty in the real problem (see the Appendix as well).

The distinction between how well the velocity in a given voxel can be determined and how well the velocities in all the sampled voxels can be determined simultaneously is important. If we are estimating many parameters (*e.g.*, velocities in different parts of the mantle) we may construct a 95% confidence interval for each parameter, with a length that depends on the variance of the estimate at that point (the variance depends on the ray sampling and the model parametrization, in addition to assumptions about the noise). If all the estimates were perfectly correlated, whenever one parameter fell within its confidence interval, all the others would too: the "simultaneous coverage probability" of the set of confidence intervals would equal the individual coverage probabilities of each interval—95%. However, since the estimates at different points are not perfectly correlated, even though each interval "covers" the true parameter value 95% of the time, the set of intervals will simultaneously include all the true velocities less than 95% of the time. Sometimes the confidence interval for the velocity at location r_j will contain the true velocity, while the confidence interval at location $r_k \neq r_j$ will not (Figure 1). To get 95% simultaneous coverage probability we must increase the lengths of the individual confidence intervals by an amount that depends on the number of parameters we estimate—see *Stark and Hengartner* [1992] for a more complete discussion. To make inferences about large-scale features of mantle models from estimates of the seismic velocity as a function of position, we need to simultaneously constrain the velocity at many positions. This paper constructs lower bounds on the width of a simultaneous confidence region for seismic velocity in the mantle, rather than confidence intervals for individual voxels.

2 Data

We relocated more than 42,000 events from the ISC catalog for the years 1964 to 1987 using both P and S arrivals and the one-dimensional *iasp91* Earth model. We kept all events with at least 25 P arrivals, with no restriction on the number of S arrivals or depth (0 - 700 km). We identified direct and reflected phases using $p-\tau$ interpolation routines provided by Ray Buland with the *iasp91* model [*Buland and Chapman*, 1983; *Kennett and Engdahl*, 1991], first picking P and then identifying *pP, PP, PcP, PKPab, PKPbc, PKPdf, S, sS, SS,* and *ScS,* arrivals simultaneously. We omitted regions of the T - Δ plane where the *iasp91* model predicts more than one of these phases. These regions depend on the depth of the event.

Generally, we kept P phases between 0° and 100° epicentral distance from the source, with travel times within 10 s of the predicted time. We kept other phases if their residuals were less than 15 s. *PP* identifications were allowed between 28° and 40° and between 44° and 180°. *PcP* had to be between 28° and 40°, or between 44° and 75°. *pP* identifications were allowed from 0° to 140°. A *PKPbc* arrival must have been identified between 146° and 154° before we allowed either *PKPab* or *PKPdf*. *PKPab* identifications had to be between 146° and 180°, and *PKPdf* could be between 114° and 180°.

We permitted S identifications between 0° and 20°, 25° and 35°, 45° and 75°, and between 85° and 105°. *SS* had to be between 51° and 180°. *ScS* was allowed between 0° and 35° and between 45° and 70°. *sS* arrivals had to be between 0° and 51°, and an S arrival must have been identified first. If an *sS* observation was outside the $15s$ window allowed for S, but within the $15s$ window allowed for *sS* and no S arrival has been identified, we discarded the *sS* observation.

These criteria are intended to overestimate the number of good data one might use to estimate mantle structure. If we incorrectly identified some phase other than *P, PP, pP, PcP, PKP, S, SS, sS, ScS* as one of those, our results are still optimistic—we have just overestimated the amount of data. If we misidentified one of those phases as another, it is hard to tell whether the results are optimistic or not, so we were especially conservative identifying phases where travel time branches cross. In regions of the T - Δ plane where only one phase we consider can have an arrival, we were more generous. For example, we kept observations of *SS* between 51° to 180°, because over that range *SS* cannot be confused with any other phase we use.

We formed summary rays by combining rays that sample nearly the same parts of the mantle, which we determine by proximity of the rays' endpoints. We binned the ray endpoints in 2° by 2° by 25 km thick voxels from Earth's surface to a depth of 700 km. Rays of the same type emanating from and ending in the same two bins were combined into a single summary ray. We did not require a minimum number of actual rays in each summary ray, so many summary rays contain just one ray. Table 1 lists the number of identified phases of each type and the number of summary rays formed from them.

3 Theory

We assume the data d (summary travel-time perturbations from the predictions of the *iasp91* spherically symmetric model) arise linearly from the model $x(r)$ with additive noise e whose components e_j are independent with zero mean and variances σ_j^2. We assume x is piecewise constant in the voxels $\{v_k\}_{k=1}^{n}$:

$$x = \sum_{j=1}^{n} \beta_j x_j(r), \tag{1}$$

where

$$x_j(r) = \begin{cases} 1, & r \in v_j \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Let the matrix A map the vector of velocity perturbations in the voxels into the linearized, ray-theoretical predictions of the travel-time perturbations they produce, so that

$$d = A\beta + e. \tag{3}$$

We use the approach of *Stark and Hengartner* [1992] to find the uncertainty in the simplified problem. This requires us to compute and invert the square "Gram matrix" Γ whose entries are

$$\Gamma_{kl} = \sum_{\text{all summary rays } j} \frac{\Delta T_j(k) \cdot \Delta T_j(l)}{\sigma_j^2}, \tag{4}$$

where $\Delta T_j(k)$ is the linearized change to the travel time along the jth ray due to a unit perturbation of the velocity in voxel k. It is easy to see that

$$\Gamma = A^T \Lambda^{-1} A, \tag{5}$$

where $\Lambda = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_N^2)$ and N is the number of summary rays. This computation is quite large since there are so many summary rays, each of which samples many voxels, but the resulting matrix has just n^2 elements, where n is the number of sampled voxels in the model. By symmetry, only $n(n+1)/2$ of these are distinct.

