Geophys. J. Int. (2011) **XXX**, 1–20

Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity

Frederik J. Simons[1], Ignace Loris[2], Guust Nolet[3], Ingrid C. Daubechies[4], S. Voronin[4], J. S. Judd[4], P. A. Vetter[4], J. Charléty[3] and C. Vonesch[4]

[1] *Department of Geosciences, Princeton University, Princeton, NJ 08544, USA. E-mail: fjsimons@alum.mit.edu*

[2] *Mathematics Department, Université Libre de Bruxelles, CP 217, Boulevard du Triomphe, 1050 Brussels, Belgium.*

[3] *GéoAzur, Université de Nice, 06560 Sophia Antipolis, France*

[4] *Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA*

Accepted 5 August 2011. Received 5 August 2011; in original form 12 April 2011

SUMMARY

We propose a class of spherical wavelet bases for the analysis of geophysical models and for the tomographic inversion of global seismic data. Its multiresolution character allows for modeling with an effective spatial resolution that varies with position within the Earth. Our procedure is numerically efficient and can be implemented with parallel computing. We discuss two possible types of discrete wavelet transforms in the angular dimension of the cubed sphere. We describe benefits and drawbacks of these constructions and apply them to analyze the information in two published seismic wavespeed models of the mantle, using the statistics of wavelet coefficients across scales. The localization and sparsity properties of wavelet bases allow finding a sparse solution to inverse problems by iterative minimization of a combination of the ℓ_2 norm of the data residuals and the ℓ_1 norm of the wavelet coefficients. By validation with realistic synthetic experiments we illustrate the likely gains of our new approach in future inversions of finite-frequency seismic data and show its readiness for global seismic tomography.

Key words: Inverse problems, seismic tomography, sparsity, spectral analysis, wavelets

1 INTRODUCTION

As long as tomographic Earth models remain the solutions to mixed-determined (Menke 1989) inverse problems (Nolet 1987, 2008) there will be disagreement over the precise location, shape, and amplitude of lateral and radial anomalies in seismic wavespeed that exist within the Earth; there will be attempts to derive the best-fitting mean structure (e.g. Becker & Boschi 2002), and the needed efforts to validate them (e.g. Capdeville et al. 2005; Qin et al. 2009; Bozdağ & Trampert 2010; Lekić & Romanowicz 2011). At the same time, patterns, second-order structure and correlations between and within models will continue to be sought with the goal of characterizing seismic heterogeneities (e.g. Passier & Snieder 1995; Bergeron et al. 1999; Hernlund & Houser 2008) or relating them to geochemical (e.g. Gurnis 1986), tectonic (e.g. Yuen et al. 2002; Becker et al. 2006), or geodynamical (e.g. Jordan et al. 1993; Piromallo et al. 2001; Houser & Williams 2009) processes. It has also become clear that model characteristics such as the power spectrum of tomographic anomalies (Chevrot et al. 1998a,b; Boschi & Dziewoński 1999) may teach us as much about the modeler's choices of parameterization and regularization as about the model, without imparting much information about the physical or statistical nature of our complex, physically and chemically differentiated system Earth — yet the latter should be our target. As pioneered by Gudmundsson et al. (1990) and Davies et al. (1992), in recent work questions about the size and scale distribution of Earth structure have more fruitfully been addressed by direct inference from the data themselves (e.g. Hedlin & Shearer 2000; Margerin & Nolet 2003; Becker et al. 2007; Garcia et al. 2009) without the detour of first deriving a global three-dimensional model and analyzing that.

By no means are the analysis and representation of volumetric properties the sole purview of seismology or geodynamics, and thus it is not surprising that there is a large literature on the subject in virtually every area of scientific inquiry (e.g. medical imaging, astronomy, cosmology, computer graphics, image processing, ...). While in prior seismological work the mathematical framework and terminology appropriate to discuss "parsimonious" parameterizations to solve inverse problems may not have been as explicit, approaches using irregular parametrizations, either in a static (chosen in advance) or dynamic sense (adapted to the data constraints during the inversion) have touched upon many aspects of the problem with which we will concern ourselves here (see, e.g. Sambridge & Rawlinson 2005, and references therein), and so have several techniques that pertain to inversions using model "simplicity" as explicit constraints (e.g. Pulliam et al. 1993; Vasco et al. 1994, 1999).

In this context much has come to be expected of the special powers of wavelets, with their built-in discriminating sensitivity to structure in the space and spatial frequency domains (Daubechies 1992; Strang & Nguyen 1997; Mallat 2008). Notwithstanding a continued interest and clear and present progress in the

field (e.g. Foufoula-Georgiou & Kumar 1994; Klees & Haagmans 2000; Freeden & Michel 2004b; Oliver 2009), the use of wavelets is still no matter of routine in the geosciences, beyond applications in one and two Cartesian dimensions. This despite, or perhaps because, there being a wealth of available constructions relevant for global geophysics, in other words: on the sphere (e.g. Schröder & Sweldens 1995; Narcowich & Ward 1996; Antoine et al. 2002; Holschneider et al. 2003; Freeden & Michel 2004a; Fernández & Prestin 2006; Hemmat et al. 2005; Schmidt et al. 2006; Starck et al. 2006; McEwen et al. 2007; Wiaux et al. 2007; Lessig & Fiume 2008; Bauer & Gutting 2011), if not on the ball. Indeed, inasmuch as they involve the analysis of cosmological data, satellite observations or computer-generated images, the above studies are mostly concerned with surfaces, not volumes.

In seismology, Chiao & Kuo (2001) were, to our knowledge, the first to develop a "biorthogonal-Haar" wavelet lifting scheme (Schröder & Sweldens 1995) for a triangular surface tesselation of the sphere suitable for multiscale global tomography. Later, these same authors formed a (biorthogonal) spline basis for a Cartesian cube useful in exploration geophysics (Chiao & Liang 2003) and for regional studies (Hung et al. 2010). Finally, Chevrot & Zhao (2007) constructed a three-dimensional (orthogonal) Haar basis on an equidistant geographical grid which was also used for a regional inversion. To this date, a truly three-dimensional wavelet basis on the ball with practical utility in the geosciences has been lacking.

Whatever the role that wavelets will play in it, the future of global seismic tomography will involve massive amounts of heterogeneous data spanning a range of resolutions, from travel times reported by global networks to waveforms of portable deployments, with strong regional concentrations of station coverage in areas such as Japan, the United States, and Europe, supplemented with sparse networks in less densely populated or oceanic regions. It is also clear that finite-frequency kernels, which allow for the correct volumetric sensitivity-based weighting of the measurements in distinct frequency bands, are here to stay, whichever the various ways in which they are calculated (see Nolet 2008, and references therein). Accounting for finite-frequency sensitivity requires an effective overparameterization if one wishes to exploit the extra resolution offered by the spatial variations in their sensitivity, something for which wavelet seem ideally suited also.

This paper documents the extensive prospective work that we have done in preparation for realistic wavelet-based global seismic inversions. Our goal remains to ensure that there exist performant and efficiently calculable, flexible wavelet methods on the three-dimensional ball, to fulfill the promise of multiresolution analysis (Mallat 1989; Jawerth & Sweldens 1994) in global seismology. Not just for the representation and analysis of seismic models after the fact, but rather for their determination, as an integral part of a parsimonious parameterization of the inverse problem — of the sensitivity matrix, of the model space, or both. Although there is no objective guarantee that Nature, or the interior of the Earth in particular, is parsimonious in character, sparsity is worth striving for. By simplifying a tomographic image to contain a relatively small number of recognizable objects we facilitate interpretation (Sambridge et al. 2006). Moreover, such models can be more accurate than their data (Gauch 2003), a point not to be overlooked in view of the large relative errors of seismic delay times and amplitudes.

By *flexibility* we mean the ability to substitute a particular wavelet design for another in any of the three coordinate directions; by *efficiency* we intend to avoid the tedious case-by-case derivation of different bases and calculation methods. By *performance* we target the ability to capture the unknown model by explaining the data (in an ℓ_2 sense) with a minimum of wavelet and scaling coefficients, both where the data require the solution to be smooth and where they necessitate the presence of sharp contrasts. It is of course in this capacity also (e.g. Donoho & Johnstone 1994, 1995) that wavelets will distinguish themselves from many other traditional methods of seismic inversion (except, perhaps, Bayesian partition modeling and related methods, see Denison et al. 2002; Bodin et al. 2009; Bodin & Sambridge 2009). As to *sparsity*, it is both numerically and philosophically attractive (Constable et al. 1987) and physically plausible or at least testable that the interior of the Earth should be sparse when expressed in a wavelet basis. Fortunately, for most large underdetermined systems of linear equations the minimal ℓ_1 norm solution is also the sparsest (Candès et al. 2006; Donoho 2006), for which (fast, iterative) algorithms exist (Daubechies et al. 2004; Loris 2009). Elsewhere, Loris et al. (2007, 2010) and Gholami & Siahkoohi (2010) explored the suitability of sparsity-seeking thresholded wavelet-based inversion approaches in two-dimensional (2D) and three-dimensional (3D) Cartesian settings relevant to seismic tomography. All of the above issues will again be the guiding principles behind the new spherical wavelet construction(s) that we present in this paper.

