

# Multiscale seismic tomography

Ling-Yun Chiao<sup>1</sup> and Ban-Yuan Kuo<sup>2</sup>

<sup>1</sup>*Institute of Oceanography, National Taiwan University, Taipei, Taiwan. E-mail: chiao@ccms.ntu.edu.tw*

<sup>2</sup>*Institute of Earth Sciences, Academia Sinica, Taipei, Taiwan*

Accepted 2000 December 5. Received 2000 September 5; in original form 2000 March 1

## SUMMARY

Seismic traveltime tomography is commonly discretized by a truncated expansion of the pursued model in terms of chosen basis functions. Whether parametrization affects the actual resolving power of a given data set as well as the robustness of the resulting earth model has long been seriously debated. From the perspective of the model resolution, however, there is one important aspect of the parametrization issue of seismic tomography that has yet to be systematically explored, that is, the space–frequency localization of a chosen parametrization. In fact, the two most common parametrizations tend to enforce resolution in each of their own particular domains. Namely, parametrization in terms of spherical harmonics with global support tends to emphasize spectral resolution while sacrificing the spatial resolution, whereas the compactly supported pixels tend to behave in the opposite manner. Some of the significant discrepancies among tomographic models are very likely to be manifestations of this effect, when dealing with data sets with non-uniform sampling. With an example of the tomographic inversion for the lateral shear wave heterogeneity of the D'' layer using *S*–*SKS* traveltimes, we demonstrate an alternative parametrization in terms of the multiresolution representation of the pursued model function. Unlike previous attempts of multiscale inversion that invoke pixels with variable sizes, or overlay several layers of tessellation with different grid intervals, our formulation invokes biorthogonal generalized Harr wavelets on a sphere. We show that multiresolution representation can be constructed very easily from an existing block-based discretization. A natural scale hierarchy of the pursued model structure constrained by the resolving power of the given sampling is embedded within the solution obtained. It provides a natural regularization scheme based on the actual ray-path sampling and is thus free from *a priori* prejudices intrinsic to most regularization schemes. Unlike solutions obtained through spherical harmonics or spherical blocks that tend to collapse structures onto ray paths, our parametrization imposes regionally varying Nyquist limits, that is, robustly resolvable local wavelength bands within the obtained solution.

**Key words:** D'' layer, inverse theory, tomography.

## 1 INTRODUCTION

Ever since the early phase of modern global tomographic study, the dichotomy among approaches that invoke different parametrizations has been obvious (Dziewonski 1982, 1984; Clayton & Comer 1983). With the large amount of seismic travel-time measurements available today, more and more detailed tomographic images of the Earth have been published each year. However, inconsistencies among these recent models with high nominal resolutions have become a controversial issue that demands to be resolved (e.g. Dziewonski & Woodhouse 1987; Morelli & Dziewonski 1987a,b; Tanimoto 1990; Woodward & Masters 1991; Pulliam & Stark 1993; Stark & Hengartner 1993; Wang & Zhou 1993; Su *et al.* 1994; Morelli & Dziewonski 1995;

Stark 1995; Masters *et al.* 1996; Zhou 1996; Grand *et al.* 1997; Bijwaard *et al.* 1998; Boschi & Dziewonski 1999). Different data sets, numerical algorithms of inversion, parametrizations and regularization schemes are among the major factors that these discrepancies might arise from. The latter two factors have been at the centre of disputes that have attracted considerable attention. It is noted that although seismic tomography is in essence a continuous inverse problem, the data kernel based on ray theory is, however, not band-limited. This precludes the direct evaluation of the *Gram* matrix, which consists of inner products of data kernels. Discretization through finite parametrization of the pursued model is thus inevitable. Rendering the continuous model function into a finite set of parameters, it is clear that any finite parametrization invokes an

implicit regularization scheme that imposes selective weightings on different model components. The intertwined effects from parametrization and regularization further complicate the interpretation and comparison among earth models obtained by different groups. Clearly, to have a solution with the resolving power that is compatible with the actual sampling while avoiding either implicit or explicit extra unjustifiable prejudices should be the main concern of choosing a particular type of basis function to execute the finite parametrization. In this study, we first review briefly some of the problems associated with the general finite parametrization. An alternative parametrization based on spherical wavelets expansion is then introduced and invoked in a tomographic study of the lateral shear wave heterogeneity of the D'' layer. Solutions obtained from parametrizations based on the three different types of basis functions, namely, spherical harmonics, spherical pixels and spherical wavelets, are compared and discussed.

## 2 THE GENERAL PARAMETRIZATION

Given a finite amount of observations,  $d_i$  with observational error  $e_i$ ,  $i = 1 \dots N$ , the goal of a continuous geophysical inverse problem is usually to estimate the model function  $m(\mathbf{r})$  with the spatial variables  $\mathbf{r}$ , governed by the data rule:

$$d_i = (g_i(\mathbf{r}), m(\mathbf{r})) + e_i, \quad (1)$$

where inner products  $(g_i(\mathbf{r}), m(\mathbf{r})) = \int g_i(\mathbf{r})m(\mathbf{r})d\mathbf{r}$  and  $g_i(\mathbf{r})$  is the  $i$ th data kernel or the Fréchet derivative of a non-linear data functional. It has been shown that the natural unbiased representation of components of  $m(\mathbf{r})$  resolvable by the given data is

$$m(\mathbf{r}) = \sum_{i=1}^N \alpha_i g_i(\mathbf{r}), \quad (2)$$

since any component orthogonal to the subspace spanned by  $g_i(\mathbf{r})$  is in the null space (Parker 1977), which makes no contribution to the data. The coefficients  $\alpha_i$  are then obtained by solving

$$\mathbf{d} = \mathbf{G}\boldsymbol{\alpha}. \quad (3)$$

The Gram matrix,  $\mathbf{G}$ , is defined with elements  $G_{ij} = (g_i(\mathbf{r}), g_j(\mathbf{r}))$ .

For seismic tomography based on ray theory, the complete evaluation of  $\mathbf{G}$  is usually not accessible. It is then inevitable that the model function is parametrized in terms of a chosen set of orthonormal basis functions  $\phi_l(\mathbf{r})$ ,  $(\phi_l(\mathbf{r}), \phi_j(\mathbf{r})) = \delta_{lj}$  such that

$$m(\mathbf{r}) = \sum_{l=1}^L \beta_l \phi_l(\mathbf{r}) + m^*(\mathbf{r}), \quad (4)$$

with the truncation level  $L$ . The expansion remainder is  $m^*(\mathbf{r}) = \sum_{k=L+1}^{\infty} \beta_k \phi_k(\mathbf{r})$ , with  $\infty$  designating higher-mode contributions beyond the truncation level. The common practice is then simply to ignore the expansion remainder and solve for the expansion coefficients  $\beta_l$ ,  $l = 1 \dots L$ , governed by

$$\mathbf{d} = \mathbf{A}_L \boldsymbol{\beta}_L. \quad (5)$$

The elements of the  $N \times L$  matrix  $\mathbf{A}_L$  are defined by  $(\mathbf{A}_L)_{il} = (g_i(\mathbf{r}), \phi_l(\mathbf{r}))$ . It should be pointed out that  $L$  is usually chosen arbitrarily, while the substitution of eq. (1) or (3) by eq. (5) is in fact not formally justified. Eqs (3) and (5) can be

related by noting that (Trampert & Snieder 1996)

$$\mathbf{G} = \mathbf{A}_L \mathbf{A}_L^T + \mathbf{A}_{\infty} \mathbf{A}_{\infty}^T, \quad (6)$$

with elements of  $\mathbf{A}_{\infty}$  defined by inner products of data kernels and truncated higher modes of the expansion basis functions. In other words, truncated expansion of the model function can be interpreted as a truncated expansion of the data kernel, which leads to an approximation of the Gram matrix  $\tilde{\mathbf{G}} = \mathbf{A}_L \mathbf{A}_L^T$ . The damped least-squares (DLS) inverse operator based on this approximation is thus

$$\tilde{\mathbf{G}}^{-1} = (\mathbf{A}_L \mathbf{A}_L^T + \theta^2 \mathbf{I})^{-1}, \quad (7)$$

and the actual solution in terms of the coefficients  $\hat{\boldsymbol{\beta}}$  obtained is

$$\hat{\boldsymbol{\alpha}} = \tilde{\mathbf{G}}^{-1} \mathbf{d}, \quad \hat{\beta}_l = (\phi_l, \hat{m}) = \sum_{i=1}^N \hat{\alpha}_i (\phi_l, g_i),$$

i.e.