The reproducing kernel $K(r, s)$ is related to Γ by

$$K(r, s) = \sum_{kl} (\Gamma^{-1})_{kl} x_k(r) x_l(s), \tag{6}$$

where $(\Gamma^{-1})_{kl}$ is the (k,l)th element of the matrix inverse of Γ. The half-width of a 95% confidence interval for the velocity perturbation at the point r in the mantle is

$$u_{.95}(r) = \sqrt{\chi^2_{.05,n}}\sqrt{K(r,r)}, \tag{7}$$

where n is the number of voxels in the inversion, i.e., the number of sampled voxels in the model, and $\chi^2_{.05,n}$ is the 95th percentile of the χ^2 distribution with n degrees of freedom. For large n (over 4000 here), the chi-square distribution is well approximated by the normal, and

$$\chi^2_{.05,n} \approx n\left(1 - \frac{2}{9n} + 1.645\sqrt{\frac{2}{9n}}\right)^3. \tag{8}$$

In linearized problems like this one, where the unknown is assumed to lie in a finite-dimensional subspace, there is no other a priori constraint on the model, and the parameters are coefficients of a voxel model, the relation between the reproducing kernel and the model covariance matrix is straightforward. For the moment assume for simplicity that the standard deviations of the errors e_j are all unity (i.e. Λ =identity).

The model covariance matrix for the ordinary least-squares estimate $\hat{\beta}_k$ is then

$$\Sigma_{kl} = E\left((\hat{\beta}_k - \beta_k)(\hat{\beta}_l - \beta_l)\right) = \left((A^T A)^{-1}\right)_{kl}. \tag{9}$$

The variance of the model at the point r can be found from the covariance of the coefficients β by multiplying the coefficients by the voxel basis functions x_k and x_l and by Σ_{kl}, then summing on k and l and evaluating the sum at the point r:

$$\sigma^2(r) = \sum_{k,l} \Sigma_{kl} x_k(r) x_l(r). \tag{10}$$

Since $\Gamma = A^T A$, this is in fact

$$\sigma^2(r) = \sum_{k,l} (\Gamma^{-1})_{kl} x_k(r) x_l(r) = K(r,r), \tag{11}$$

and thus that the standard deviation of the model at the point r is

$$\sigma(r) = \sqrt{K(r,r)}. \tag{12}$$

See Figure 1 for a schematic illustration. The normalizing factor $\sqrt{\chi^2_{.05,n}}$ buys simultaneous coverage probability; if we were interested in the value of the model only at the point r and no other points, we could normalize instead by 1.96, the 97.5th percentile of the standard normal distribution.

4 Computations

The computational limitation is not time but storage. To compute Γ we use sparse storage techniques to minimize the required storage and allow us to compute Γ efficiently. To invert Γ we use LAPACK [*Anderson et al.* 1992] routines SSPTRF and SSPTRI, which store only the upper triangle of the matrix. The LAPACK routines call BLAS subroutines that have been optimized and parallelized by Cray Research, Inc.

We compute Γ by forming $A^T\Lambda^{-1/2}\Lambda^{-1/2}A$. The matrix A that maps model parameters into travel-time perturbations is stored as three entirely dense vectors: a real vector containing the nonzero elements of A, and two integer vectors that keep track of the locations of the nonzero elements in A. The first of the integer vectors indicates to which column of A each element of the real vector belongs, and the second integer vector keeps track of where new rows start in the other two vectors. The number of elements in the real vector and the first integer vector is the total number of voxels sampled by all the rays in the data set, counting duplications (29 million for compressional phases and 4.5 million for shear phases). The dimension of the second integer vector is just the number of data. To make the multiplications more efficient, we sort the indices so that elements corresponding to the same voxel are adjacent in the vector. (Actually, we create a third integer vector whose elements are sorted pointers to elements of the second vector, which avoids actually moving any of the stored numbers.) Since we know where new rows begin, the algorithm can just loop through the sorted vector and do the necessary products and sums, referencing only elements that need to be multiplied or added. The whole procedure vectorizes and is highly efficient.

We performed the major calculations on the Cray Y-MP C90 at Cray Research, Inc. in Chippewa Falls, Wisconsin. This is Cray Research's fastest machine. It has 256 Mwords (about 2 Gb) of RAM, and 16 cpus. Each cpu can perform a billion floating-point operations per second (1 Gflop), and can do two floating-point adds and two floating-point multiplies in each 4.1 nanosecond clock cycle. Our jobs typically ran in parallel on 8 cpus. The maximum memory required was 120 Mwords (about 960 Mb) to form Γ, and 24 Mwords (192 Mb) to invert it.

For the largest problem (using all the compressional phases) forming Γ took about 39 s, and inverting it took about 188 s. For the smallest problem (S alone), forming Γ took 9 s and inverting it took 157 s. Input and output took on the order of a minute.

5 Results

The results are summarized in Tables 3 through 6, which give percentage points of the spatial distribution of uncertainty. They tabulate the smallest values such that the fraction of the mantle's volume for which the half-width of a 95% confidence interval is less than those values is 0, 25%, 50%, 75% and 100%. Tables 3 and 4 juxtapose, for a fixed volume fraction of the mantle, the uncertainties with more and more phases, to show how much each phase helps to constrain mantle structure. Tables 5 and 6 give the distribution of the uncertainties using all the compressional phases and all the shear phases, respectively, layer by layer within the mantle.

The tables show that once the entire mantle is sampled (e.g., once both P and PP phases are used), including additional phases does not greatly effect the minimum, 25th percentile, median, or 75th percentile of the uncertainty. The maximum uncertainty decreases with the addition of new phases, but remains large for P (± 99.8km s^{-1}), and infinite for S, even though the voxels in the parametrization are so large.

It is interesting that some of the S velocity uncertainties are smaller than the corresponding P velocity uncertainties, although there are fewer shear phases and their local sampling density is smaller than for P. This is due to the way we parametrize the problem, which excludes unsampled voxels from the least-squares estimation (their uncertainties are infinite). The value of n for the chi-square percentage point is just the number of sampled voxels. Fewer voxels are sampled by shear phases than by compressional phases. Since fewer parameters are estimated for S than for P, the value of $\chi^2_{.05,n}$ is smaller, even though the reproducing kernel (analogous to the model covariance) is larger.