This paper is organized as follows. In Section 2 we develop a first class of wavelet constructions on the sphere via a well-known Cartesian-to-spherical mapping known as the "cubed sphere" (Ronchi et al. 1996; Komatitsch & Tromp 2002). As this surface tesselation has "seams" separating each of six subdivisions or "chunks", we acknowledge these boundaries in the construction by using so-called "wavelets on the interval". These revert to the classical compactly supported (bi)orthogonal Cartesian constructions of Daubechies (1988) and Cohen et al. (1992) in the interior domains but receive special consideration on the edges as put forth by Cohen et al. (1993). In Section 3 we study the sparsity of two global seismic tomographic Earth models by thresholded reconstructions of their wavelet transforms applied to the angular coordinates of the cubed sphere, at constant depth intervals, and considering a variety of goodness-of-fit criteria. We furthermore characterize, in Section 4, the scale lengths of heterogeneity in these models by reporting the absolute and relative contributions of their wavelet and scaling coefficients in the expansion as a function of depth and location in the Earth. We calculate the correlation of both models as a function of scale and position in the Earth, and where this appears sensible, present estimates for the variable $\delta \ln V_S / \delta \ln V_P$ ratios that can be derived from it. Sections 3 and 4 are relatively self-contained and discuss aspects of geophysical interest that are not properly part of the mathematical treatment of the tomographic inverse problem. Readers with an interest limited to the latter will appreciate the motivation for our approach from these sections but may wish to skip them for a first reading. In Section 5 we review the main approach to obtain sparse wavelet-based solutions to the *inverse* problem of seismic tomography, which were previously discussed in a Cartesian framework by Loris et al. (2007, 2010). As using the initial construction with such schemes led to undesirable artifacts at the edges between the chunks, we derive a second wavelet construction in Section 6, which appears to be free of such artifacts, as we show using realistic synthetic tests in Section 7. As we envisage it, the first wavelet construction remains the tool of choise for the analysis of seismic Earth models, whereas the second construction is an adaptation that should be used in the inversion for such models from primary data. While in this paper we focus on the angular part of the cubed *sphere* we generalize our construction to including the case of the *ball* and provide an outlook for further research in global seismic tomography in the concluding Section 8.

2 A FIRST CONSTRUCTION

Simple latitude-longitude grid coverings of the sphere are very inhomogeneous and suffer from meridian convergence and singularities at the poles (e.g. Swinbank & Purser 2006; González 2010; Lauritzen et al. 2010). For this reason no wavelet transform is ever expected to be naturally formulated in such traditional coordinates. As Ronchi et al. (1996), we define the coordinate quartet (ξ, η, r, κ) for each of the $\kappa = 1 \rightarrow 6$ chunks. The angular coordinates $-\pi/4 \leq \xi, \eta \leq \pi/4$ and the radial coordinate r are mapped to the usual Cartesian triplet (x, y, z) using the transformation

$$(x,y,z) = \begin{cases} r\,(\tan\eta, -1, -\tan\xi)/s & \text{if} \quad \kappa = 1, \\ r\,(-1, -\tan\xi, \tan\eta)/s & \text{if} \quad \kappa = 2, \\ r\,(\tan\eta, -\tan\xi, 1)/s & \text{if} \quad \kappa = 3, \\ r\,(-\tan\xi, \tan\eta, -1)/s & \text{if} \quad \kappa = 4, \\ r\,(1, \tan\eta, -\tan\xi)/s & \text{if} \quad \kappa = 5, \\ r\,(-\tan\xi, 1, \tan\eta)/s & \text{if} \quad \kappa = 6, \end{cases} \tag{1}$$

whereby $s = \sqrt{1 + \tan^2\xi + \tan^2\eta}$. The inverse mapping is obtained, for $t = \max(|x|, |y|, |z|)$,

$$(\xi,\eta,\kappa) = \begin{cases} [\,\mathrm{atan}(z/y), \mathrm{atan}(-x/y), 1\,] & \text{if} \quad t = -y, \\ [\,\mathrm{atan}(y/x), \mathrm{atan}(-z/x), 2\,] & \text{if} \quad t = -x, \\ [\,\mathrm{atan}(-y/z), \mathrm{atan}(x/z), 3\,] & \text{if} \quad t = z, \\ [\,\mathrm{atan}(x/z), \mathrm{atan}(-y/z), 4\,] & \text{if} \quad t = -z, \\ [\,\mathrm{atan}(-z/x), \mathrm{atan}(y/x), 5\,] & \text{if} \quad t = x, \\ [\,\mathrm{atan}(-x/y), \mathrm{atan}(z/y), 6\,] & \text{if} \quad t = y, \end{cases} \tag{2}$$

whereby $r = \sqrt{x^2 + y^2 + z^2}$. This parametrization is non-smooth across the edges separating the chunks. The above formulas correspond to the drawing in Fig. 1, where only one of the chunk faces is gridded to reveal the angular coordinate lines (ξ, η) at a resolution that divides this face into $2^4 \times 2^4$ distinct surface elements. Throughout this paper we will quote N as the angular resolution level of our cubed sphere, which implies that it has 6×2^{2N} such elements, with typical tomography grids having $N = 7$.

In principle there are many possibilities to choose the surficial coordinates (ξ, η) in each chunk. We picked ours so as to minimize the splitting of continents over more than one chunk. Our choice differs from the canonical version of Ronchi et al. (1996) by a rigid rotation of the coordinate system, as can be seen by comparing our Fig. 2 with their Figs 15–16. The Euler angles used in our construction are $\alpha = 0.0339$, $\beta = 1.1705$, and $\gamma = 1.1909$, respectively. It is important to note that within a chunk ξ and η are *not* spherical coordinates; a shift in ξ (with η fixed) or in η (with ξ fixed) does *not* correspond to a rotation on the sphere. This is apparent from the pinching of coordinate lines in Fig. 1.

Armed with the coordinate conversions of eqs (1) and (2) we are able to regard the problem of designing a wavelet transform for the sphere as simply requiring the selection of a certain Cartesian wavelet transform which is mapped to and from the sphere. Such an approach is philosophically related to those involving stereographic projection (Antoine & Vandergheynst 1999; Antoine et al. 2002; Wiaux et al. 2005), though the fundamental domain of our transform remains a single chunk. Within each such chunk, the surface Jacobian of our mapping is given by the smoothly varying

$$\mathcal{J}(\xi,\eta) = (1 + \tan^2\xi)(1 + \tan^2\eta)/s^3, \quad \sqrt{2}/2 \leq \mathcal{J} \leq 1. \tag{3}$$

For each of the chunks then the area is given by

$$\int_{-\pi/4}^{\pi/4} \int_{-\pi/4}^{\pi/4} \mathcal{J}(\xi,\eta)\, d\xi\, d\eta = \frac{4\pi}{6}. \tag{4}$$

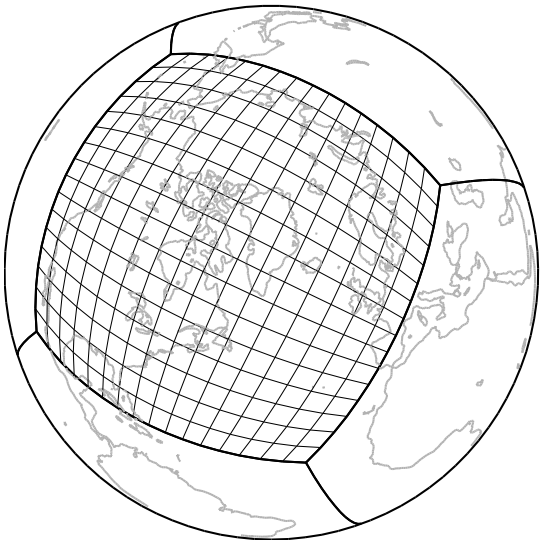

Figure 1. Aerial view showing our first adaptation of the cubed sphere of Ronchi et al. (1996). Of the front-facing four of the in total six "chunks", one is gridded to reveal its 2^{2N} distinct surface elements ($N = 4$).

Without this being a uniform mapping, one of the main advantages of the chosen coordinate system is thus that the meshes defined on each region span the surface of the sphere with an almost constant spatial resolution, as noted by Ronchi et al. (1996).

Ignoring any and all such distortions we are able to unlock the power of popular Cartesian wavelet constructions, of which we choose the two best known: the orthogonal construction of Daubechies (1988) and the biorthogonal construction of Cohen et al. (1992). Both of these lead to compactly supported wavelets and scaling functions, though only the biorthogonal ones can be (anti)symmetric (except for Haar). Examples of scaling functions and wavelets at scales of decreasing dominant wavelength are shown in Fig. 3 for the four-tap Daubechies basis (D4) and in Fig. 4 for the Cohen-Daubechies-Feauveau family with four and two vanishing moments (CDF 4–2) in analysis and synthesis, respectively. To get a sense of the physical size of wavelets and scaling functions at a certain scale, which depends on the original cubed-sphere resolution parameter N, we calculate reference circles of various angular radii Δ, and quote their values in degrees above each panel.

The literature on Cartesian wavelet analysis is vast, and it is not our intention to repeat any of it here. Most useful for the practicing geophysicist will perhaps be the treatises by Mallat (2008) and Strang & Nguyen (1997); texts focused on algorithms are Press et al. (1992) and, in particular, Jensen & la Cour-Harbo (2001). All of the computer code required to reproduce the figures and conduct the analyses presented in this paper is molded after these general references and will be available from the authors.