$$\begin{bmatrix} \hat{\beta}_L \\ \hat{\beta}_{\infty} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_L^T \\ \mathbf{A}_{\infty}^T \end{bmatrix} \tilde{\mathbf{G}}^{-1} \mathbf{d} = \mathbf{R} \begin{bmatrix} \beta_L \\ \beta_{\infty} \end{bmatrix}, \quad (8)$$

$$\mathbf{R} = \begin{bmatrix} \mathbf{A}_L^T \tilde{\mathbf{G}}^{-1} \mathbf{A}_L & \mathbf{A}_L^T \tilde{\mathbf{G}}^{-1} \mathbf{A}_{\infty} \\ \mathbf{A}_{\infty}^T \tilde{\mathbf{G}}^{-1} \mathbf{A}_L & \mathbf{A}_{\infty}^T \tilde{\mathbf{G}}^{-1} \mathbf{A}_{\infty} \end{bmatrix}.$$

Note that on the right-hand side of eq. (8), the resolving relation between the estimated  $\hat{\boldsymbol{\beta}}_L$  and the actual  $\boldsymbol{\beta}_L$  (upper submatrix,  $\mathbf{R}_{11}$  and  $\mathbf{R}_{12}$ ) corresponds to the familiar solution of the discretized inverse problem, since

$$\begin{aligned} \mathbf{A}_L^T (\mathbf{A}_L \mathbf{A}_L^T + \theta^2 \mathbf{I})^{-1} (\mathbf{A}_L + \mathbf{A}_{\infty}) \\ = (\mathbf{A}_L^T \mathbf{A}_L + \theta^2 \mathbf{I})^{-1} (\mathbf{A}_L^T \mathbf{A}_L + \mathbf{A}_L^T \mathbf{A}_{\infty}). \end{aligned} \quad (9)$$

Although it is clear from eq. (8) that  $\mathbf{R}$  is the complete *resolving operator*, only  $\mathbf{R}_{11}$  has been conventionally examined in appraising the solution obtained. The second term ( $\mathbf{R}_{12} = (\mathbf{A}_L^T \mathbf{A}_L + \theta^2 \mathbf{I})^{-1} \mathbf{A}_L^T \mathbf{A}_{\infty}$ ) contributes to the previously discussed potential spectral leakage (Trampert & Snieder 1996). It stems from the unclear resolution of the  $\hat{\boldsymbol{\beta}}_L$  component from  $\boldsymbol{\beta}_{\infty}$ , arising from non-orthogonality among the adopted modes and the truncated higher modes for the sampling portrayed by the given data set (i.e. elements of  $\mathbf{A}_L^T \mathbf{A}_{\infty} \neq 0$ ).

### 2.1 Comparison between different parametrizations

Interestingly, since the truncated expansion is now interpreted simply as a means of approximating the Gram matrix (eq. 6), it is possible to examine the obtained solution in terms of other expansions. For example, suppose that in evaluating the approximating Gram matrix  $\tilde{\mathbf{G}}$ , the matrix  $\mathbf{A}^1$  with elements  $A_{il}^1 = (g_i(\mathbf{r}), \phi_l^1(\mathbf{r}))$  in eq. (7) is implemented by choosing the spherical harmonic functions,  $\phi_l^1(\mathbf{r})$ , in a global tomography problem. Once  $\tilde{\mathbf{G}}$  and thus the appropriate  $\tilde{\mathbf{G}}^{-1}$  are calculated, a different set of basis functions, for example, spherical blocks ( $\phi_l^2(\mathbf{r})$ ), can be used to calculate a different  $\mathbf{A}^2$  matrix in computing the final model structure using eq. (8). That is, all the matrices  $\mathbf{A}_L$ ,  $\mathbf{A}_{\infty}$  in eq. (8) are substituted by the second

parametrization, while  $\tilde{\mathbf{G}}^{-1}$  is calculated by the first parametrization. This suggests that the final solution obtained through a particular chosen discretization can be inspected in terms of other different expansions up to any truncation level. This might be useful in comparing solutions obtained through different discretizations. More importantly, it indicates that good resolution obtained through one particular parametrization does not imply that the resolving kernel ( $\mathbf{R}$ ) will behave satisfactorily in another domain. For example, parametrization in terms of spherical harmonics tends to have better spectral resolution but usually poor spatial resolution whereas the pixels discretization is vice versa. This is another way of interpreting the spurious artefacts within sampling gaps associated with spherical harmonic discretization that lacks spatial resolution. Since model components represented by these two common parametrizations are theoretically equivalent when the sampling density is spatially uniform (Senso 1990), the inconsistency among inversion results derived from different parametrizations manifests itself only with non-uniform sampling. However, Boschi & Dziewonski (1999) showed that over-damping can effectively diminish the inconsistency even for highly non-uniform sampling. One explanation is that over-damping will in fact severely sacrifice the resolving power of the given data set such that only relatively long-wavelength model components that are equally effectively represented by the two different parametrizations are retained.

Another interesting point is that although the approximation of the Gram matrix (eq. 7) is formulated with the expansion truncated at the level  $L$ , it does not prevent the evaluation of those expansion coefficients of the higher modes. Keeping only  $\hat{\beta}_L$  in the final solution, as has been common practice, is thus invoking unjustifiable *a posteriori* smoothing that is responsible for unrealistic spurious structures appearing within data gaps (Pulliam & Stark 1993). In fact, the  $\hat{\beta}_\infty$  part of the solution is usually not negligible. While a solution based only on  $\hat{\beta}_L$  that embeds spurious structures within data gaps might indicate a satisfactory fit to the data, incorporating contributions from  $\hat{\beta}_\infty$  will diminish those artefacts but usually deteriorate the fit significantly. As will be discussed later, this commonly exercised procedure, which retains only  $\hat{\beta}_L$  is very different from parametrizing the inverse problem with a high truncation level ( $L$ ) and then synthesizing the final solution up to a lower level ( $l_{\text{syn}} < L$ ). The lower synthesizing level might be based on the aim of further reducing the model variance of the obtained solution (Kuo & Wu 1997).

## 2.2 Spectral leakage

Trampert & Snieder (1996) proposed an anti-leakage operator by minimizing both the data misfit and the contribution from the  $\mathbf{R}_{12}$  term. They have shown with a surface wave tomographic example that the construction (their eq. 9),

$$\hat{\beta}_L = [\mathbf{A}_L^T \mathbf{W}^{-1} \mathbf{A}_L + (\theta_L^2 / \theta_\infty^2) \mathbf{I}]^{-1} \mathbf{A}_L^T \mathbf{W}^{-1} \mathbf{d}, \quad (10)$$

has been very successful in annihilating the spectral leakage, where

$$\mathbf{W}^{-1} = (\mathbf{A}_\infty \mathbf{A}_\infty^T + \theta_\infty^2 \mathbf{I})^{-1} \quad (11)$$

is the anti-leakage operator. However, let  $\mathbf{C}_M = (\theta_\infty^2 / \theta_L^2) \mathbf{I}$  and thus  $\mathbf{C}_M^{-1} = (\theta_L^2 / \theta_\infty^2) \mathbf{I}$ . A little algebraic manipulation of eq. (10)

yields

$$\begin{aligned} \hat{\beta}_L &= [\mathbf{A}_L^T \mathbf{W}^{-1} \mathbf{A}_L + \mathbf{C}_M^{-1}]^{-1} \mathbf{A}_L^T \mathbf{W}^{-1} \mathbf{d} \\ &= \mathbf{C}_M \mathbf{A}_L^T (\mathbf{A}_L \mathbf{C}_M \mathbf{A}_L^T + \mathbf{W})^{-1} \mathbf{d} \\ &= \mathbf{A}_L^T [\mathbf{A}_L \mathbf{A}_L^T + (\theta_L^2 / \theta_\infty^2) \mathbf{A}_\infty \mathbf{A}_\infty^T + \theta_L^2 \mathbf{I}]^{-1} \mathbf{d}. \end{aligned} \quad (12)$$

In other words, the solution obtained through the anti-leakage operation (10) is equivalent to the solution obtained by implementing an alternative inverse operator of the Gram matrix of the form

$$\tilde{\mathbf{G}}^{-1} = [\mathbf{A}_L \mathbf{A}_L^T + (\theta_L^2 / \theta_\infty^2) \mathbf{A}_\infty \mathbf{A}_\infty^T + \theta_L^2 \mathbf{I}]^{-1} \quad (13)$$

Numerically, the computational advantage of eqs (10) and (11) over the schematic expression eq. (12) is obvious. However, it is interesting to note that for  $0 < (\theta_L^2 / \theta_\infty^2) < 1$ , the formulation can be interpreted as special cases of the weighted least-squares solution of the original continuous inverse problem. In addition to the minimum-norm damping (the third term in eq. 13), the second term describes the relative weighting among the adopted modes and those truncated. In fact, differences among formulations with different parametrization and regularization schemes can usually be compared in terms of having different forms of quelling (Backus 1970; Chou & Booker 1979) that implement each of their own particular weighting functions (Meyerholtz *et al.* 1989). Most of these particular weighting functions are, however, based on *a priori* prejudice that is seldom physically justifiable.