Our analysis up to this point has not used the *ISC* travel time residuals themselves: the model uncertainties depend only on the geographical distribution of ray paths and our idealization of travel time errors. However, in order to find velocity estimates we must use the travel time residuals, and there may be problems with this particular data set: *ISC* residuals may have a low signal-to-noise ratio, as suggested by *Gudmundsson et al.* [1990] and *Davies et al.* [1992], or systematic biases from any of a number of factors (e.g., subduction zone events may be mislocated, there may be anomalous structure in the outer or inner core, etc.).

In either case our estimate of the width of a confidence region remains optimistic, but our velocity estimate depends on *ad hoc* assumptions about the quality of *ISC* data. For example, the inferred velocity anomalies depend on how we treat very large travel time residuals. If we decide that residuals greater than, say, ±10s are unreliable, we will probably choose to downweight them somehow, which will tend to reduce the magnitude of the inferred velocity perturbation. We have implicitly completely downweighted extreme residuals by using time windows to identify phases. Choosing a level at which to downweight is rather subjective, as are the details of the procedure.

For example, we might set all residuals beyond ±10 s to ±10 s (truncation), or we might simply discard the larger residuals. Alternatively, or in addition, we might try to reduce spurious large deviations in the fitted model due to erroneous data by using damped least-squares.

To get a sense of the effects of these choices, we fitted models by least-squares to the S data with (a) residuals larger in absolute value than 7 s discarded, (b) residuals larger than 10 s discarded, and (c) residuals truncated at ±7 s. We also estimated a model using damped least-squares (d) by adding a small constant (0.01, or approximately 0.1%) to the diagonal of the Gram matrix.

When we truncated a residual, we kept the ray in computing the confidence region, and kept the truncated residual (of ±10 s or ±7 s) in estimating the velocity model. When we discarded a residual, we discarded the ray as well, and did not include it in computing the confidence region. Thus the number of data is smaller when we discard than when we truncate. Consequently, the confidence region tends to be narrower when we truncate, while the estimated model tends to be smaller when we discard, since there are fewer large residuals. When we used damped least-squares, we incorporated the damping in the reproducing kernel as well, which made the confidence regions narrower. (Note that incorporating the damping into the reproducing kernel implies an *ad hoc* assumption about the norm of the velocity model; this assumption is not justified on physical grounds. *Stark and Hengartner* [1993] discuss the relation between norm bounds, smoothness assumptions and the reproducing kernel approach in more detail.)

The voxels least significantly different from the *iasp91* reference model were, of course, the unsampled voxels: their uncertainty is infinite. Tables 5 and 6 show the approximate distributions of the locations of the unconstrained voxels by depth. The results are essentially the same for the discarding/truncating/damping treatments (a)-(d). Using just P or S phases, the coverage is worst in the upper mantle, so perturbations in the upper mantle tend to be worse-constrained than in the lower mantle. Overall, the voxels most significantly different from the reference model were in the lower mantle, slightly deeper for the P data alone than for the other data sets. Table 7 gives the locations and "signal-to-noise" ratios for the best constrained voxels using the four extreme data sets. The relatively larger values of the best signal-to-noise ratio for shear versus compressional phase data results from the unrealistically small uncertainties we assigned to shear phases, and from the fact that shear phases sample fewer voxels than compressional phases do, so we estimate the velocity in fewer voxels and thus can use a smaller value of χ^2. The best-constrained regions are rather scattered throughout the mantle. We do not plot them here, since, as a result of our research (and taking into account the large number of optimistic assumptions we have made here), we are skeptical of the reliability of least-squares ray-theoretical travel-time tomography. We emphasize that the uncertainties we find are optimistic lower bounds on the true uncertainty.

Table 8 compares weighted least-squares estimates for treatments (a)-(d) with the corresponding 95% confidence regions. We solved the normal equations directly. This is likely to give larger estimated anomalies than iterative procedures such as LSQR, since stopping after only a few iterations (typically 30 to 50) is equivalent to regularizing the least-squares problem in an unusual way, which tends to reduce the estimated velocity perturbations.

For case (a), discarding residuals beyond ±10 s, the model fit using S phase data is within a 95% confidence envelope around the spherically-symmetric *iasp91* model in about 90% of the mantle's volume: the differences from *iasp91* in 90% of the mantle are statistically insignificant. Incorporating all the shear phases considered here (*S, SS, ScS, sS*) increases the "significant" volume to 11.1%. Discarding residuals beyond ±7 s (case (b)) decreases the constrained volume, more for the model fitted to all the shear phases. In case (b), the size of the anomalies tends to decrease since there are fewer large residuals, while the width of the confidence region increases since there are fewer rays. As expected, this results in a smaller volume in which the velocity is estimated to be significantly different from *iasp91*.

Truncating the residuals at ±7 s (case (c)) decreases the "significant" volume compared with the previous case using S alone, but increases the "significant" volume when all shear phases are used. Damping (case (d)) reduces the width of the confidence region, but also decreases estimated velocities. In net effect, damping does not increase the volume of statistically significant departures from the *iasp91* model.

To compute all the different cases for compressional phases was prohibitively expensive, so we did only case (a). Using P alone, discarding residuals larger than ±10 s, the estimated velocity in about 10% of the mantle's volume differs significantly from *iasp91* at the 95%

confidence level. Using all the compressional phases, discarding residuals beyond ± 10 s), 13.4% differs significantly from *iasp91*.

6 Discussion

6.1 The Bad News

The uncertainties we find are all extremely optimistic: they depend on assumptions that are surely false, and whose violation makes the true uncertainty larger. The artificial restriction with the largest effect on the uncertainty is the relatively crude voxel model. If we allowed seismic velocity in the mantle to be truly nonparametric, within ray theory the uncertainty in velocity would be formally infinite except along ray paths, which have zero volume. The "smoothness constraint" implicit in a voxel basis is the only restriction that makes the uncertainty finite over positive volumes. Refining the parametrization by using smaller voxels increases the uncertainty two ways: the number of estimated parameters tends to grow, so there is a larger normalization constant for the uncertainties, and the parametrization gives more freedom to trade off anomalies in different places and still fit the data adequately. Furthermore, a smaller volume of the mantle would be sampled in a finer discretization, so more of the mantle would have infinite uncertainty.