Two aspects of wavelet analysis bear specific mentioning here. The first intricacy is how we treat the seams between the chunks. In agreement with Cohen et al. (1993) the argument is easily made that neither ignoring the seams nor periodization or reflection are viable options, as each of these leads to artifacts in the representation. We thus follow their suggestion to the letter and construct a multiresolution basis requiring 2^{2N} wavelet and scaling coefficients for each of the chunk faces having 2^{2N} surface elements. For this we switch to special boundary filters at each of the edges, and apply preconditioners to the data prior to transformation in order to guarantee the usual polynomial cancellation throughout the

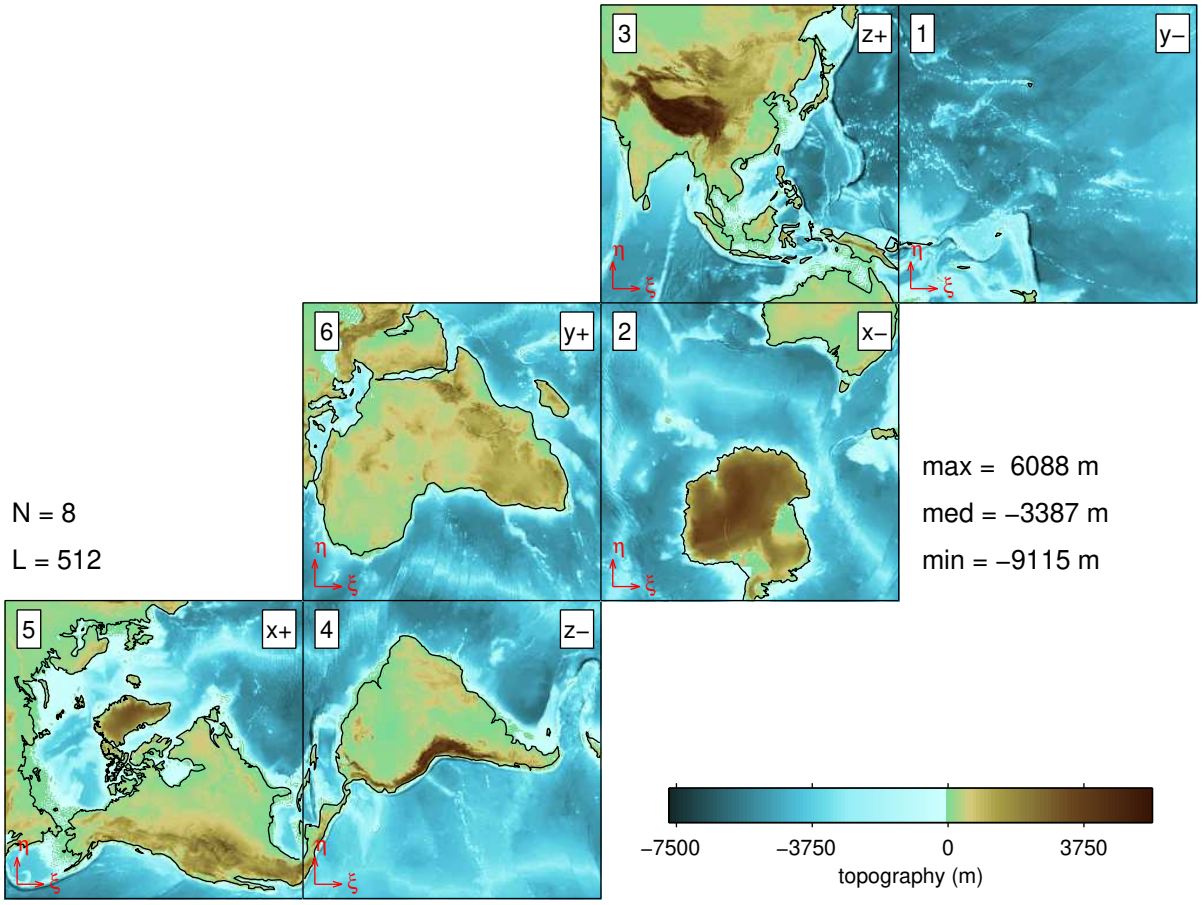

Figure 2. Geometry, nomenclature, and numbering of the six faces of our first adaptation of the cubed sphere of Ronchi et al. (1996) in a two-dimensional "unfolded" view. Rendered is the Earth's topography from the model ETOPO5, courtesy of NOAA's National Geophysical Data Center. The projection was obtained by spherical-harmonic expansion of the coefficients from this model (Georg Wenzel, *pers. comm.*) truncated at degree and order $L = 2^{N+1}$, evaluated at the 6×2^{2N} cubed-sphere grid points ξ, η, for $N = 8$. Minimum, median, and maximum values in this approximation are shown in the legend.

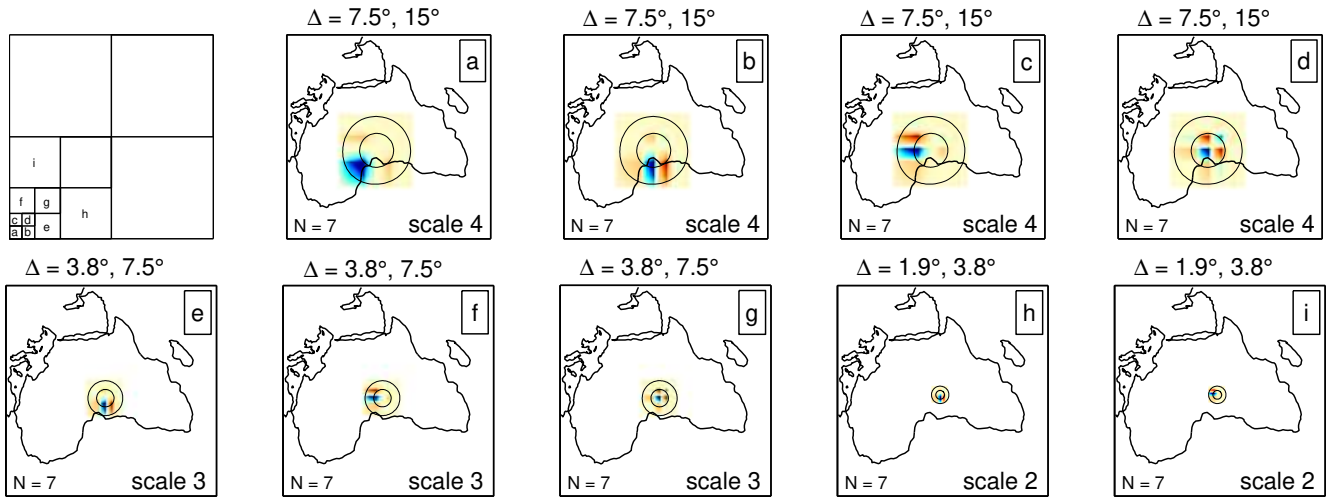

Figure 3. Wavelet and scaling functions of the D4 construction in the angular coordinates of the $N = 7$ cubed sphere, at various scales. Continental outlines and circles of varying angular radii Δ are plotted to make reference to physical spatial scales. The positions of the coefficients belonging to the functions in the lettered panels are shown in the diagram in the top left. The scaling function (*a*), which is averaging in nature, captures what remains to be explained after the breakdown into wavelets down to scale 4 is complete. Each of the wavelets, which pick up detailed, derivative, structure, is sensitive in a particular direction: to ξ in (*b*), to η in (*c*), or diagonally in (*d*). In the interior domain, away from the edges where boundary functions (not shown) live, the patterns repeat exactly, with the footprint at each successive scale half that of the preceding scale. The diagonally sensitive wavelet at scale 2 is not shown. Every function shown is orthonormal in (ξ, η) and their inner products with respect to every other one vanish.

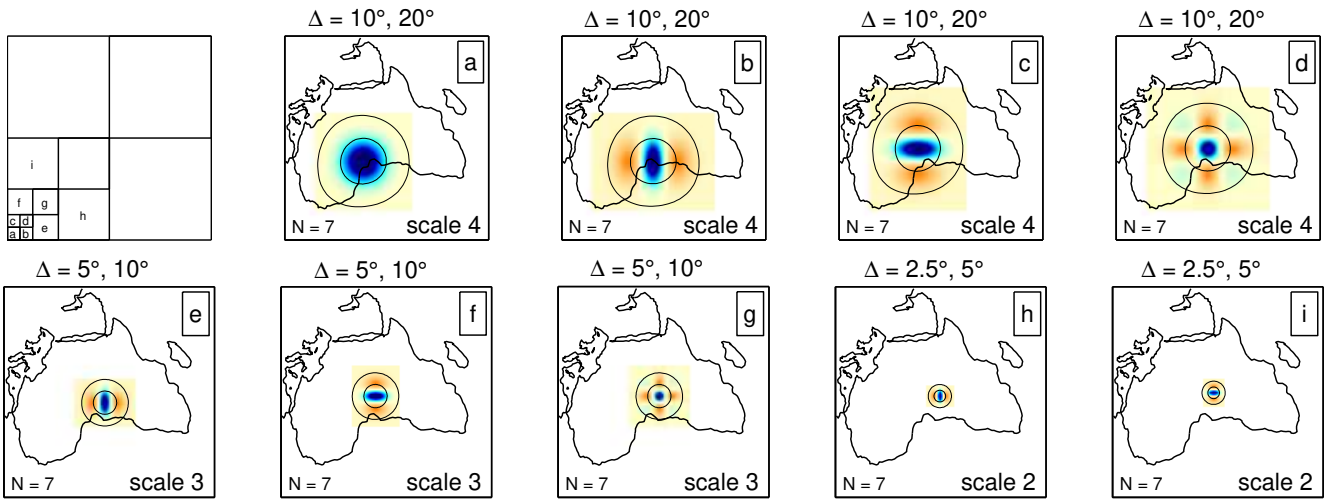

Figure 4. Wavelet and scaling functions arising from the CDF 4–2 construction in the angular coordinates of the cubed sphere, with their scale levels indicated. The layout is identical to that of Fig. 3. As opposed to the D4 wavelets, the CDF 4–2 construction is biorthogonal, which renders every shown synthesis function orthogonal in (ξ, η) to its dual, which is used for analysis; none of the dual functions are shown. Unlike the D4 functions the CDF 4–2 have mirror symmetry.

closed rectangular interval. The acknowledgment of the edges in this way is the hallmark of the wavelet construction in this section here (which we call the First Construction). This is as easily done for the orthonormal as for the biorthogonal constructions, though we have limited the implementation and illustration of this procedure, in Fig. 5, to the compactly supported two-tap (Haar), four-tap, and six-tap orthonormal families (D2, D4 and D6).