Another interesting practice implied by eq. (13) is that instead of actually carrying the parametrization to infinitely high modes, the  $\infty$  might indicate a level higher than the truncation level ( $L$ ) that we are actually interested in. In other words, the anti-leakage scheme of Trampert & Snieder (1996) is, in essence, consistent with our general consensus that even though we are only comfortable with conservatively inverting for relatively lower-mode components of the model function, the discretization still has to be carried up to sufficiently high modes to avoid spectral leakage as much as possible. Kuo & Wu (1997) invoked such a procedure to obtain what they thought were robust features of lateral shear wave heterogeneity within the  $D''$  layer. As discussed earlier, this is very different from invoking only low modes in the parametrization and evaluating the solution obtained in terms of those modes that were actually activated in the initial parametrization. In contrast, Boschi & Dziewonski (1999) favoured the conclusion that with sufficient damping, long-wavelength components of an over-parametrized model are consistent with those obtained from an underparametrized model, and that aliasing (our spectral leakage) problems are negligible. Again, we argue that this is possibly due to the effect of over-damping essentially giving up resolving power of shorter-wavelength model components of the data such that the non-orthogonality among higher and lower modes no longer manifests itself in the final solution. One risk, however, is that over-damping might also underestimate the long-wavelength components significantly.

In summary, the aim of devising a parametrization scheme that is equipped with an intrinsic natural regularization and takes into consideration the dual resolution of frequency-space motivates us to explore the fast-developing theory of multi-resolution analysis (Mallat 1989). The seismic tomography example considered in this study utilizes *S-SKS* differential

traveltime residuals to constrain lateral shear wave heterogeneity within the D" layer. This problem has been tackled by Kuo & Wu (1997) and Kuo *et al.* (2000) by spherical harmonic functions so that interesting comparisons can be executed. To proceed with the alternative multiresolution parametrization on the spherical surface, an implementation of spherical wavelet functions has to be devised.

### 3 FINITE PARAMETRIZATION BY SPHERICAL WAVELETS

Discrete wavelets in the 1-D and 2-D plane settings have developed rapidly in the past two decades. However, the characteristics of classical wavelets constructed from dyadic translates and dilates of one particular function have prohibited its generalization to general manifolds. In a series of studies, Sweldens (1995, 1996) proposed the lifting scheme to build so-called second-generation wavelets that are fast to compute and easily applicable to general manifolds. Schröder & Sweldens (1995) constructed a suite of spherical wavelets for efficiently representing functions on the sphere. Based on their work, we modify the construction to be compatible with seismic tomography around the spherical surface.

The construction of spherical wavelets starts with building almost equal-area patches or faces that cover the spherical surface. Beginning with a regular geodesic polyhedron (e.g. an icosahedron in Fig. 1a), successive refinement by subdividing each spherical triangle into four children leads to a spherical tessellation depending on the refinement level (Fig. 1b). For example, from coarse levels ( $j$ ) to finer levels ( $j+1$ ), spherical triangle  $\Delta_T^j$  consists of four subtriangles  $\Delta_0^{j+1}$ ,  $\Delta_1^{j+1}$ ,  $\Delta_2^{j+1}$ ,  $\Delta_3^{j+1}$ . Each subtriangle can then be further refined to finer levels  $j+2$ ,  $j+3$ ... etc. For a particular ray path (bold dotted trace in Fig. 1a), the traveltime residual ( $d$ ) is related to the average slowness ( $v_i^{j+1}$ ,  $i=1, 2, 3$ ) and the path length ( $l_i^{j+1}$ ) by

$$\dots + l_0^{j+1} v_0^{j+1} + l_1^{j+1} v_1^{j+1} + l_2^{j+1} v_2^{j+1} + l_3^{j+1} v_3^{j+1} + \dots = d. \quad (14)$$

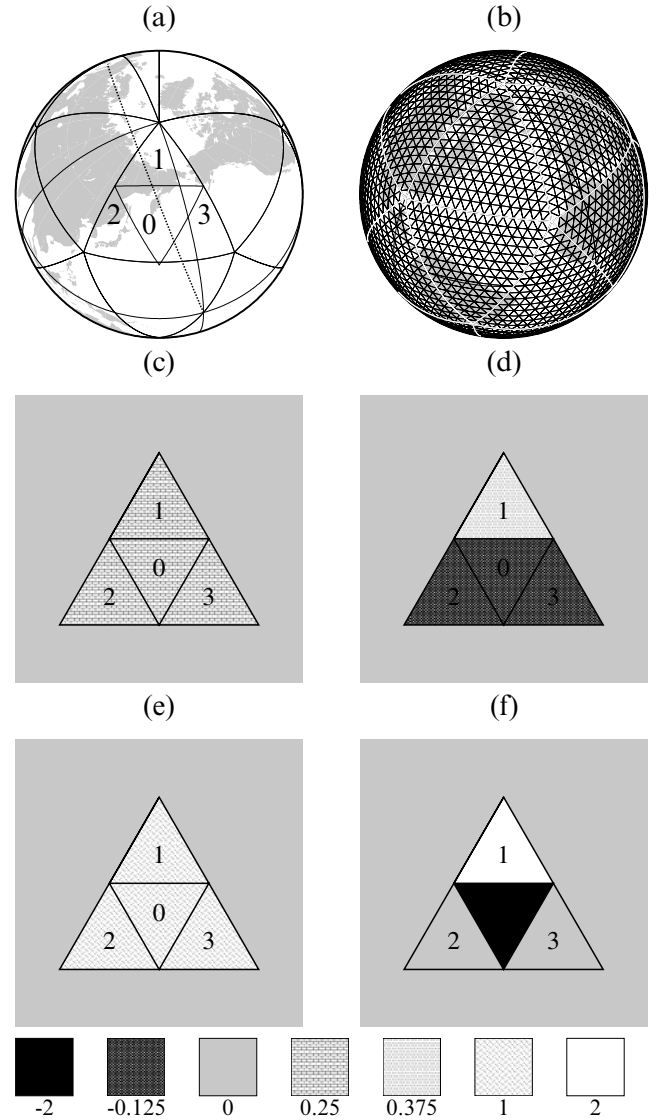
Define a multiresolution representation of the slowness,

$$\begin{cases} \lambda_T^j = \frac{1}{A_T^j} (A_0^{j+1} v_0^{j+1} + A_1^{j+1} v_1^{j+1} + A_2^{j+1} v_2^{j+1} + A_3^{j+1} v_3^{j+1}) \\ \gamma_m^j = \frac{1}{2} (v_m^{j+1} - \lambda_T^j), \end{cases} \quad (15)$$

where  $\lambda_T^j$  is the average slowness in the level  $j$ , and  $\gamma_m^j$  is used to parametrize the deviation of the slowness at the level  $j+1$  interpolated from level  $j$ . Because the areas  $A_0^{j+1}$ ,  $A_m^{j+1}$  of  $\Delta_0^{j+1}$ ,  $\Delta_m^{j+1}$ ,  $m=1, 2, 3$ , are slightly different, compensating weights are incorporated.  $A_T^j$  is the spherical area of  $\Delta_T^j$ . Similarly, the resident path length can be spanned with the dual basis and the expansion coefficients would be

$$\begin{cases} \tilde{\lambda}_T^j = (l_0^{j+1} + l_1^{j+1} + l_2^{j+1} + l_3^{j+1}) \\ \tilde{\gamma}_m^j = 2 \left( l_m^{j+1} - \frac{A_m^{j+1}}{A_0^{j+1}} l_0^{j+1} \right) \end{cases} \quad (16)$$

Simple decomposition (from finer scale level  $j+1$  to coarse-scale  $j$ ) and synthesizing (coarse to fine) algorithms defined in eqs (15) and (16) correspond to the Bio-Harr basis functions



**Figure 1.** Spherical tessellation constructed from the root level with a spherical icosahedron. (a) Geodesic triangles (20 faces of the icosahedron) are successively subdivided into four subtriangles. An example of this refinement from coarse level (level  $j$ ) to finer level (level  $j+1$ ) is indicated by  $\Delta_T^j = \Delta_0^{j+1} + \Delta_1^{j+1} + \Delta_2^{j+1} + \Delta_3^{j+1}$ . The dotted line represents an example of a ray path. (b) 5120 spherical triangles at refinement level 5. (c) Magnitude response of the primary scaling function used for multiresolution representation of the slowness model. (d) Magnitude of one of the three primary wavelets at the same scale level as in (c). (e), (f) Magnitudes of the dual scaling function and one of the dual wavelets.