The next two assumptions likely to reduce the apparent uncertainty profoundly are the restriction to radially symmetric structure except in the mantle, and the assumption that travel-time errors are independent with such small standard deviations, regardless of distance. Allowing three-dimensional structure in the core and near stations (e.g. station corrections) would allow a larger range of mantle models to fit the data adequately. Similarly, the notion of adequate fit to the data depends on the noise level, which tends to grow with distance, and is probably closer to 1 s than to 0.25 s or 0.5 s for many of the phases in the data set we used. Increasing the standard deviations increases the uncertainty in direct proportion. It is also unlikely that the data errors are independent. Some of the "error" is unmodeled velocity structure, which may be correlated through station effects and on the spatial scale over which summary rays are formed. It is hard to predict the effect incorporating the true error dependence would have on the uncertainty (however, see the Appendix).

The next approximation we think is likely to reduce the apparent uncertainty is using ray theory to model the sensitivity of travel times to changes in velocity. *Stark and Nikolayev* [1993] show that travel times are much more sensitive to near-source and near-receiver structure than to structure along the raypath, due to near-field terms that ray theory omits. As a result, near-source and near-receiver structure are likely to show up all along the ray path when ray theory is used to model the physics. Since earthquakes and stations are both in relatively inhomogeneous regions, this effect could be large.

The approximations we think have the smallest effect on the uncertainty are using raypaths and source locations derived from the *iasp91* model, and linearizing the problem around the *iasp91* model. The likely deviations from these locations and raypaths are not large enough, relative to the size of voxels, to have a significant effect. In order to change the uncertainty much, the effects of linearization and of using the *iasp91* model would have to be big enough to cause the inferred raypaths to sample substantially different voxels.

Within ray theory, the spatial variation of the uncertainty is dominated by the sampling, by the noise level in the summary rays, and by the parametrization. The actual travel times do not enter into the calculation of the uncertainties, so the quality of ISC picks, and possible biases in the ISC data set are not at issue. We do not think that other, higher quality sets of travel time data could improve much on the uncertainties we find here. For example, using differential travel times eliminates some uncertainties in origin times and event locations [*Woodward and Masters*, 1991ab], but as long as the distribution of raypaths is similar to that we used, the uncertainty will be about the same, since we neglect errors in origin time, locations, near-source structure, and station effects in our analysis (see the Appendix).

Using regularized least-squares reduces the model covariance, which yields formally smaller uncertainties, at the expense of introducing additional bias of unknown magnitude. Using only a small number of iterations in LSQR (a conjugate-gradient approach; see *Paige and Saunders* [1982]) , or *van der Sluis and van der Vorst* [1987]), SIRT (a linear iterative technique; see *van der Sluis and van der Vorst* [1987]) has a similar effect. We are unaware of any quantitative argument (e.g., from gravimetry, geomagnetism, or mineral physics) suggesting a bound on the spatial variation of seismic velocity in the mantle. Such an argument is needed to bound the bias regularization introduces. Regularization enhances the numerical stability of algorithms that construct tomographic images and reduces the variance of the estimation procedure, but it is not clear whether regularization increases or decreases the reliability of the images, since the additional bias it introduces could be arbitrarily large.

6.2 Promising Directions

It seems plausible that using surface wave data and normal mode data in addition to travel-time data could improve the reliability of images of long wavelength features of the mantle, since, within the linearization of the problem, those data provide linearly independent measurements sensitive to large-scale departures from spherical symmetry. However, the possibility of small-scale structure substantially affects the uncertainty of simultaneous estimates when the data measure long wavelength averages. Thus the magnitude of the improvement might not be large. The reproducing kernel approach advocated here and in *Stark and Hengartner* [1993] can be used to study tomographic imaging using combinations of travel time, surface wave, and normal mode data as well.

The direction we find most promising is to use a more accurate representation of the physics than ray theory (*e.g.*, "tubular tomography" as advocated by *Stark and Nikolayev* [1993]), and to focus on particular, geophysically meaningful and interesting parameters and hypotheses instead of trying to image the entire mantle. For example, it should be possible to discern differences in the average velocity in moderate-volume regions below 670 km beneath subduction zones, if they exist. Inferences about such functionals would yield strong evidence for or against slab penetration and whole-mantle convection.

Similarly, if the degree 2 pattern of heterogeneity is of primary importance in testing convection models (see, *e.g.*, *Silver et al.* [1988] for a discussion), the 5 $l = 2$ coefficients could be estimated directly using techniques advocated by *Stark* [1992a] in a way that takes into account the uncertainty from nonuniqueness in infinite-dimensional problems. The approach of fitting a small number of spherical harmonics or voxels can be patched up to give conservative uncertainties using finite-dimensional models (see *Stark* [1992a] for general methods, *Stark* [1992b] for a method using smoothness assumptions, and *Pulliam and Stark* [1993] and *Stark* [1993] for examples of how the uncertainties in truncated spherical harmonic models can be misleading, especially when the spatial sampling is uneven).

When such an analysis is performed, we will be able to tell whether and to what extent mantle tomography can constrain mantle convection models. In the meanwhile, we think suggestions (such as that of *Olson et al.* [1990]) that tomography has essentially revealed the large-scale pattern of flow in the mantle are premature.

Another promising direction is to abandon travel times in favor of other functionals of the seismogram that are more sensitive to structure in the parts of the mantle we wish to image (e.g., *Gee and Jordan* [1992]). We feel this is likely to enhance the sensitivity of hypothesis tests about features of mantle velocities, but we do not think the signal-to-noise and sampling are adequate to image the entire mantle with any confidence, no matter how sophisticated the data analysis.