Before we discuss Fig. 5 in any more detail we should introduce the second important feature that renders wavelet transforms in general useful for the analysis and representation of (geophysical) data. This second topic is the idea of *thresholding*, or shrinkage. In many applications the wavelet transformation amounts to a projection under which many of the expansion coefficients are very small: so small that we might as well throw them away; the resulting reconstruction will still be close to the original (Donoho & Johnstone 1994). Intuitively, the "best" wavelet basis that we can select to represent our data is the one that yields the most near-zero coefficients. When these are *replaced* by zeroes prior to reconstruction, as under the definition of *hard* thresholding (Mallat 2008), we obtain highly compressed versions of the data at hand, with only negligible degradation.

Fig. 5 explores the effects of thresholding, coefficient statistics, and reconstruction errors for a model of terrestrial topography, a general proxy for the length scales of heterogeneities to be found not only at the surface, but also in the interior of the Earth. We focus on the sixth, or "African" chunk of our cubed sphere, and use the D2, D4 and D6 wavelet bases (on the interval, with preconditioning). The top row uses the (common) conventions introduced in Fig. 3 in plotting the wavelet and scaling coefficients in each of the basis after (hard) thresholding them such that only the coefficients larger than their value at the 85^{th} percentile level survive. The coefficients that have now effectively been zeroed out are left white in these top three panels. The middle series of panels of Fig. 5 plots the spatial reconstruction after thresholding at this level; the root mean squared (rms) error of these reconstructions are quoted as a percentage of the original root mean squared signal strength. The thresholded wavelet transforms allow us to discard, as in these examples, 85% of the numbers required to make a map of African topography in the cubed-sphere pixel basis: the percentage error committed is only 5.8%, 4.9% and 6.7% according to this energy criterion in the D2, D4 and D6 bases, respectively. From the map views it is clear that despite the relatively small error, the D2 basis leads to unsightly block artifacts in the reconstruction, which are largely avoided in the smoother and more oscillatory D4 and D6 bases. A view of the coefficient statistics is presented in the lowermost three panels of Fig. 5. The coefficients are roughly log-normally distributed, which helps explain the success of the thresholded reconstruction approach. While the example here was strictly designed to illustrate our algorithms and procedures, we conclude that the D4 basis is a good candidate for geophysical data representation, provided the edges between cubed-sphere chunks have properly been accounted for.

3 EARTH MODEL SPARSITY

In tomographic studies, either as an integral part of the inversion or after a solution has been found, the target model is parameterized by local or global basis functions (Nolet 2008). Blocks, cells, nodes, or voxels (e.g., Aki et al. 1977; Zhang & Tanimoto 1993; Spakman & Bijwaard 2001; Simons et al. 2002; Debayle & Sambridge 2004; Nolet & Montelli 2005) are all strictly local functions. Cubic B-splines (e.g., Wang & Dahlen 1995; Wang et al. 1998; Boschi et al. 2004) or wavelets (e.g., Chiao & Kuo 2001; Chevrot & Zhao 2007; Loris et al. 2007) are more generally localized functions. Spherical harmonics (e.g., Dziewoński 1984; Woodhouse & Dziewoński 1984; Ekström et al. 1997; Trampert & Woodhouse 1996, 2001) are ideally localized spectrally but have global support (Freeden & Michel 1999). An intermediate approach that combines spatial and spectral localization was developed using spherical harmonic splines by Amirbekyan & Michel (2008) and Amirbekyan et al. (2008), but this produces an inverse problem that scales with the square of the number of data collected, rendering it impractical for the large-scale tomographic systems of the future.

In preparing for the study of the suitability for solution of such massive inverse problems of the wavelet transforms that we introduced in the previous section, we take a detour in this section by addressing the question: is the Earth sparse in a wavelet basis? Of

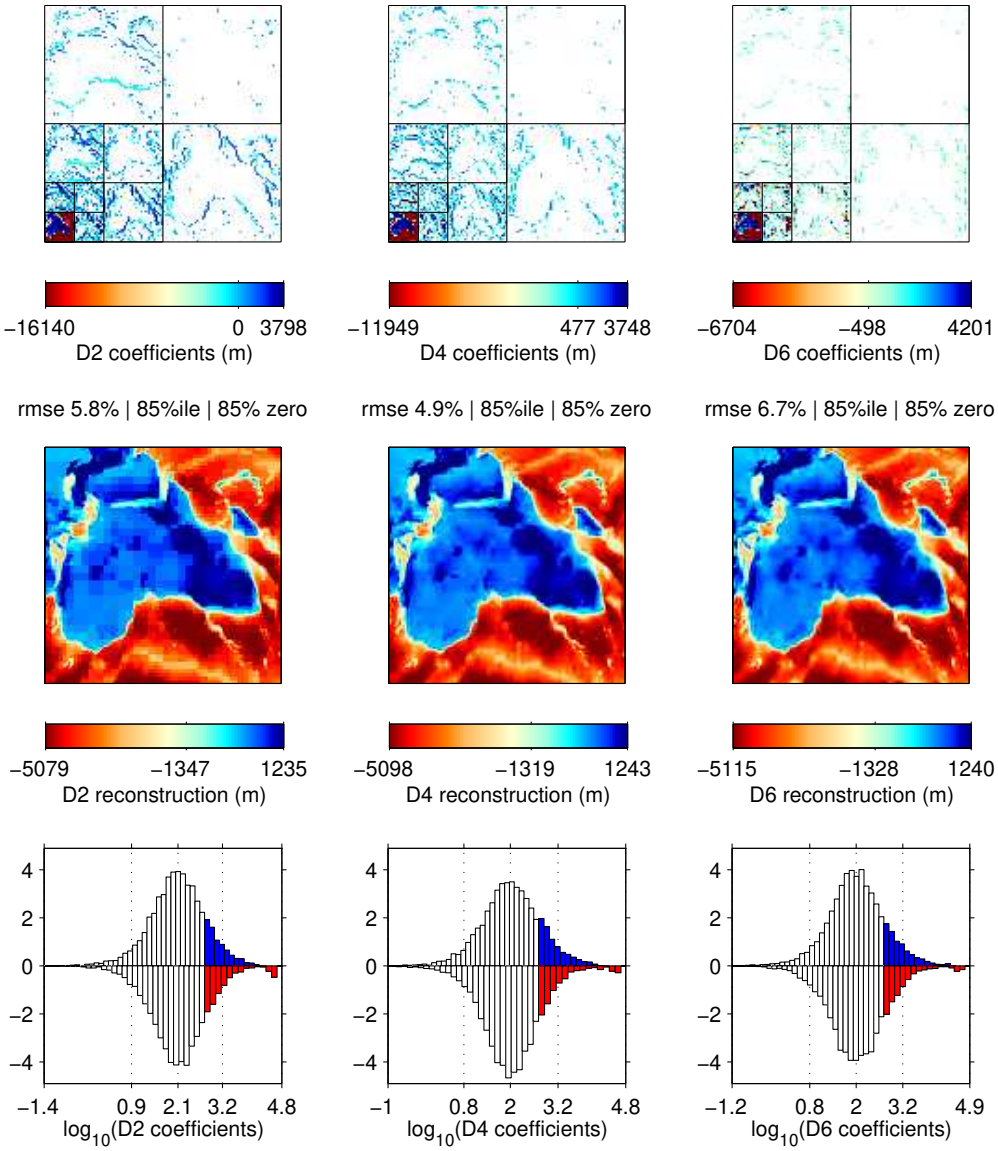

Figure 5. Wavelet and scaling coefficients (*top*), space-domain reconstructions after thresholding (*middle*), and "signed" histograms (*bottom*) of the wavelet and scaling coefficients of the "African" (sixth) face of the cubed-sphere version of the Earth's topography first shown in Fig. 2. We have used the preconditioned interval wavelet transforms on the faces of the cubed sphere, as described in the text. All coefficients were hard-thresholded at the 85^{th} percentile level, retaining only the 15% largest coefficients by absolute value. In the top row, the locations of zeroed coefficients are rendered white; those are also captured by the white bars in the histograms. The root mean squared (rms) error of the reconstruction after thresholding is indicated as a percentage of the signal rms. Tick marks on the color bars identify the 5^{th}, 50^{th} and 95^{th} percentile of the coefficients or the spatial reconstructions after thresholding, respectively. Interior ticks on the histograms roughly coincide with these same percentiles as applied to either the positive and negative coefficients when expressed on a logarithmic scale. Histograms for the positive coefficients point up and have ordinates in positive percentages, histograms for the negative coefficients point down and have ordinates in negative percentages; these percentages are with respect to the total number of positive and negative coefficients. The blue and red shaded areas of the histograms reflect the coefficients retained at the global 85^{th} thresholding level.