(Schröder & Sweldens 1995) of a particular multiresolution representation on the sphere.  $\lambda_T^j$  in eq. (15) is the expansion coefficient of the primary scaling function  $\phi_T^j$  (Fig. 1c) at scale level  $j$ , whereas the three  $\gamma_m^j$  are coefficients of the primary wavelets  $\psi_m^j$  (Fig. 1d). Similarly,  $\tilde{\lambda}_T^j$  and  $\tilde{\gamma}_m^j$  in eq. (16) are coefficients of the dual scaling function  $\tilde{\phi}_T^j$  (Fig. 1e) and the dual wavelets  $\tilde{\psi}_m^j$  (Fig. 1f). Schröder & Sweldens (1995) showed that although Bio-Harr wavelets have only one vanishing moment, more vanishing moments could actually be added to a new multiresolution analysis by the dual lifting scheme.

Let the inner product on the sphere be represented by  $(f, g) = \int f g d\Omega$ ; it is straightforward to verify that

$$\begin{cases} (\phi_T^j, \tilde{\phi}_T^j) = \delta_{T,T'} \\ (\psi_m^j, \tilde{\psi}_{m'}^j) = \delta_{m,m'} \delta_{j,j} \\ (\phi_T^j, \tilde{\psi}_m^j) = 0 \\ (\psi_m^j, \tilde{\phi}_T^j) = 0 \end{cases} \quad (17)$$

Based on the biorthogonality of eq. (17), eq. (14) can be written as

$$\dots + \tilde{\lambda}_T^j \lambda_T^j + \sum_{m=1}^3 \gamma_m^j \tilde{\gamma}_m^j + \dots = d. \quad (18)$$

Biorthogonality implied by eqs (17) and (18) is very important. It suggests that the actual implementation of the multi-resolution parametrization can be easily incorporated in an existing block parametrization. The direct implementation follows the general expansion by going through eqs (4) and (5) and constructing the Gram matrix by evaluating the path-length for each of the basis functions. This requires considerable efforts of tedious book keeping. Alternatively, we can construct the Gram matrix for the spherical blocks by evaluating the path integral (eq. 14) on the finest level (leaf level). Note that each row of this matrix corresponds to one particular ray. The elements of each column are path lengths within one particular block on the finest level. Simply performing wavelet decomposition for each row by the dual decomposition (eq. 16) to the coarse level (root level) is equivalent to rewriting eq. (14) as eq. (18). That is, after this reconfigured matrix is inverted, the solution automatically gives the coefficients of the multiresolution representation of the slowness model. Synthesis from the root level to the leaf level can then be performed to inspect the spatial distribution. Both the decomposition and the synthesis procedures are straightforward to implement and numerically fast to compute. Note that in this implementation, the slowness decomposition (eq. 15) leads to an unbiased estimate of the average slowness at each scale-level, whereas the length of the ray path increases from fine to coarse scales (eq. 16). That is, depending on the sampling, local long-wavelength components will be better constrained. This implies a natural, data adaptive weighting scheme toward robust and relatively long-wavelength components. Unlike imposing other *a priori* smoothness preferences on the pursued model variation, no external criterion on the form of a particular smoothing is invoked along with the parametrization. Furthermore, since the weighting depends on the accumulation of path length and is thus sampling-adaptive, it imposes non-stationary, heterogeneous smoothing in accord with the locally resolvable bandwidth instead of enforcing a stationary band throughout the whole region of interest.

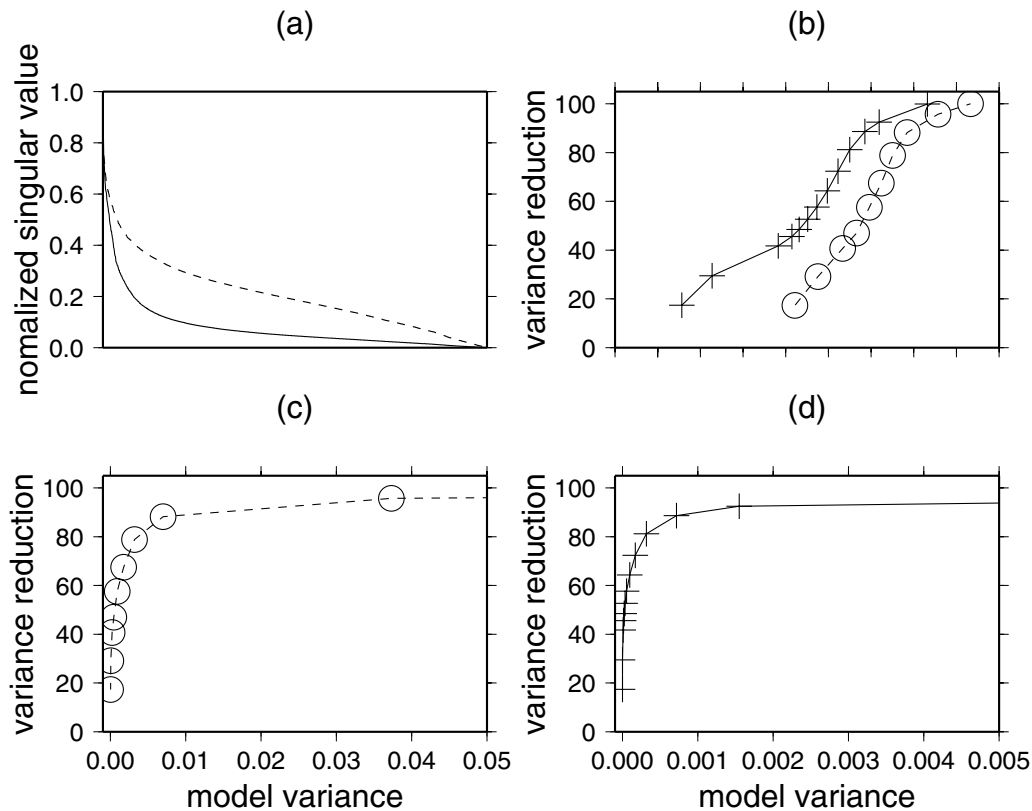
#### 4 AN EXAMPLE OF MULTISCALE TOMOGRAPHY: SHEAR VELOCITY HETEROGENEITY WITHIN THE D" LAYER

Tomographic experiments using the *S-SKS* differential travel-time data set (Kuo & Wu 1997; Kuo *et al.* 2000) to constrain the lateral shear wave heterogeneity of the D" layer are carried

out using the multiscale parametrization described above. A five-level refinement of the initial icosahedron results in a tessellation with 5120 ( $= 20 \times 4^{(5-1)}$ ) spherical triangles covering the entire spherical surface (Fig. 1). A simple block-type inversion is first performed based on a discretization using these spherical triangles as pixels. The approximated Gram matrix, with elements defined by lengths of ray paths within each triangle (eq. 14), is then reconfigured through the decomposition process from the leaf (finest) level to the root level according to eqs (16) and (18). The reconfigured, or decomposed, Gram matrix is constructed very quickly in a straightforward, *in situ* fashion. The reconfiguration does not hinder the matrix inversion via an iterative scheme such as the LSQR method (Paige & Saunders 1982) for sizeable problems. However, in this study, for the sake of exploring other theoretical characteristics, all the solutions are obtained by singular value decomposition (SVD), no matter which parametrization the approximated Gram matrix is based upon. After the SVD, a threshold that is a fraction of the maximum singular value is set up, such that only those singular values, greater than the threshold are kept to synthesize the inverse operator.

Based on the truncated singular-values construction, some characteristics of the solution obtained from the reconfigured Gram matrix (based on the multiresolution expansion) and the original matrix (based on the original spherical triangles discretization) can be compared. Note first that the singular-value spectrum (Fig. 2a) is modified as a result of regrouping the resolvable components in terms of coefficients of the chosen basis functions. Standard variance reduction versus model covariance trade-off curves for both solutions are generated by adjusting different thresholds of the singular-value truncation (Figs 2b, c and d). The variance reduction is simply defined by data fitting expressed as a percentage of the  $L_2$  norm of the data. Assuming uniform, uncorrelated error statistics for traveltimes measurements, the model covariance is evaluated by the diagonal sum of the square matrix  $\mathbf{G}^{-1}(\mathbf{G}^{-1})^T$  (Menke 1984), where  $\mathbf{G}^{-1}$  is the generalized inverse operator of the Gram matrix  $\mathbf{G}$  formed by the singular-value truncation described above. Note that the model covariance obtained through inverting the reconfigured Gram matrix is significantly smaller than that obtained by inverting the original Gram matrix. This is not unexpected, since the coefficients of the longer-wavelength basis functions embedded within the multi-scale hierarchy are always better constrained. This is one of the virtues of the proposed parametrization that leads to solutions more robust with respect to the potential data errors. Based on the trade-off curve (Fig. 2c) we conservatively pick the robust solution with a relatively low variance reduction (48 per cent).