7 Conclusions

Even if one assumes that mantle velocity structure is relatively smooth, piecewise constant in 4872 10° by 10° voxels, and that the uncertainties of travel times are relatively small (0.25 s for direct P and 0.5 s for all other phases), the departure of least-squares estimates of mantle velocity from a radially symmetric model is smaller than the 95% uncertainty in 87% to 90% of the mantle's volume. Based on 889,909 summary P rays, 84,046 summary PP rays, 67,228 summary pP rays, 23,024 summary PcP rays, 11,239 summary $PKPab$ rays, 47,227 summary *PKPbc* rays and 141,843 summary *PKPdf* rays), the 95% uncertainty in P velocity exceeds 0.193km s^{-1} in half of the mantle, and exceeds 0.593km s^{-1} in a quarter of the mantle, by volume. Based on 163,354 summary S rays, 13,781 summary SS rays, 7,441 summary ScS rays and 12,494 summary sS rays, the 95% uncertainty in S velocity exceeds 0.254km s^{-1} in half of the mantle, and exceeds 0.554km s^{-1} in a quarter of the mantle.

It might be possible to constrain seismic velocity fairly well in small areas if one assumes seismic velocity varies smoothly with position, but the uncertainty of estimates of global features of the mantle from travel-time data using least-squares and ray theory are larger than plausible three-dimensional variations. New approaches using more accurate approximations of the physics than ray theory to estimate specific functionals of mantle seismic velocity, rather than the entire velocity structure, appear to hold the greatest promise for making useful inferences about the mantle from travel-time data.

Acknowledgements. We are grateful to Cray Research, Inc. for time on a Cray Y-MP C90, to Steve Larson, Steve Luzmoor and Michael Stewart for help with the computations, and to Christopher Genovese, Niklaus Hengartner and Terry Speed for helpful discussions. Part of this work was performed while RJP was a postdoctoral researcher at the Geophysics Program, Univ. of Washington and the IRIS DMC, Seattle and he gratefully acknowledges their support. This work was supported by the NSF through PYI award DMS-8957573 and grants DMS-8810192 and INT-9205103.

A Neglecting Correlated Errors

We claimed that assuming that the summary-ray travel time errors are independent with standard deviation 0.25 s for direct P phases or 0.5 s for S and core and reflected phases, and ignoring the correlated errors introduced by source and station corrections, inaccuracies in the *iasp91* model, etc., gives optimistic uncertainties. We prove here that provided the travel time errors we include are independent of those we neglect, the uncertainty is optimistic.

The uncertainty as measured by the reproducing kernel is biggest perturbation to the model that does not change the predicted data by more than χ^2 in the norm weighted by the covariance matrix; i.e., we are concerned with the maximum of a linear functional over the set

$$\mathcal{X}_\Lambda \equiv \left\{\beta \in \mathbf{R}^n : \beta^T A^T \Lambda^{-1} A\beta \leq \chi^2\right\}, \tag{13}$$

where Λ is the (diagonal) covariance matrix of the errors we consider. Suppose there are additional errors independent of the first set, with covariance matrix Σ. By independence, the covariance matrix for the combined errors is $\Lambda+\Sigma$, so the "true" set we ought to consider is

$$\mathcal{X}_{\Lambda+\Sigma} \equiv \left\{\beta \in \mathbf{R}^n : \beta^T A^T (\Lambda+\Sigma)^{-1} A\beta \leq \chi^2\right\}. \tag{14}$$

The apparent uncertainty neglecting the second set of errors will certainly be smaller than the "true" uncertainty incorporating those errors if $\mathcal{X}_\Lambda \subset \mathcal{X}_{\Lambda+\Sigma}$. This in turn holds if

$$\beta^T A^T \Lambda^{-1} A\beta \geq \beta^T A^T (\Lambda+\Sigma)^{-1} A\beta, \tag{15}$$

which we show is true. Since Λ and Σ are covariance matrices, they are positive definite and symmetric, and we can invert them and take their square-roots. Define $y = \Lambda^{-1/2} A\beta$. Note that

$$(\Lambda+\Sigma)^{-1} = (\Lambda^{1/2}(I+\Lambda^{-1/2}\Sigma\Lambda^{-1/2})\Lambda^{1/2})^{-1} \tag{16}$$

$$= \Lambda^{-1/2}(I+\Lambda^{-1/2}\Sigma\Lambda^{-1/2})^{-1}\Lambda^{-1/2}. \tag{17}$$

The matrix $\Lambda^{-1/2}\Sigma\Lambda^{-1/2}$ is positive definite, and so has an eigenvalue-eigenvector decomposition

$$\Lambda^{-1/2}\Sigma\Lambda^{-1/2} = \Omega^T\Delta\Omega,$$

where Ω is an orthogonal matrix and Δ is a diagonal matrix with strictly positive diagonal elements. Since $(I+\Lambda^{-1/2}\Sigma\Lambda^{-1/2})$ is positive definite, we can write

$$(I+\Lambda^{-1/2}\Sigma\Lambda^{-1/2})^{-1} = I - \Omega^T\Delta\Omega + (\Omega^T\Delta\Omega)^2 - \cdots \tag{18}$$

$$= \sum_{j=0}^{\infty}\Omega^T(-\Delta)^j\Omega \tag{19}$$

$$= \Omega^T\left(\sum_{j=0}^{\infty}(-\Delta)^j\right)\Omega \tag{20}$$

$$\equiv \Omega^T\Upsilon\Omega, \tag{21}$$

where Υ is a diagonal matrix whose diagonal elements are strictly between zero and one. Since Ω is orthogonal,

$$\beta^T A^T\Lambda^{-1}A\beta = \|y\|^2 \tag{22}$$

$$= \|\Omega y\|^2 \tag{23}$$

$$= y^T\Omega^T I\Omega y \tag{24}$$

$$\leq y^T\Omega^T\Upsilon\Omega y \tag{25}$$

$$= y^T\Lambda^{1/2}\Lambda^{-1/2}(I+\Lambda^{-1/2}\Sigma\Lambda^{-1/2})^{-1}\Lambda^{-1/2}\Lambda^{1/2}y \tag{26}$$

$$= \beta^T A^T(\Lambda+\Sigma)^{-1}A\beta, \tag{27}$$

as claimed.