course we will never be able to answer this question with any degree of certainty, but we can investigate, at the very least, whether Earth *models* are sparse in such bases. Because they are, as we shall see, we will gain from parameterizing the inversion for future Earth models using the spherical wavelets developed in this paper. The expected gains are with respect to numerical efficiency but also in terms of regularization. Since wavelets are not global functions (ours, as can be seen from Figs. 3 and 4, are compactly supported, i.e. vanishing outside their scale-dependent footprint), and yet, (bi)orthogonal, the function basis will not dictate the model structure in areas of poor data coverage as is the case with spherical harmonics (Trampert & Snieder 1996; Boschi & Dziewoński 1999; Amirbekyan et al. 2008). Moreover, though this depends on precisely what wavelet construction is being used, they are capable of representing both smoothly varying functions as well as preserving sharp edges, and their natural multi-resolution nesting will allow for the model resolution to vary spatially, as required by the data.

There is, however, another reason to find out how seismic Earth models behave under wavelet transformation: because it enables us to study the relative importance of model heterogeneity at

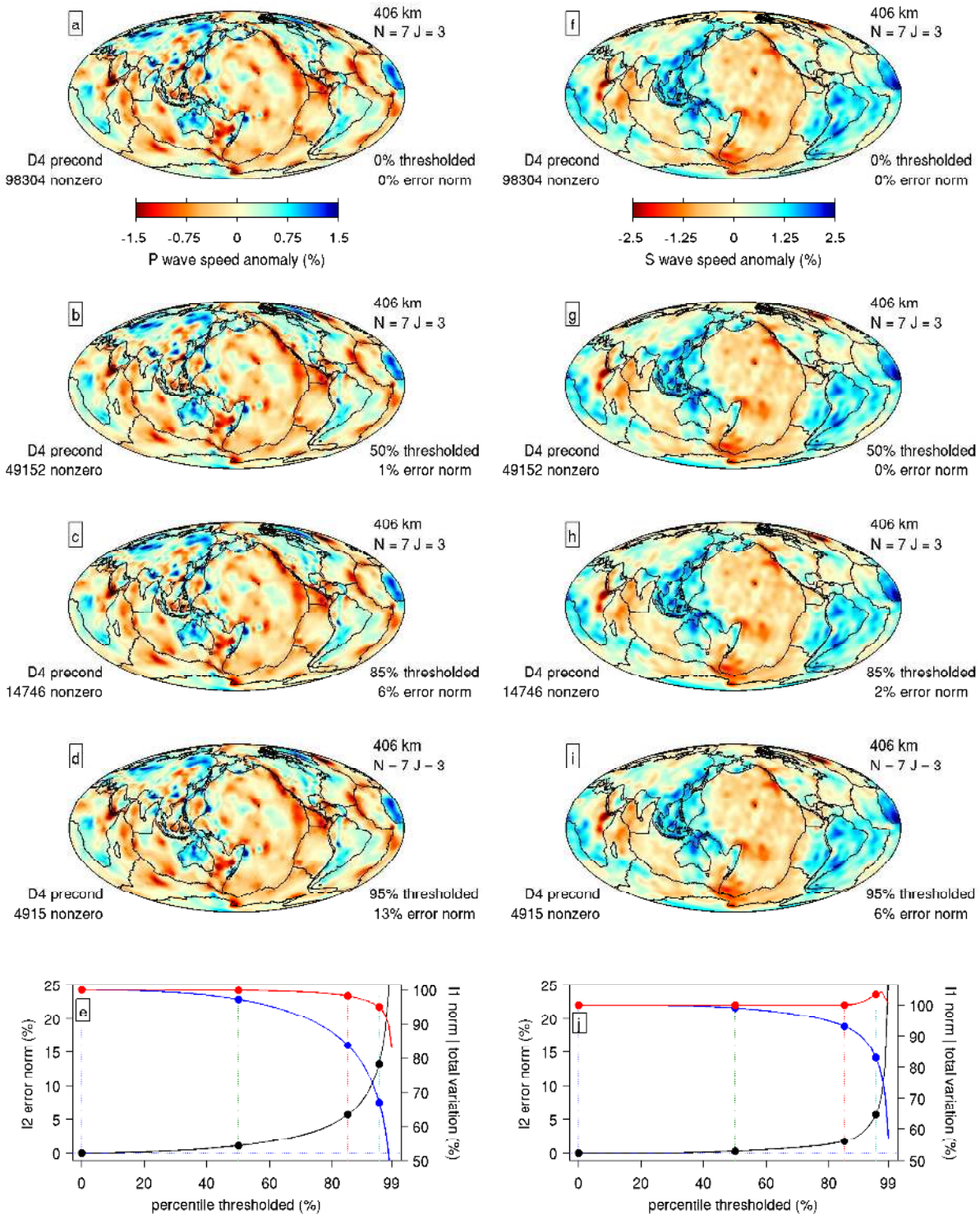

Figure 6. Sparsity and reconstruction stability of two global seismic wavespeed models under incremental hard thresholding of their wavelet and scaling coefficients using the preconditioned edge-cognizant D4 wavelet basis (Daubechies 1988; Cohen et al. 1993) in the angular coordinates of the cubed sphere, as developed in this paper. *(a–e)* Results for the *P*-wave seismic model of Montelli et al. (2006) and *(f–j)* for the *S*-wave seismic model of Ritsema et al. (2010), at the same depth of 406 km below the surface of the Earth, for cubed spheres with 6×2^{2N} elements ($N = 7$), and to a 2^J dyadic subdivision ($J = 3$). As a function of the percentage of the coefficients that are being thresholded, and relatively to the original unthresholded values, the bottom panels quote the spatial ℓ_2 norms of the reconstruction error (in black), the total variation norms of the reconstructed images in the space domain (in red), and the ℓ_1 norms of the coefficients that remain (in blue). The values obtained for the cases shown in map view are shown as filled circles on these graphs, and the corresponding metrics in the D2, D4 and D6 bases are tabulated in Table 1. The reconstructions remain faithful to the originals even at elevated levels of thresholding.

depth (km)	thresholding percentile	Montelli et al. (2006) relative ℓ_2 error norm (%) D2	D4	D6	Ritsema et al. (2010) relative ℓ_2 error norm (%) D2	D4	D6
203	50	1.816	0.808	3.025	1.014	0.236	0.229
	85	9.212	4.653	6.324	5.028	1.360	0.722
	95	18.721	11.214	11.456	10.073	4.351	3.172
406	50	2.294	1.107	3.983	1.267	0.311	0.297
	85	10.559	5.757	7.701	6.182	1.786	0.968
	95	20.689	13.231	13.481	12.393	5.717	4.125
609	50	2.661	1.244	3.499	1.562	0.397	0.393
	85	11.471	6.419	7.775	7.428	2.211	1.230
	95	21.622	14.145	14.384	14.589	7.121	5.162
1015	50	3.099	1.311	3.884	2.083	0.533	0.531
	85	12.727	6.896	8.440	9.517	2.775	1.592
	95	23.296	15.107	15.139	18.621	9.009	6.462
2009	50	1.995	0.461	1.799	1.582	0.379	0.372
	85	8.890	3.662	5.174	7.363	2.021	1.145
	95	16.946	9.208	9.104	14.527	6.572	4.695

Table 1. A companion to Fig. 6, this table lists the ℓ_2 error norms, relative to the original, of the reconstructions of the *P*-wave speed model of Montelli et al. (2006) and the *S*-wave model of Ritsema et al. (2010) under hard wavelet thresholding in the angular coordinates. See Fig. 6 and text for more details.

different scale lengths, which is important to help constrain geochemical and geodynamical models and interpretations of Earth structure. The Earth is heterogeneous at all scales but not likely everywhere to the same degree; thermally induced deviations from the radial average one-dimensional Earth structure are expected to be smoother and with longer wavelengths than those due to compositional variations; the presence of distinct scatterers further complicates this picture (Shearer & Earle 2004). In short, we are interested in obtaining a power spectral density of sorts (Chevrot et al. 1998a,b; Boschi & Dziewoński 1999), as applied to seismic structure and how it may vary spatially within the Earth. As we are not in the position to return to direct measurements of the energy distribution of heterogeneity (Hedlin & Shearer 2000; Margerin & Nolet 2003; Becker et al. 2007; Garcia et al. 2009) we will instead study the sizes and scales within reported tomographic Earth models.