The resulting shear wave velocity perturbation is displayed by synthesis from the root level, gradually incorporating finer details of higher scale levels (from top to bottom in the left column of Fig. 3). Note that the image constructed by synthesis up to scale-level 5 is essentially the same as synthesis up to scale-level 4. That is, details finer than the wavelength of scale-level 5 (approximately  $1.32^\circ$ ) are not robustly resolvable by the data. A solution with the same variance reduction but obtained by inverting the original Gram matrix, which is based on the parametrization in terms of spherical triangular pixels, is also displayed for comparison. Note that this solution can also be decomposed into the scale hierarchy (right column in Fig. 3) in a straightforward fashion according to eq. (15). The difference



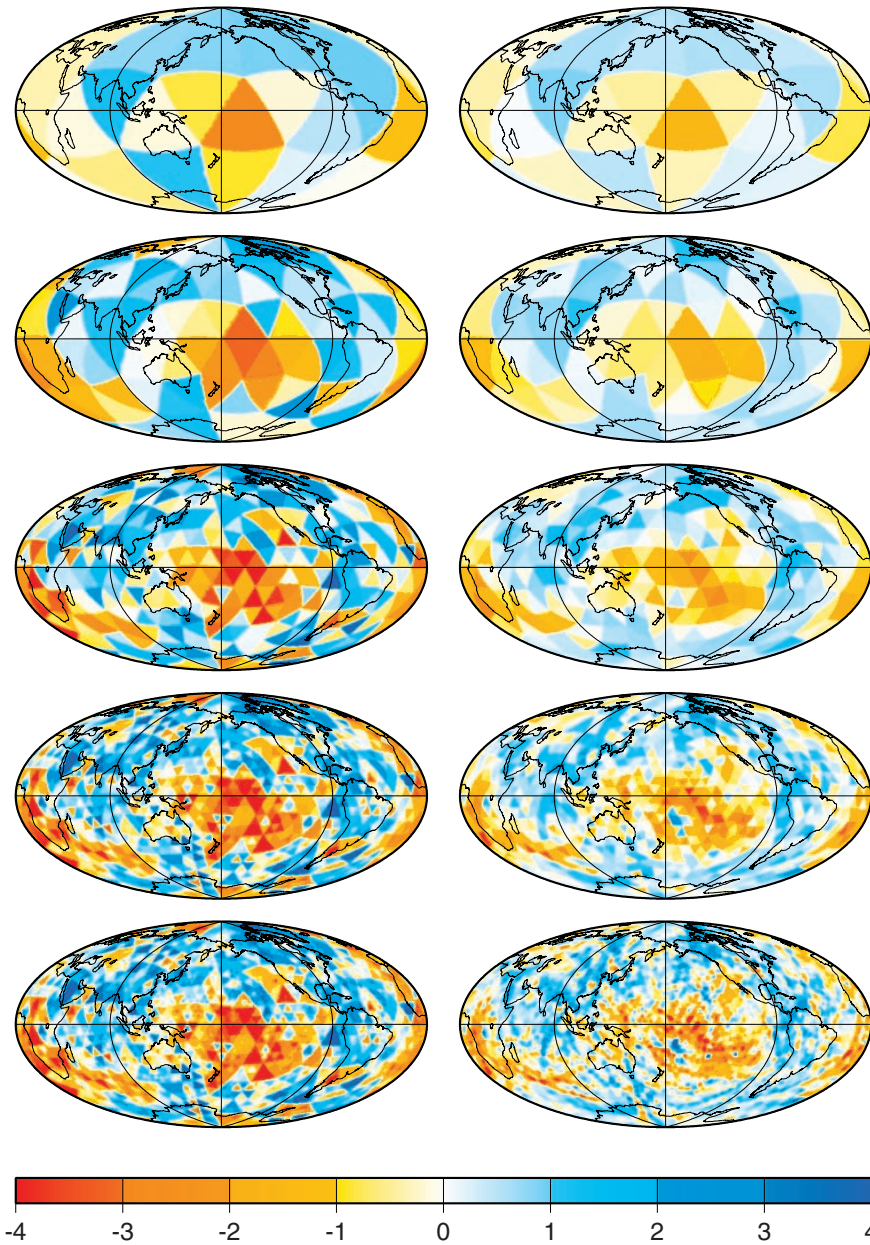
**Figure 2.** Comparison of tomography of  $S$ -SKS differential traveltimes parametrized by spherical pixels (dash line and circle symbol) and spherical wavelets (solid line and cross symbol) that is constructed (see the text) from spherical pixels. (a) Normalized singular value spectrum (by the maximum singular value) indicates the effect of reconfiguration of the Gram matrix. (b) Multiscale parametrization yields models with model variance significantly lower (an order of magnitude) at the same level of variance reduction. Note that the horizontal scale is model variance on a logarithmic scale. (c), (d) The variance reduction versus model variance trade-off curves for the spherical pixels and the spherical wavelets formulations.

between solutions obtained via these two different parametrizations is clear (compare the left and right columns in Fig. 3). We find that unlike the spherical wavelets solution, which has heterogeneous resolvable scales, the spherical pixels solution tends to collapse significant structures gradually along the ray paths. Furthermore, magnitudes of long-wavelength (low scale-level) components of the spherical pixels solution are considerably lower than the spherical wavelets solution, a point we will come back to for more discussions.

Another solution with a similar variance reduction that is parametrized in terms of the globally supported spherical harmonics up to the 40th degree is also obtained. The comparison among solutions obtained via the three different parametrizations is shown in Fig. 4. The overall spatial patterns of the three solutions are similar, with remarkable clustering of the calculated plume roots (Steinberger & O'Connell 1998) around the low-velocity anomalies. However, there are significant discrepancies among these images. Note that the data set has been carefully sorted to ensure that the sampling coverage is as uniform as possible (Kuo *et al.* 2000) which explains the consistency between the spherical harmonics solution and the spherical pixels solution. Otherwise, with the presence of large data gaps or regionally redundant sampling, it is well known that considerable spurious artefacts will appear within data gaps for solutions parametrized by globally supported basis functions (Pulliam & Stark 1993), unless the inversion is heavily damped (Boschi & Dziewonski 1999). The major difference between the three solutions, however, is that while

both the spherical harmonics solution and the spherical pixels solution tend to collapse structures along the ray paths, the grouping of local structures into longer wavelengths in the spherical wavelets solution is different.

Other than the overall spatial pattern, the level-wise contributions to the variance reduction, the root-mean-square model norm of the three solutions as well as the power spectrum when projected onto the spherical harmonics expansion are also compared (Fig. 5). Note that from the discussion in Section 2.1, it is possible to project the spherical harmonics solution onto a representation in terms of spherical pixels and thus perform the level-wise decomposition. Inspecting the variance reduction of the spherical harmonics solution, projected and decomposed, from the root level and gradually incorporating higher scale-level details, it is found that contributions from scale-level 5 actually deteriorate the data fitting. This is caused by the fact that the spherical harmonics parametrization was carried only up to the 40th degree, with the degrees of freedom less than the level 5 refinement of the spherical pixels discretization. From eq. (8) and the subsequent discussions, the projection of the spherical harmonics solution onto the spherical pixels representation invokes components higher than the truncation level (degree 40) that are usually assumed negligible when constructing the final solution. It is also noted that there are significant contributions on the variance reduction from fine structures in scale-level 5 for both the spherical harmonics and the spherical pixels solution (Fig. 5a). In fact, if the discretization had been carried to even higher scale levels, this



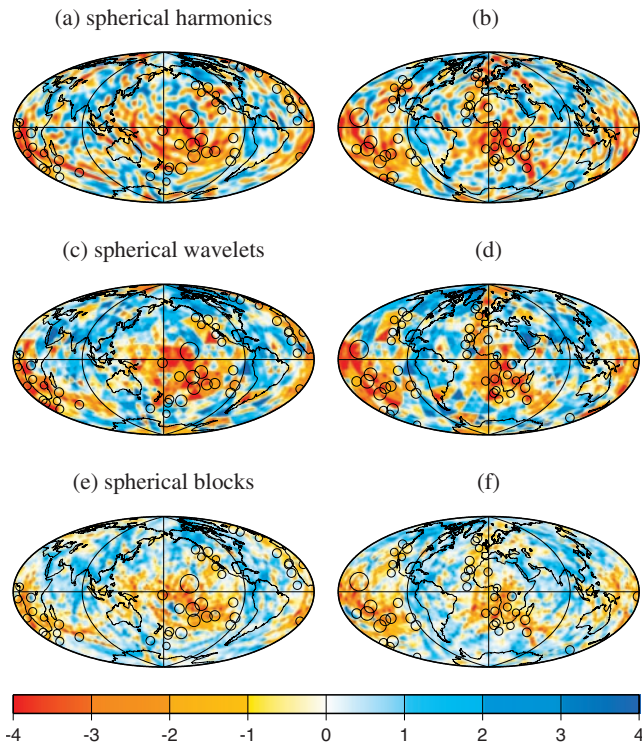
**Figure 3.** Based on the trade-off portrayed in Fig. 2(d), a robust solution with variance reduction of 48 per cent is chosen to examine images obtained by model parametrizations in terms of spherical wavelets (left column) and spherical blocks (right column). From top to bottom, the shear wave perturbation within the  $D''$  layer is synthesized up to higher scale levels to incorporate structures with finer details. Note that structures up to scale-level 5 are almost identical to the image portrayed up to the level 4 on the left column. That is, finer details above scale-level 4 are not robustly resolvable with the given data set. On the other hand, the contribution from finer details up to scale-level 5 is still significant for the spherical pixels parametrization on the right.

trend would still persist such that eventually all structures are collapsed onto the ray paths. This is, however, not the case for the spherical wavelets solution, where the model contributions needed to satisfy the data essentially peak at scale-level 4, suggesting that a global maximum resolvable wavelength is no shorter than the characteristic wavelength of this scale level.