References

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen. *LAPACK User's Guide.* SIAM, Philadelphia, 1992.

[2] R. Buland and C.H. Chapman. The computation of seismic travel times. *Bull. Seis. Soc. Am.*, 73:1271–1302, 1983.

[3] R.W. Clayton and R.P. Comer. A tomographic analysis of mantle heterogeneities from body wave travel time data. *EOS Trans. Am. Geophys. Un.*, 64:776, 1983.

[4] J.H. Davies, O. Gudmundsson, and R. W. Clayton. Spectra of mantle shear wave velocity structure. *Geophys. J. Intl.*, 108:865–882, 1992.

[5] A.M. Dziewonski. Mapping the lower mantle: Determination of lateral heterogeneity in P velocity up to degree and order 6. *J. Geophys. Res.*, 89:5929–5952, 1984.

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

[7] L.S. Gee and T.H. Jordan. Generalized seismological data functionals. *Geophys. J. Intl.*, 11:363–390, 1992.

[8] O. Gudmundsson, 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. Intl.*, 102:25–43, 1990.

[9] H. Inoue, Y. Fukao, K. Tanabe, and Y. Ogata. Whole mantle P-wave travel time tomography. *Phys. Earth Planet. Inter.*, 59:294–328, 1990.

[10] B.L.N. Kennett and E.R. Engdahl. Traveltimes for global earthquake location and phase identification. *Geophys. J. Intl.*, 105:429–465, 1991.

[11] P. Olson, P.G. Silver, and R.W. Carlson. The large-scale structure of convecton in the Earth's mantle. *Nature*, 344:209–215, 1990.

[12] C. C. Paige and M.A. Saunders. LSQR, an algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Software*, 8:43–71, 1982.

[13] R.J. Pulliam. *Imaging Earth's interior: Tomographic inversions for mantle P-wave velocity structure.* PhD thesis, University of California, Berkeley, 1991.

[14] R.J. Pulliam and P.B. Stark. Bumps on the core-mantle boundary: Are they facts or artifacts? *J. Geophys. Res.*, 98:1943–1956, 1993.

[15] R.J. Pulliam, D. Vasco, and L.R. Johnson. Tomographic inversions for mantle P-wave velocity structure based on the minimization of l^2 and l^1 norms of ISC travel time residuals. *J. Geophys. Res.*, 98:699–734, 1993.

[16] M.K. Sengupta and M.N. Toksõz. Three dimensional model of seismic velocity variation in the earth's mantle. *Geophys. Res. Lett.*, 3:84–86, 1976.

[17] P.G. Silver, R.W. Carlson, and P. Olson. Deep slabs, geochemical heterogeneity, and the large-scale structure of mantle convection: Investigation of an enduring paradox. *Ann. Rev. Earth Planet. Sci.*, 16:477–541, 1988.

[18] P.B. Stark. Inference in infinite-dimensional inverse problems: Discretization and duality. *J. Geophys. Res.*, 97:14,055–14,082, 1992.

[19] P.B. Stark. Minimax confidence intervals in geomagnetism. *Geophys. J. Intl.*, 108:329–338, 1992.

[20] P.B. Stark. Uncertainty of the COBE quadrupole detection. *Ap. J. Lett.*, 408:L73, 1993.

[21] P.B. Stark and N.W. Hengartner. Reproducing Earth's kernel: Uncertainty of the shape of the core-mantle boundary from PKP and PcP travel times. *J. Geophys. Res.*, 98:1957–1972, 1993.

[22] P.B. Stark and D.I. Nikolayev. Toward tubular tomography. *J. Geophys. Res.*, 98:8095–8106, 1993.

[23] A. van der Sluis and H.A. van der Vorst. Numerical solution of large, sparse linear systems arising from tomographic problems. In G. Nolet, editor, *Seismic Tomography With Applications in Global Seismology and Exploration Geophysics*, pages 53–87. D. Reidel Publishing Co., Dordrecht, 1987.

[24] D.W. Vasco, R.J. Pulliam, and L.R. Johnson. Formal inversion of ISC arrival times for mantle P-velocity structure. *Geophys. J. Intl.*, in press, 1993.

[25] R.L. Woodward and G. Masters. Global upper mantle structure from long-period differential travel times. *J. Geophys. Res.*, 96:6351–6377, 1991.

[26] R.L. Woodward and G. Masters. Lower mantle structure from ScS-S differential travel times. *Nature*, 352:231–233, 1991.

Figure Captions

Fig. 1. Schematic of least-squares simultaneous inference. The data are fitted by least-squares to estimate the velocity at a set of positions r_j in the mantle. In general, the estimates have different variances σ_j^2, depending on details of the data mapping matrix (e.g., how many summary rays pass through the voxel containing r_j). The square-root of the "diagonal" of the reproducing kernel, $\sqrt{K(r_j, r_j)}$, renormalizes the estimates to have the same variance. I.e., if $\hat{v}_j$ is the (random) estimate of the velocity at position r_j, $\hat{v}_j/\sqrt{K(r_j, r_j)}$ has the same variance for all j. Thus a 95% confidence interval for any $\hat{v}_j/\sqrt{K(r_j, r_j)}$ has the same length. The estimates v_j are correlated, but not perfectly, so not all confidence intervals will contain the "true" values v_j at the same time. As a result, the simultaneous coverage probability of the set of individual 95% confidence intervals is less than 95%. To ensure that all the confidence intervals simultaneously contain the true values v_j 95% of the time, we need to enlarge all the intervals to account for the lack of perfect correlation. The enlargement factor depends on the number of estimated parameters in the model.

phase	rays	summary rays
P	4,569,294	889,909
PP	141,526	84,046
pP	133,102	67,228
PcP	53,891	23,024
$PKPab$	34,571	11,239
$PKPbc$	305,241	47,227
$PKPdf$	436,666	141,843
total P	5,674,291	1,264,516

S	591,369	163,354
SS	25,510	13,781
ScS	12,226	7,441
sS	20,054	12,494
total S	649,159	197,070