From the plethora of seismic Earth models that are available to study, we select two mantle models: one by Montelli et al. (2006) of compressional (*P*) wavespeed heterogeneity and another by Ritsema et al. (2010) of shear (*S*) wavespeed perturbations. Neither model has much at all in common with the other in terms of its construction, and from the point of view of parameterization, Montelli's model has a tetrahedral grid underlying it, whereas Ritsema's expands wavespeed anomalies in a spherical harmonic basis complete to degree and order 40. At a depth of about 400 km, Figs 6a and 6f show *P*-wave (Montelli et al. 2006) and *S*-wave (Ritsema et al. 2010) anomalies from the average at that depth. Montelli's model was *interpolated* (from the tetrahedral grid on which it was built) onto the 6×2^N ($N = 7$) points of our cubed sphere, whereas Ritsema's model was *evaluated* (from the listed spherical harmonic and radial spline expansion coefficients) at these same points. Subsequently, the wavelet transform in the D4 basis (with special boundary filters and after preconditioning, and up until scale $J = 3$) was thresholded and the results re-expanded to the spatial grid, identically as we did for the topography in Fig. 5. The results for specific values of the thresholding (quoted as the percentile of the original wavelet coefficients) are shown in Figs 6a and 6g for the $50^{\rm th}$, Figs 6c and 6h for the $85^{\rm th}$, Figs 6d and 6i for the $95^{\rm th}$ percentile, respectively. At each level of thresholding the number of nonzero wavelet/scaling expansion coefficients is quoted: at 0% thresholding this number is identical to the number of pixels in the surficial cubed sphere being plotted.

As we have written before, the wavelet transformation does not change the *number* of pieces of information with which it is presented. Rather, it dramatically redistributes information in a manner that allows us to simply omit those coefficients with low values, with limited degradation to the field being represented. This reconstruction "error" can be visually assessed from the pictures; it is also quoted next to each panel as the percentage of the root mean squared error between the original and the reconstruction, normalized by the root mean squared value of the original in the original pixel representation, in percent. Specifically, we calculate and quote the ratio of ℓ_2 norms in the pixel-basis model vector $\mathbf{m}$,

$$100 \times \left\| \mathbf{m} - S\left\{\mathcal{T}[A(\mathbf{m})]\right\} \right\|_2 \Big/ \left\| \mathbf{m} \right\|_2 , \tag{5}$$

which, in the lower-right annotations is called the "% error norm". We have written A for any of the wavelet (analysis) transforms that are used and S (synthesis) for their inverses, and $\mathcal{T}$ for the "hard" thresholding (Mallat 2008) of the wavelet and scaling coefficients.

In Figs 6e and 6j, this same misfit quantity (5) is represented as a black line relevant to the left ordinate labeled "ℓ_2 error norm", which shows its behavior at 1% intervals of thresholding; the filled black circles correspond to the special cases shown in the map view. Only after about 80% of the coefficients have been thresholded does the error rise above single-digit percentage levels, but after that, the degradation is swift and inexorable. The blue curves in Figs 6e and 6j show another measure of misfit relevant in this context, namely the ratio of the ℓ_1 norms of the thresholded wavelet coefficients compared to the original ones, in percent, or

$$100 \times \left\| \mathcal{T}[A(\mathbf{m})] \right\|_1 \Big/ \left\| A(\mathbf{m}) \right\|_1 . \tag{6}$$

As we can see from the figure the ℓ_2 ratios (5) in the black curves (and the left ordinate) evolve roughly symmetrically to the ℓ_1 ratios (6) in the blue curves (and the right ordinate), though evidently their range is different.

Finally, a third measure that is being plotted as the red curve is the "total variation" norm ratio, in percent, namely

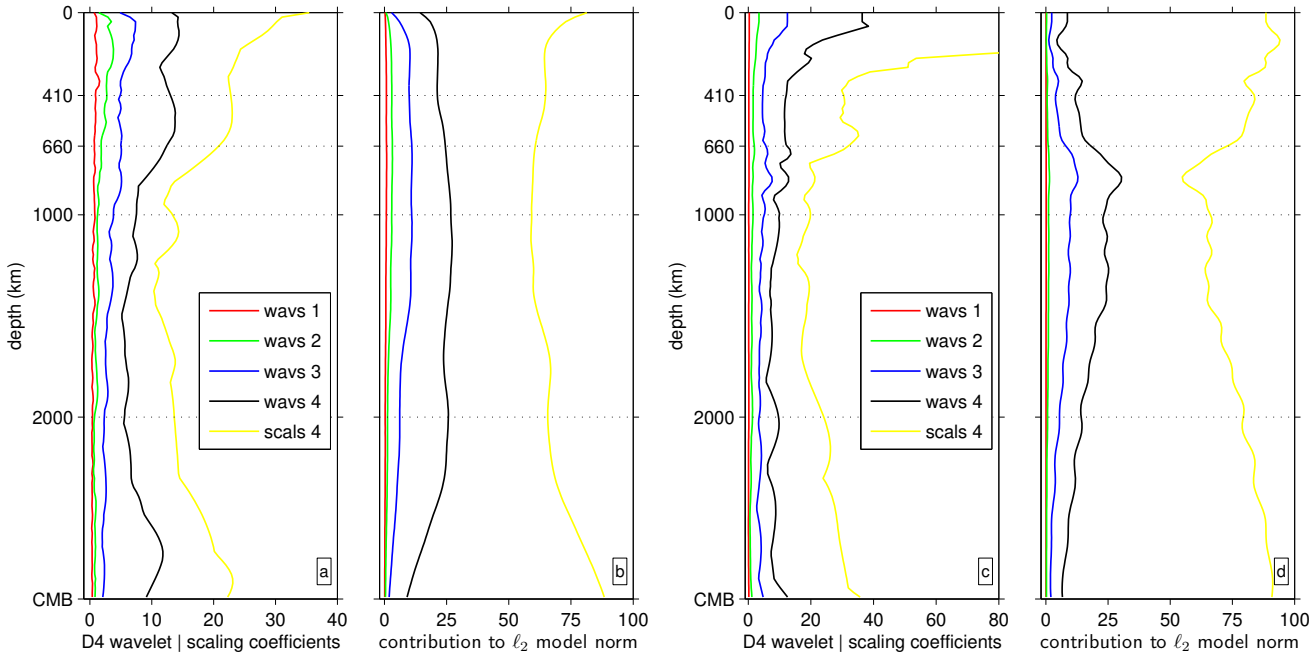

Figure 7. Scale lengths of seismic heterogeneity as a function of depth in the Earth, to the core-mantle boundary (CMB), obtained from the complete angular expansion in the D4 wavelet basis of (*a–b*) the *P*-wave speed model of Montelli et al. (2006) and (*c–d*) the *S*-wave speed model of Ritsema et al. (2010). See Fig. 3 for the wavelet and scaling functions and Fig. 6 for the seismic models: all calculations are with reference to cubed spheres with 6×2^{2N} elements ($N = 7$), and to a 2^J dyadic subdivision ($J = 4$). Panels *a* and *c* show the maximum absolute values of the wavelet or scaling coefficients (*wavs* and *scals* in the legend, respectively) at the scales quoted, differentiated by color. The scaling coefficients at the fourth scale have the largest values: at all depths the maximum at this scale and the overall maximum (not shown) coincide. Panels *b* and *d* show the proportion (in %) of the contribution to the overall ℓ_2 norm of the seismic models at every depth by the ensemble of the coefficients at each of the scales. Ritsema's model has much more structure in the top 410 km of the Earth (not shown because of the axis truncation is a peak with a value of 137.2 centered at 135 km) compared to the bottom 1000 km, as opposed to Montelli's model which has a more uniform distribution of heterogeneity. Both models are characterized by minima of seismic structure at mid-mantle depths.

$$100 \times \|\boldsymbol{\nabla}\, S\,\{\mathcal{T}[A(\mathbf{m})]\}\|_1 \big/ \|\boldsymbol{\nabla}\mathbf{m}\|_1\,, \qquad (7)$$

whereby $\|\boldsymbol{\nabla}\mathbf{m}\|_1$ is the sum over all voxels of the length of the local gradient of $\mathbf{m}$. By this measure, which is popular in image restoration applications (Rudin et al. 1992; Dobson & Santosa 1996; Chambolle & Lions 1997), the quality of the reconstruction stays very high even at very elevated levels of thresholding; we note that its behavior is not monotonic and may exceed 100%.

As with terrestrial topography in Fig. 5 we conducted all of the experiments on the seismic models that are presented in Fig. 6 in the D2, D4 and D6 wavelet bases. A summary of the ℓ_2 error norm ratios as a function of thresholding levels for each of those bases is presented in Table 1. On the strength of its behavior under the criteria (5)–(7) and upon visual inspection of the results, we conclude that the D4 basis remains a very appropriate choice for the efficient representation of seismic models. To this choice we adhere in the geophysically motivated study of mantle structure in those same models which follows below.

4 TOMOGRAPHIC MODEL STRUCTURE

There is much geophysical interest in tying seismic observations of mantle structure to models incorporating geodynamic modeling and mineral physics observations (e.g. Jordan et al. 1993; Kárason & van der Hilst 2000; Becker & Boschi 2002; Bull et al. 2009). Our study is an attempt to provide a flexible, quantitative, multiresolution framework for such analyses that may add to the more traditional power-spectral (e.g. Becker & Boschi 2002; Houser & Williams 2009; Schuberth et al. 2009) and statistical analyses (e.g. Hernlund & Houser 2008). In obliterating the phase of the anomalies, the former line of inquiry largely loses the relative spatial location of seismic structure, while the latter type of study is no longer sensitive to its scale and wavelength dependence. While in this paper we do not explicitly study the radial correlation of mantle structure (Puster et al. 1995; van der Hilst & Kárason 1999), the analysis below readily lends itself to adaptation in the third dimension: our study is thus as much an initial exploration into the richness of the wavelet transform as a way of characterizing terrestrial heterogeneity as an encouragement to further study.