Another way of further inspecting these three solutions is to examine their level-wise contribution of the  $L_2$  model norm. In Fig. 5(b), root mean squares of the  $L_2$  model norm are presented for each scale level. Again, it is clear that the contribution from scale-level 5 is insignificant for the spherical wavelets solution, while it tends to be important for the other

two common parametrizations. All three solutions are also projected into the spherical harmonics domain to compare their power spectra (Fig. 5c). They are all clearly dominated by the degree 2 component. However, magnitudes of the degree 2 power for each solution are quite different. First, it is noted that the spherical pixels solution has lower degree 2 magnitude than the spherical harmonics solution. The reason for this discrepancy is that the degrees of freedom invoked by the former is greater than the latter (5120 versus 1681). It is already mentioned that the data kernels based on ray theory are not band-limited in the across-ray directions. In other words, there will always be energy spread through shorter and shorter





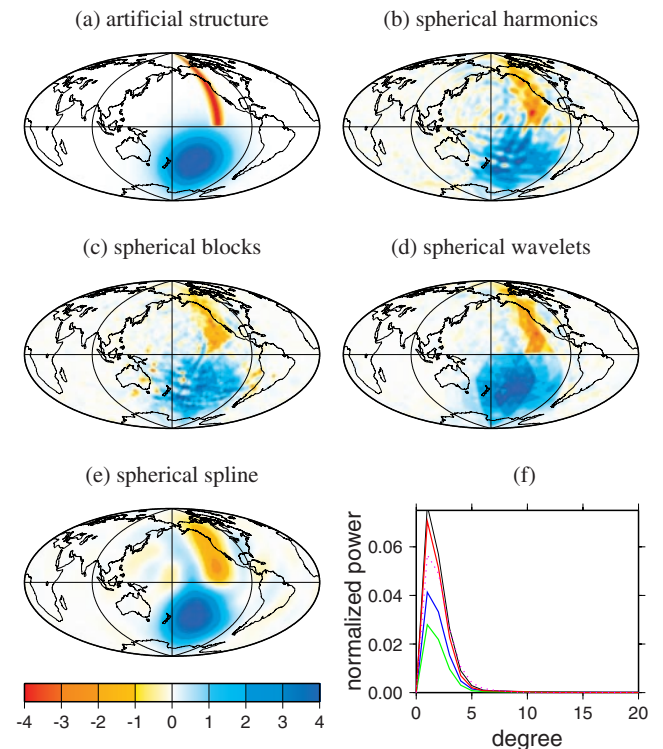
**Figure 4.** Comparison of solutions with the same variance reduction (48 per cent), but parametrized in terms of different basis functions. (a), (b) Tomographic inversion parametrized by spherical harmonics up to the 40th degree. (c), (d) Spherical wavelets parametrization (the same as those in the left column of Fig. 3). (e), (f) Spherical pixels (same as the right column of Fig. 3). Images in the left column (a, c, e) have map projections centred on longitude  $180^\circ$ , while those in the right column are centred on  $0^\circ$ . Circles are positions of the plume roots at the core-mantle boundary calculated by Steinberger & O'Connell (1998). Note the apparent effect of the ray traces within solutions based on both spherical harmonics and spherical pixels, while the grouping of structures within the spherical wavelets solution is quite different.

wavelengths to put structures onto the ray paths. Since these short-wavelength components are not orthogonal to the long-wavelength components with a non-uniform sampling coverage, the more the energy spread through higher modes, the more the energy that belongs to the lower modes will be peeled away along with the increasing damping. That is, with the same variance reduction, the solution parametrized with higher degrees of freedom will always end up with lower power in the lower modes (see also examples in Kuo *et al.* 2000). One noticeable difference of the proposed multiscale parametrization from the other two is that this alternative parametrization distinctly preserves the magnitudes of the long-wavelength components.

#### 4.1 Inversion experiment for synthesized data generated from artificial structures

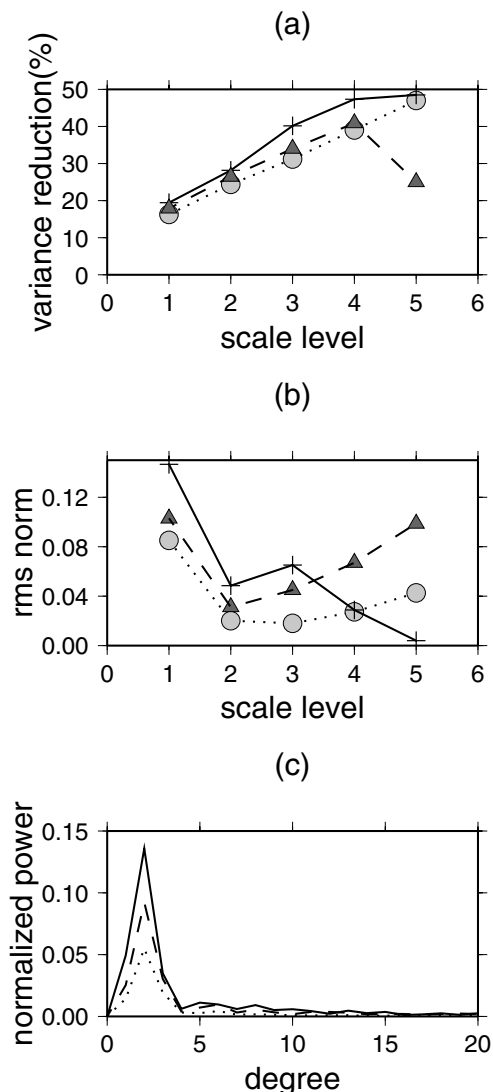
The *S-SKS* traveltime data set is inevitably noisy since it is collected through real geophysical observations with considerable effort. To further clarify the characteristics of the tomographic solutions obtained through these different parametrizations, it is suggested that a comparison experiment should be executed on a synthesized data set that is generated

from known artificial structures. We have designed such an experiment with a relatively broad positive shear wave speed perturbation centred at the south Pacific, and a narrow belt of negative perturbation across the north Pacific (Fig. 6a). The reason for choosing such structures instead of the commonly adopted checkerboard test is that we wish to explore the potentially non-stationary resolving capability intrinsic to standard tomographic problems. We utilize the same ray-path sampling as in the real data set to generate traveltime anomalies to be inverted by the three formulations based on different parametrizations. The solutions obtained are compared in Fig. 6. For the relatively broad positive structure, the scattered images obtained through spherical pixels and truncated spherical harmonics indicate the obvious effect of the ray sampling, whereas the multiscale solution shows a natural grouping into locally coherent structures but still reflects the rigours of the local sampling. For the relatively narrow belt of negative perturbation, the grouping is also obvious. However, note that at places along the belt where the sampling is adequate, both the conventional block and spectral solutions are capable of recovering high-resolution variations in the across-belt direction. This capability is not sacrificed with the multiscale parametrization. The heterogeneous, sampling-adaptive resolving



**Figure 6.** Comparison of solutions obtained by different parametrizations for artificial, noise-free data. The variance reductions are all around 80 per cent. (a) The artificial structures used to generate the synthesized data. There is a broad Gaussian variation with  $\sigma = 30^\circ$  centred at  $(150^\circ\text{W}, 40^\circ\text{S})$  and a narrow belt of negative perturbation of shear wave speed across the north Pacific. (b) Tomographic inversion parametrized by spherical harmonics up to the 40th degree. (c) Spherical pixels parametrization. (d) Spherical wavelets. (e) Spherical splines (see text). (f) Power spectra of the original structure and from solutions of tomographic inversions (black: original structure; red: spherical wavelets; magenta: spherical spline; blue: spherical harmonics; green: spherical pixels).