Table 1: Number of phases of various types in the data set, identified from over 46,000 events in the ISC catalog from years 1964-1987. Column 1: phase. Column 2: number of rays. Column 3: number of summary rays derived from this phase.

layer	depth	volume
1	0 – 35	0.039
2	35 – 200	0.084
3	200 – 400	0.103
4	400 – 660	0.123
5	660 – 860	0.069
6	860 – 1060	0.075
7	1060 – 1260	0.081
8	1260 – 1460	0.087
9	1460 – 1860	0.128
10	1860 – 2260	0.102
11	2260 – 2460	0.089
12	2460 – CMB	0.020
total		1.000

Table 2: Parameters of the voxel model, and sampling of the voxels by compressional and shear phases. The angular dimensions of the voxels are roughly 10° by 10° at the equator. Column 1: depth range of the layer. Column 2: fraction of the mantle's total volume in the layer. There were 406 voxels in each of the twelve layers, giving a total of 4872 voxels in the mantle model. Voxels comprising about 97% of the mantle's volume were sampled by a P summary ray. All voxels were sampled by some compressional summary ray. Voxels comprising about 95% of the mantle's volume were sampled by an S summary ray. Voxels comprising 99.96% of the mantle's volume were sampled by some shear summary ray.

fraction	P	$+PP$	$+PcP$	$+pP$	$+PKP_{bc}$	$+PKP_{ab}$	$+PKP_{df}$
0.00	0.025	0.025	0.025	0.025	0.025	0.025	0.025
0.25	0.109	0.102	0.101	0.099	0.098	0.098	0.096
0.50	0.230	0.207	0.205	0.203	0.197	0.196	0.193
0.75	0.723	0.598	0.597	0.594	0.594	0.594	0.593
1.00	∞	107.1	107.1	107.0	100.3	100.1	99.8

Table 3: Distribution of the half-width of a 95% simultaneous confidence region for the mantle using compressional-wave observations. Column 1: Fraction of the mantle's volume for which the half-width of a 95% simultaneous confidence region is less than or equal to the entries in subsequent columns. Column 2: Region half-width (in km s^{-1}) based on *P* summary rays alone. Column 3: Half-width using *P* and *PP* summary rays. Column 4: Half-width using *P, PP* and *pP* summary rays. Column 5: Half-width using *P, PP, pP* and *PcP* summary rays. Column 6: same as column 5, but with *PKPbc* summary rays too. Column 7: same as column 6, but with *PKPab* summary rays too. Column 8: same as column 7, but with *PKPdf* summary rays too. The voxels sampled by *P* alone comprise 97.18% of the mantle's volume; when any of the other compressional phases are included, all the voxels in the model are sampled.

fraction	S	$+SS$	$+ScS$	$+sS$
0.00	0.036	0.035	0.034	0.031
0.25	0.217	0.148	0.142	0.134
0.50	0.585	0.281	0.265	0.254
0.75	1.986	0.594	0.568	0.554
1.00	∞	∞	∞	∞

Table 4: Distribution of the half-width of a 95% simultaneous confidence region for the mantle using shear observations. Column 1: Fraction of the mantle's volume for which the half-width of a 95% simultaneous confidence region is less than or equal to the entries in subsequent columns. Column 2: Region half-width (in km s^{-1}) based on S summary rays alone. Column 3: Region half-width based on S and SS summary rays. Column 4: same as column 3, but with ScS summary rays as well. Column 5: same as column 4, but with sS summary rays too. The fraction of the mantle's volume in this parametrization sampled by S alone is 94.9%; 99.88% is sampled by S and SS together, and 99.96% is sampled by S, SS, ScS (with or without sS).

layer	min	25%	50%	75%	max
1	0.856	3.400	7.233	17.405	99.809
2	0.094	0.519	1.387	4.612	21.496
3	0.100	0.385	1.017	2.728	9.001
4	0.060	0.186	0.448	1.069	3.479
5	0.050	0.120	0.290	0.588	2.527
6	0.045	0.113	0.239	0.474	1.593
7	0.048	0.116	0.229	0.403	1.267
8	0.052	0.107	0.204	0.352	1.039
9	0.029	0.052	0.098	0.158	0.333
10	0.025	0.048	0.081	0.130	0.344
11	0.027	0.054	0.091	0.131	0.347
12	0.029	0.061	0.095	0.133	0.348
all	0.025	0.096	0.193	0.593	99.809

Table 5: Layer by layer uncertainties (in km s^{-1}) of P velocity in the mantle using all the compressional summary rays: *P, PP, pP, PcP, PKPab, PKPbc,* and *PKPdf.* Together these phases sample voxels comprising 100% of the mantle's volume.

layer	min	25%	50%	75%	max
1	0.349	2.337	5.315	12.563	56.795
2	0.058	0.335	1.007	3.011	15.417
3	0.086	0.319	0.798	1.938	7.990
4	0.062	0.192	0.423	0.872	3.475
5	0.052	0.129	0.272	0.488	1.663
6	0.065	0.140	0.259	0.414	1.375
7	0.052	0.139	0.240	0.422	1.139
8	0.049	0.146	0.238	0.413	0.871
9	0.031	0.080	0.124	0.214	0.410
10	0.038	0.088	0.150	0.222	0.568
11	0.046	0.115	0.203	0.321	2.634
12	0.046	0.147	0.250	0.455	∞
all	0.031	0.134	0.254	0.554	∞