The first breakdown is as a function of depth and by scale of the D4 decomposition, as shown in Fig. 7. To aid in the interpretation we remind the reader of the dominant wavelengths that are represented at a specific scale by referring to Fig. 3, where of course it should be noted that the area of the panels decreases with the square of the depth in the Earth.

The main observations relevant to both the Montelli et al. (2006) and the Ritsema et al. (2010) models are that seismic wavespeed heterogeneity has a dominantly "red spectrum" (Chevrot et al. 1998a,b; Boschi & Dziewoński 1999). Figs. 7a and 7c show the maximum absolute values of the wavelet and scaling coefficients at each of the four scales in the D4 decomposition, as a function of depth. The scaling functions at scale 4 (denoted "scals 4" in the legend; these are depicted in Fig. 3a) require the largest expansion coefficients; the maxima of the coefficients corresponding to the wavelets at scale 4 ("wavs 4", see Figs 3b–d) are only about half as large; those at scale 3 ("wavs 3", see Figs 3e–g)

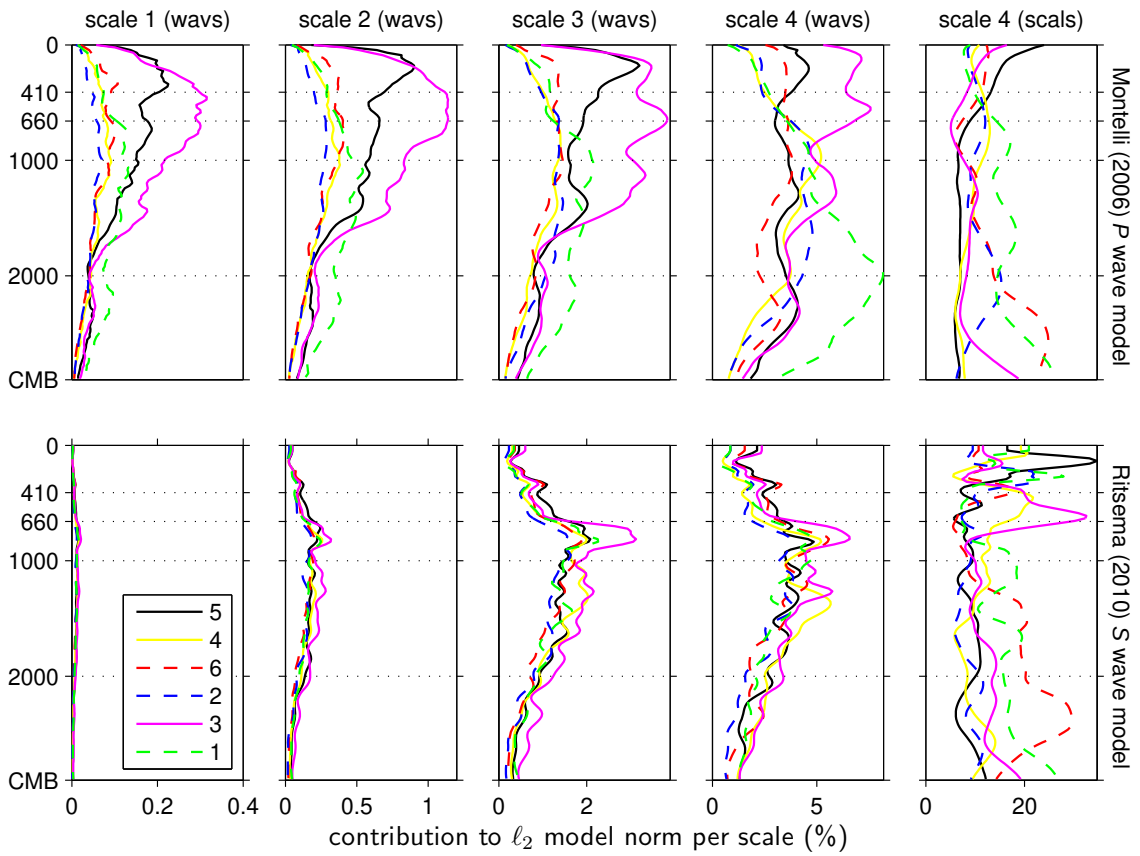

Figure 8. Scale lengths of seismic heterogeneity as a function of depth in the Montelli et al. (2006) *P*-wave and the Ritsema et al. (2010) *S*-wave models. The calculations are identical to those reported in Fig. 7 but they are now broken per cubed-sphere chunk to reveal geographical variations in seismic mantle structure. See Fig. 2 for the numbering scheme used in the legend identifying the colored lines: roughly speaking, 1 corresponds to the Pacific, 2 to Antarctica, 3 to Asia, 4 to South America, 5 to North America, and 6 to Africa. The relative lack of fine structure at scales 1 and 2 and the less geographically differentiated character at scales 3 and 4 of the Ritsema model clearly distinguishes it statistically from the Montelli model. Other features are more persistent between models, such as the predominantly large-scale structure near the core-mantle boundary underneath Africa and the Pacific, and the predominantly smaller-scale features in the shallow mantle and crust underneath Asia and North America.

peak at about half that; and so on. While noting that the Montelli model has peak amplitudes that are about half as large as the ones in the Ritsema model, in both models the overall largest values are in the lithosphere, which encompass the crust and shallowmost mantle down to about 250 km. The upper mantle (down to 660 km) and the transition zone (410–660 km) in particular are characterized by strong maxima that fluctuate with depth. Both seismic models have a somewhat different take on this measure of mantle structure: the maxima in the Ritsema model (Fig. 7c) are more oscillatory with depth and have a strong peak around the 660 km mantle discontinuity which is broader than the corresponding one in the Montelli model (Fig. 7a). Each of the curves in Figs. 7a and 7c decays sharply with increasing depth in the lower mantle below 660 km depth to reach their smallest maxima in the mid-mantle before increasing again in the bottom 1000 km, near the core-mantle boundary (CMB). This identification of dominantly long-wavelength structure near the core-mantle boundary (see also Wysession 1996; van der Hilst & Kárason 1999) is relatively more pronounced in the Montelli model than in Ritsema's. In Montelli's *P* wave model (Fig. 7a) both scales 3 and 4 have significant "bumps" near the CMB, while the corresponding increase in maximum structure in Ritsema's *S* wave model (Fig. 7c) is more gradual and confined mostly to the longest-wavelength scaling functions at scale 4 (see also Wysession et al. 1999).

The maximum values of the expansion coefficients in the wavelet basis provide but one part of interpretation of mantle structure, thus in Figs. 7b and 7d we plot the percentage-wise relative contribution of the wavelet and scaling coefficients at each scale to the overall ℓ_2 norm of the respective seismic models. These curves again reveal the scale and depth dependence of mantle heterogeneity, but now in terms of how much variance is explained by each scale at every depth individually: each of the curves sums to very nearly 100% at every depth. Their failure to sum to *exactly* 100% arises from the preconditioning of the wavelet transforms at the edges, which renders even the D4 transforms slightly non-orthonormal overall; however, these small ($< 1\%$) deviations are not sufficiently important to influence any of the interpretations. In this analysis we note that once again the relative contributions to model structure are more variable with depth in the Ritsema model (Fig. 7d) than in the Montelli model (Fig. 7b), which is particularly smooth in this regard. In both, however, the importance of the structure at scale 3 grows as a function of depth to reach a maximum about one third of the way down. This maximum is particularly well pronounced in Ritsema's model where it is well localized at the top of the lower mantle, between 660 km and 1000 km depth.