**Figure 5.** Further comparison of characteristics of solutions obtained by the three different parametrizations. The spherical pixels solution is marked by a dotted line and grey circles, while the spherical harmonics solution is marked by a dashed line with solid triangles and the spherical wavelets solution by a solid line with crosses. (a) Successive contributions on the variance reduction from structures of different scale levels. For the spherical wavelets solution, detailed structures finer than scale-level 4 have negligible contributions, while the finer details are still significant for the spherical pixels solution (see also Fig. 3). Note that scale-level 5 contributions of the spherical harmonics solution actually deteriorate the data fit (see the text). (b) Root-mean-square norms of the three different solutions. Again, note that structures from scale-level 5 of the spherical wavelets solution is insignificant, while this is not the case for the other two solutions. (c) Power spectrum of the three solutions. While they are all characterized by the obvious degree 2 dominance, the absolute degree 2 magnitudes are quite different (see discussions in the text).

capability of the multiscale parametrization is revealed in this experiment. In contrast, if *a priori* smoothness preference is invoked, a uniform global smoothing will be enforced. We test this effect by a spherical spline parametrization that is implemented by simply adding an  $l(l+1)$  weighting to each degree of the expansion harmonics (see Parker 1994). The smoothing effect on the resulting image (Fig. 6e) is obvious.

Although it seems to offer a better recovery of the positive structure, the narrow belt of the negative anomaly broadens considerably, and loses the peak amplitude along the belt in general.

## 5 DISCUSSION

It was pointed out earlier that the conventional DLS solution, equipped with the minimum  $L_2$  norm regularization, of the common spherical harmonics or the spherical pixels parametrization will tend to collapse all structures onto the ray paths. In the early years of global seismic tomography, relatively fewer degrees of freedom were invoked in studies parametrized with the globally supported spherical harmonics. The rationale was to aim at resolving conservatively yet robustly the long-wavelength components of Earth structures. However, it is becoming clear that such practices are prone to contamination by spectral leakage effects. General consensus has been established that even if only the relatively low modes or long-wavelength features are the pursued structures of interest, the parametrization still has to be carried to finer details as far as possible. However, as indicated previously, since the travel-time data kernel is not band-limited in the across-ray direction based on ray theory, finer parametrization of tomographic problems with non-uniform sampling will always result in significant structures in the short-wavelength components that eventually collapse structures completely on the ray paths. One misleading consequence of the obtained earth model with high nominal resolution is then interpreting the high-frequency components resulting from the ray distribution as meaningful Earth structures. This phenomenon might not be easily detected since ray coverage in standard tomographic problems is usually highly non-uniform, which means that the locally resolvable bandwidths are regionally variable. Consequently, one important point worth re-emphasizing is that, even with the currently available abundance of traveltimes measurements and the continually improving computational resources that enable finer and finer parametrizations, the nominal resolution of a parametrized tomographic image does not necessarily imply the actual resolution of the pursued earth structure.

Another related problem with fine parametrization inverted along with the minimum-norm regularization is that, with finer and finer parametrizations and thus increasing degrees of freedom, the magnitudes of the long-wavelength components tend to decrease rapidly. As discussed previously, this can be attributed to spreading energies throughout the spectrum. It might be argued that this phenomenon and the collapsing onto the ray paths are both due to the regularization scheme based on the model-norm minimization. Invoking alternative regularization schemes, such as the first or second derivative type of minimization (e.g. VanDeCar & Snieder 1994), or more generally the quelling operator (Chou & Booker 1979; Meyerholtz *et al.* 1989), might avoid these effects. However, minimum norm regularization bears merit because it results in solutions embedding the least contributions from the null space. In contrast, quelling types of regularization interpolate the model function where it is not constrained by the data. In essence, the procedure of quelling modifies the data rule such that the modified data kernels are convolved with a smooth spatial function  $w(\mathbf{r})$  with local supports in both spatial and wavenumber domains. The pursued model is now the roughened

model function deconvolved by  $w(\mathbf{r})$ , that is,  $m(\mathbf{r}) * w^{-1}(\mathbf{r})$ . The implicit regularization scheme embedded with the DLS solution of the modified system is now based on the minimization of the roughened model function, that is,  $\|m * w^{-1}\|$  rather than  $\|m\|$ . It is conceivable that the modified data kernel now has an approximated finite width depending on the function  $w(\mathbf{r})$ . For example, the regularization based on minimization of the second spatial derivatives of the model function is essentially the convolution with  $w^{-1}(\mathbf{r})$  that has a spectral response characterized by  $k^2$ ,  $k$  being the wavenumber. In the case of spherical harmonics, a weighting is defined by  $l(l+1)$ , with  $l$  being the degree (Parker 1994) that is implemented in our spherical spline parametrization (Fig. 6e). This convolutional quelling procedure has been reported to be effective in avoiding the collapse of images onto the ray paths and in preserving magnitudes of long-wavelength components (Meyerholtz *et al.* 1989; Delprat-Jannaud & Lailly 1993). Although it has been argued that unlike Bayesian inference, which requires *a priori* probabilistic information about the model function (Gouveia & Scales 1997), an appropriate roughening operator  $w^{-1}(\mathbf{r})$  has to be pre-determined for this regularization. A similar formulation that invokes an *a priori* model covariance operator depending on the 'correlation length' has also been proposed (Tarantola & Nercissian 1984). It is noted that some formulation in invoking the quelling is derived directly from the original continuous problem (e.g. Chou & Booker 1979; Tarantola & Nercissian 1984) instead of applying the quelling after the finite parametrization. However, as long as the discretization is carried up to a truncation level such that the 'correlation length' is significantly wider than the shortest discretization wavelength (e.g. the grid interval of a pixels-based parametrization), the effect will be the same. The preferred  $w^{-1}(\mathbf{r})$ , or the appropriate correlation length, will strongly affect the solution obtained. In fact, the general truncated expansion, in terms of either spherical harmonics or spherical pixels, can be treated as a particular case of convolutional quelling with the characteristics of  $w(\mathbf{r})$  being a boxcar function in the spectral or spatial domain. In this interpretation, there is no problem of the infinite convergence onto the rays. However, it does not seem to be physically justifiable to choose a particular width of the weighting function  $w(\mathbf{r})$ , unless the formulation is based on other finite theory (e.g. Marquering *et al.* 1999; Hung *et al.* 2000) rather than the ray theory. More importantly, once the preferred smoothing operator is chosen, it enforces stationary weighting on the local scale structure. This might in fact impede the ability to image critical short-wavelength structures at places where the data is capable of resolving (Megnin *et al.* 1997).

The multiscale parametrization proposed in this study can also be interpreted as a particular type of quelling that is implicitly invoked in all inversion schemes equipped with a regularization scheme. However, instead of choosing a particular *a priori* weighting function, the intrinsic scale hierarchy of the ray-path sampling determines locally the model bandwidth that is robustly resolvable. The effect of natural grouping into long-wavelength features rather than onto the ray paths and the preservation of the long-wavelength components have been demonstrated to be natural consequences of this parametrization. One important potential advantage of this parametrization is the ability to examine the non-stationary variation of spectral contents (Bergeron *et al.* 1999) directly from the obtained solution. Furthermore, although the formulation in this paper is centred on the multiresolution decomposition

of the path length, it does not prohibit similar decomposition of other general forms of data kernel. In fact, combined with the convolution quelling, which renders seismic rays into ray tubes with finite width, the same reconfiguration of the resulting Gram matrix can be undertaken. In other words, *a priori* model information based on viable physical grounds can still be incorporated.

## 6 CONCLUSIONS

Parametrization in terms of basis functions with global support tends to focus on frequency resolution and sacrifices spatial resolution. Parametrization in terms of basis functions with local support, on the other hand, does the opposite. Invoking curvature type and other smoothing regularization by way of quelling the data kernel with a finitely supported smoothing function might bring the two extremes closer together since a finite width will be imposed on the rays. However, *a priori* bandwidth has to be determined. Furthermore, this bandwidth is not flexible with respect to data sampling that varies regionally. To confront this, pixels with spatially variable size have been implemented (Fukao *et al.* 1992; Wang *et al.* 1998; Bijwaard *et al.* 1998). Zhou (1996), on the other hand, overlaid a series of tessellations with cell sizes of different scales that are not mutually orthogonal. The wavelet parametrization demonstrated in this study is data-adaptive. The spatially varying bandwidth, which is robustly resolvable by the given data, is automatically adapted by the local hierarchy portrayed by the multiresolution representation of the pursued model variation. The example of *S-SKS* traveltime tomography utilizing the multiscale parametrization has been shown to be very easily implemented. Based on an existing parametrization in terms of spherical pixels, straightforward reconfiguration of the Gram matrix yields a robust solution that is less prone to the apparent pattern of the ray distribution but will still faithfully reflect the sampling density. We intend to extend this parametrization to 3-D tomographic problems as well as to explore alternative, more sophisticated types of wavelet basis functions than the generalized Harr wavelet.