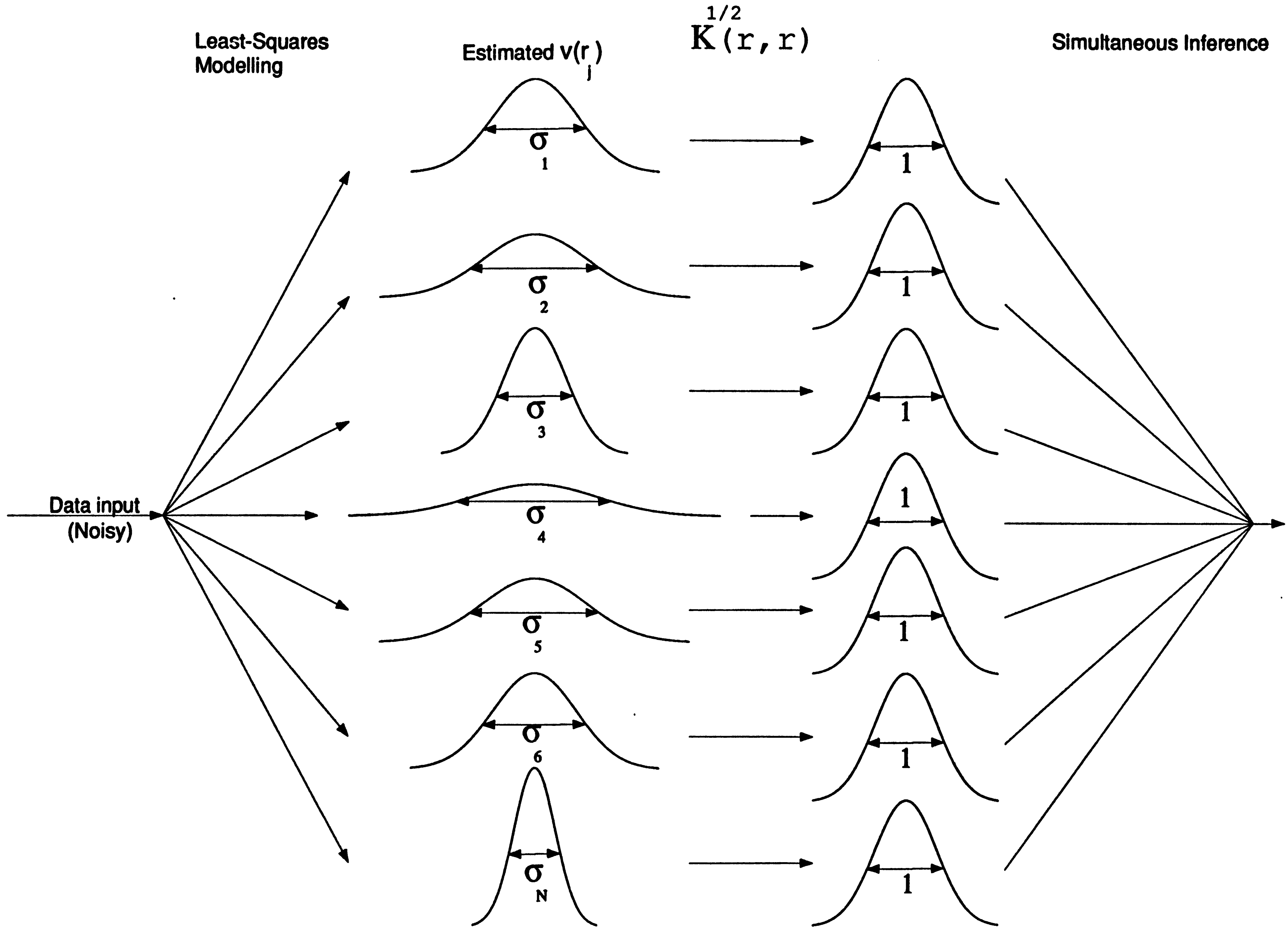
Table 6: Layer by layer uncertainties (in km s^{-1}) of S velocity in the mantle using all the shear summary rays: *S, SS, ScS,* and *sS*. Together these phases sample voxels comprising 99.96% of the mantle's volume.

data	bottom	top	N edge	S edge	E edge	W edge	S/N ratio
P	2270km	1870km	0°	−10°	180°	−170°	4.70
all P	1870km	1470km	0°	−10°	−60°	−70°	4.47
S	1870km	1470km	20°	10°	−106°	−116°	392
all S	1870km	1470km	−60°	−70°	24°	0°	22.6

Table 7: Locations of voxels most significantly different from the spherically symmetric reference model, using weighted least-squares and discarding residuals larger than ± 10 s and the corresponding rays . Column 1: data set. First row uses only direct P phases; second row refers to results using the compressional phases *P, PP, pP, PcP, PKPab, PKPbc,* and *PKPdf*; third row uses only direct S phases; and the last row uses *S, SS, ScS,* and *sS* phases. The next four columns give the boundaries of the voxels: bottom radius, top radius, North edge (latitude), South edge (latitude), East edge (E longitude) and West edge (E longitude). The final column gives the ratio of the inferred velocity anomaly (in km s^{-1}) to the half-width of the 95% simultaneous confidence envelope at that location. Note that the ratio is larger for the shear data than for the P data; this reflects in part the fact that fewer voxels are sampled by shear than by compressional phases, so fewer parameters are estimated, and also reflects the generally lower quality of ISC shear phase picks, which we optimistically underestimate to be 0.5 s.

case	S alone	all shear phases
a	9.99%	11.1%
b	9.45%	3.45%
c	7.63%	4.64%
d	7.68%	3.45%

Table 8: Volume fraction of the mantle for which the least-squares estimate of S perturbations lies outside a 95% simultaneous confidence region around the 1-dimensional *iasp91* model. Column 1: treatment of travel time residuals. In case (a), residuals greater in absolute value than 10s were discarded, along with the rays corresponding to them. Case (b) is the same as case (a), except that the residuals beyond 7s were discarded. In case (c) residuals beyond $\pm$7s were truncated to $\pm$7s, and their corresponding rays were retained in computing the confidence regions. Case (d) is the same as case (b) except a small constant was added to the diagonal of the Gram matrix before inverting it (damped least-squares). In case (d), the damping was included in computing the confidence region, which reduces its width. Column 2 shows results for models fitted to the direct S summary residuals alone. Column 3 shows results for models fitted simultaneously to all the shear summary rays: *S, SS, ScS,* and *sS*.



Least-Squares Modelling
Estimated $v(r_j)$
$K^{1/2}(r, r)$
Simultaneous Inference
Data input (Noisy)
σ_1
σ_2
σ_3
σ_4
σ_5
σ_6
σ_N
1
1
1
1
1
1
1