## ACKNOWLEDGMENTS

We thank Jeannot Trampert and R. Jay Pulliam as well as the editor for their valuable comments, which improved the manuscript significantly. All figures were prepared using GMT software version 3. (Wessel & Smith 1991). This research was supported by the National Science Council of the Republic of China under grant NSC 89-2611-M-002-043 and NSC 90-2116-M-001-002.

## REFERENCES

- Backus, G., 1970. Inference from inadequate and inaccurate data II, *Proc. Nat. Acad. Sci.*, **65**, 281–287.
- Bergeron, S.T., Vincent, A.P., Yuen, D.A., Tranchant, B.J.S. & Tchong, C., 1999. Viewing seismic velocity anomalies with 3-D continuous Gaussian wavelets, *Geophys. Res. Lett.*, **26**, 2311–2314.
- Bijwaard, H., Spakman, W. & Engdahl, E.R., 1998. Closing the gap between regional and global travel time tomography, *J. geophys. Res.*, **103**, 30 055–30 078.

- Boschi, L. & Dziewonski, A.M., 1999. High- and low-resolution images of the Earth's mantle: Implications of different approaches to tomographic modeling, *J. geophys. Res.*, **104**, 25 567–25 594.
- Chou, C.W. & Booker, J.R., 1979. A Backus-Gilbert approach to inversion of travel-time data for three-dimensional velocity structure, *Geophys. J. R. astr. Soc.*, **59**, 325–344.
- Clayton, R.W. & Comer, R.P., 1983. A tomographic analysis of mantle heterogeneity from body wave travel time data, *EOS, Trans. Am. geophys. Un.*, **64**, 776.
- Delprat-Jannaud, F. & Lailly, P., 1993. Ill-posed and well-posed formulations of the reflection travel time tomography problem, *J. geophys. Res.*, **98**, 6589–6605.
- Dziewonski, A.M., 1982. Mapping the lower mantle, *EOS, Trans. Am. geophys. Un.*, **63**, 1035.
- Dziewonski, A.M., 1984. Mapping the lower mantle: Determination of lateral heterogeneity in P velocity up to degree and order 6, *J. geophys. Res.*, **89**, 5929–5952.
- Dziewonski, A.M. & Woodhouse, A.H., 1987. Global images of the Earth's interior, *Science*, **236**, 37–48.
- Fukao, Y., Obayashi, M., Inoue, H. & Nenbai, M., 1992. Subducting slabs stagnant in the mantle transition zone, *J. geophys. Res.*, **97**, 4809–4822.
- Gouveia, W.P. & Scales, J.A., 1997. Resolution of seismic waveform inversion: Bayes versus Occam. *Inver. Prob.*, **13**, 323–349.
- Grand, S.P., van der Hilst, R.D. & Widiyantoro, S., 1997. High resolution global tomography: a snapshot of convection in the Earth, *Geol. Soc. Am. Today*, **7**, 1–7.
- Hung, S.-H., Dahlen, F.A. & Nolet, G., 2000. Frechet kernels for finite-frequency traveltimes—II. Examples, *Geophys. J. Int.*, **141**, 175–203.
- Kuo, B.-Y. & Wu, K.-Y., 1997. Global shear velocity heterogeneities in the D'' layer: Inversion from Sd-SKS differential travel times, *J. geophys. Res.*, **102**, 11 775–11 788.
- Kuo, B.-Y., Garnero, E.J. & Lay, T., 2000. Tomographic inversion of S-SKS times for shear velocity heterogeneity in D'': Degree 12 and hybrid models, *J. geophys. Res.*, **105**, 28 139–28 158.
- Mallat, S.G., 1989. A theory of multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Machine Intell.*, **11**, 674–693.
- Marquering, H., Dahlen, F.A. & Nolet, G., 1999. Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, *Geophys. J. Int.*, **137**, 805–815.
- Masters, G., Johnson, S., Laske, G. & Bolton, H., 1996. A shear-velocity model of the mantle, *Phil. Trans. R. Soc. Lond.*, **354**, 1385–1411.
- Megnin, C., Bunge, H.-P., Romanowicz, B. & Richards, M.A., 1997. Imaging 3-D spherical convection models: What can seismic tomography tell us about mantle dynamics? *Geophys. Res. Lett.*, **24**, 1299–1302.
- Menke, W., 1984. *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, Orlando, FL.
- Meyerholtz, K.A., Pavlis, G.L. & Szpakowski, S.A., 1989. Convolutional quelling in seismic tomography, *Geophysics*, **54**, 570–580.
- Morelli, A. & Dziewonski, A.M., 1987a. The harmonic expansion approach to the retrieval of deep Earth structure, in *Seismic Tomography*, pp. 251–274, ed. Nolet, G., Reidel, Dordrecht.
- Morelli, A. & Dziewonski, A.M., 1987b. Topography of the core–mantle boundary and lateral homogeneity of the liquid core, *Nature*, **325**, 678–683.
- Morelli, A. & Dziewonski, A.M., 1995. Comment on Reproducing Earth's kernel: Uncertainty of the shape of the core–mantle boundary from PKP and PcP travel times by P.B. Stark and N.W. Hengartner, *J. geophys. Res.*, **100**, 15 393–15 397.
- Paige, C.C. & Saunders, M.A., 1982. LSQR: an algorithm for sparse linear equations and sparse least squares, *Trans. Mart. Software*, **8**, 43–71.
- Parker, R.L., 1977. Understanding inverse theory, *Ann. Rev. Earth planet. Sci.*, **5**, 35–64.
- Parker, R.L., 1994. *Geophysical Inverse Theory*, Princeton University Press, Princeton, NJ.
- Pulliam, R.J. & Stark, P.B., 1993. Bumps on the core–mantle boundary: Are they facts or artefacts?, *J. geophys. Res.*, **98**, 1943–1956.
- Schröder, P. & Sweldens, W., 1995. Spherical Wavelets: Efficiently representing functions on the sphere, *SIGGRAPH: Computer Graphics, Proc. Ann. Conf. Ser.*, 161–172.
- Senso, F., 1990. On the aliasing problem in the spherical harmonic analysis, *Bull. Geod.*, **64**, 313–330.
- Stark, P.B., 1995. Reply to comment 'Reproducing Earth's kernel: uncertainty of the shape of the core–mantle boundary from PkP and PcP travel times' by P. B. Stark and N. W. Hengartner, *J. geophys. Res.*, **100**, 15 399–15 402.
- Stark, P.B. & Hengartner, N.W., 1993. 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.
- Steinberger, B. & O'Connell, R.J., 1998. Advection of plumes in mantle flow: implications for hotspot motion, mantle viscosity and plume distribution, *Geophys. J. Int.*, **132**, 412–434.
- Su, W.J., Woodward, R.L. & Dziewonski, A.M., 1994. Degree 12 model of shear velocity heterogeneity in the mantle, *J. geophys. Res.*, **99**, 6945–6980.
- Sweldens, W., 1995. *The Lifting Scheme: A Construction of Second Generation Wavelets*, Tech. Rept. Ind. Maths Initiative, Dept Mathematics, University of South Carolina, Columbia.
- Sweldens, W., 1996. The lifting scheme: Accustom-design construction of biorthogonal wavelets, *Appl. Comput. Harmon. Anal.*, **3**, 186–200.
- Tanimoto, T., 1990. Long-wavelength S-wave velocity structure throughout the mantle, *Geophys. J. Int.*, **100**, 327–336.
- Tarantola, A. & Nercessian, A., 1984. Three-dimensional inversion without blocks, *Geophys. J. R. astr. Soc.*, **76**, 299–306.
- Trampert, J. & Snieder, R., 1996. Model estimations biased by truncated expansions: possible artifacts in seismic tomography, *Science*, **271**, 1257–1260.
- VanDeCar, J.C. & Snieder, R., 1994. Obtaining smooth solutions to large, linear, inverse problems, *Geophysics*, **59**, 818–829.
- Wang, H. & Zhou, H., 1993. A comparison between spherical harmonic inversion and block inversion in whole mantle tomography, *EOS, Trans. Am. geophys. Un.*, **74**, 418.
- Wang, Z., Tromp, J. & Ekstrom, G., 1998. Global and regional surface-wave inversions: a spherical-spline parameterization, *Geophys. Res. Lett.*, **25**, 207–210.
- Wessel, J.K. & Smith, H.F., 1991. Free software helps map and display data, *EOS, Trans. Am. geophys. Un.*, **72**, 441, 445–446.
- Woodward, R.L. & Masters, G., 1991. Lower mantle structure from ScS-S differential travel times, *Nature*, **352**, 231–233.
- Zhou, H.-W., 1996. A high-resolution P wave model for the top 1200km of the mantle, *J. geophys. Res.*, **101**, 27 791–27 810.