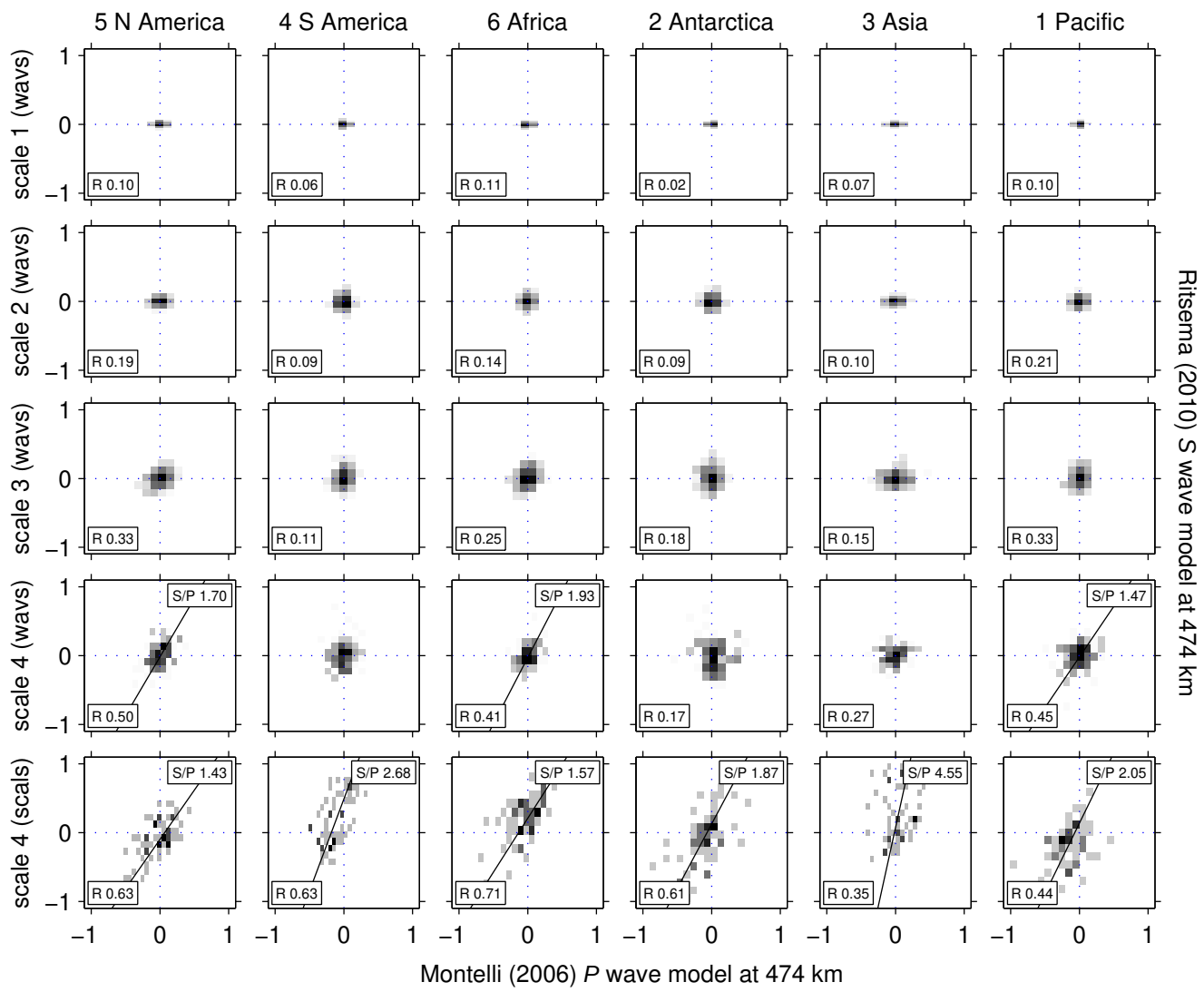

Figure 9. Joint properties of seismic mantle structure in the Montelli et al. (2006) *P*-wave and Ritsema et al. (2010) *S*-wave speed models, at 474 km depth in the Earth. Every row corresponds to a different scale in the D4 wavelet decomposition of the models. Each panel shows the logarithmic density of observations. Black shading corresponds to the maximum density in each panel, all patches that account for less than 1% of the observations are rendered white. Using total least squares a regression line was fit to all sets with correlation coefficients exceeding 0.35. The slope of the line, a measure of the $\delta \ln V_S / \delta \ln V_P$ ratio appears in the top right corners. Correlation coefficients are quoted in the bottom left corners when they are deemed significant at the 95% level.

The growth of scale 3 structure comes at the expense of scale 4 structure, suggesting that in that depth range long-wavelength heterogeneity is broken down to smaller scales.

Serving as a window into the Earth's structural heterogeneity as well as a useful comparison between models comes in the form of Fig. 8, where we are able to deconstruct both of the seismic models under consideration on a chunk-by-chunk basis. The (arbitrary and thus easily modified) choice we made in Fig. 2 to deviate from the canonical Ronchi et al. (1996) orientation of the cubed sphere by approximately centering each of the faces on a major continental landmass now allows us to study the relative contributions of the depth-dependent seismic structure broken down by preponderant scale length as a function of location in the Earth. Each of the curves originally plotted in Figs 7b and 7d degenerates to six individual ones with their own geographical affiliation. The numbering scheme is the one introduced in Fig. 2, thus in order of appearance, 1 corresponds to the Pacific realm, 2 to Antarctica, 3 to most of Asia, 4 to South America, 5 to North America and parts of Eurasia, and 6 to Africa, the middle East and the Arabian Peninsula. In the computer code that accompanies this paper any other wholesale rotation may be applied to the master grid, e.g. to undo the somewhat unfortunate splitting of Australia over chunks 2 and 3 and of Eurasia over chunks 3, 5 and 6. In other words, the cubed-sphere wavelet transform may be applied in "detector" mode by rigid rotation to center on any point of interest. Moreover, provided the scales to be analyzed allow it, any geographical portion of the wavelet-transformed coefficients may be zeroed out to provide even more geographical selectivity without compromise. Such is the power derived from multi-resolution and scale-space localization under the wavelet transform.

Among other features the results presented in Fig. 8 reveal how the dominantly long-wavelength structure near the core-mantle-boundary is mostly due to what lies beneath Africa and the Pacific: indeed these are regions that have been long known for

being the source of various long-wavelength mantle upwellings or (super-)plumes (Ni & Helmberger 2003). As to Ritsema's model, it is surprising how little mantle structure is present at the very shortest wavelengths of scale 2 (see Figs 3h–i) and scale 1 (whose footprint, not shown in Fig 3, is exactly half that of scale 2). While also in Montelli's model the heterogeneity at these scales remains limited at the sub-percentage level, there is considerable more energy that contributes to the model norm, and there is much more geographical variability between chunks in this latter model. The relative lack of a geographical signature when comparing Ritsema's to Montelli's model continues to be apparent at the larger scales 3 and 4; only at scale 4 do both models ascribe mantle structure with significant difference to each of the six gross mantle domains. Presumably this rather different character between both models is due to the data selection and model parameterization: Ritsema's model contains the effect of the whole-mantle sensitivity of normal-mode splitting functions and the spread-out influence of long-period surface waves. Moreover, this model is derived in terms of global spherical harmonics (Ritsema et al. 2010), although the resolution gains from including spherical harmonic basis functions to degree and order 40 as compared to an earlier iteration of this model (Ritsema et al. 1999, 2004; Ritsema 2005) appear modest. Montelli's model, in contrast, contains only body-wave observations, albeit using finite-frequency sensitivity theory which noticeably "fattens" their traditional, ray-theoretical, zone of influence (Montelli et al. 2004, 2006), and it is parameterized on a grid of tetrahedral nodes that, while globally distributed throughout the Earth's volume, allows for more degrees of freedom and hence spatial variability in the recovered seismic model. Undoubtedly the scale- and space-dependent breakdown of both models is also influenced by the different choices of damping and smoothing in the inverse problem that led to their construction (Boschi & Dziewoński 1999). Thus, while our analysis cannot claim to uncover the "truth" in characterizing Earth structure, it does however, endow us with a measurement tool for the multi-scale dependence of seismic *model* structure. This will serve as a target to reconcile such models with what we can learn from forward geodynamical modeling or in their confrontation with mineral physics observations (e.g. Mégnin et al. 1997; Piromallo et al. 2001; Cammarano et al. 2005; Piazzoni et al. 2007; Ritsema et al. 2007; Bull et al. 2009).

A more detailed multi-scale comparison between both seismic models involves the joint properties of their wavelet coefficients at all depths in the Earth. An example is shown in Fig. 9, where we report the correlation between wavelet coefficients in the Montelli and Ritsema models as a function of scale and approximate geographical position (see Fig. 2 for the numbering scheme of the cubed-sphere chunk). A rendering of the two-dimensional density of the data is accompanied by the value of their correlation coefficient (lower left labels) where this is deemed significant at the 95% level, and the slope of the total-least-squares based fit in this space (upper right labels), which is only quoted when the correlation coefficients exceeded 0.35. This should provide an estimate of the logarithmic ratio of shear-wave to compressional-wave speed perturbations, $\delta \ln V_S / \delta \ln V_P$, an important discriminant in the interpretation of the (thermal or chemical) cause of seismic velocity anomalies (Masters et al. 2000; Trampert & van der Hilst 2005). The variation of this ratio as a function of scale and chunk position yields information that will be of use for geochemical and geodynamical studies, and the orthogonality of the wavelet basis in scale and physical space removes some of the arbitrariness in the calculation. The depth of 474 km that we selected for illustration displays the largest overall correlation between both models. The absence of information at the smallest scales 1 and 2 in Ritsema's model is now apparent from the top two rows in Fig. 9: invariably, the *P* anomalies from Montelli's model map to much smaller *S* perturbations in Ritsema's. From scale 3 onward a positively correlated pattern begins to emerge, though at this particular scale, the correlation coefficients remain below the somewhat stringent 0.35 level that we have set for ourselves. Wavelets and scaling coefficients are rather well correlated at the largest scale 4 considered, with several of the correlation coefficients comfortably exceeding our threshold. The value of the $\delta \ln V_S / \delta \ln V_P$ ratios vary between about 1.5 and 4.5, which represents about half the range that they reach when all depths are being considered. This is in agreement with previous studies, e.g. those by Tkalčić & Romanowicz (2002), Saltzer et al. (2001) and Deschamps & Trampert (2003). With these last two authors we emphasize how regionally variable such ratios are, and how ultimately, they remain dependent on data availability and modelling assumptions entering the production of independently derived *P* and *S* mantle models Trampert & van der Hilst (2005). Robust as these determinations of $\delta \ln V_S / \delta \ln V_P$ ratios are using the wavelet decomposition scheme, our analysis has now identified an additional dependence, on scale length.

5 THE INVERSE PROBLEM

In the previous sections we have constructed a new wavelet transform on the three-dimensional ball. We have shown that, in a suitably chosen wavelet basis, Earth models require few significant coefficients. We have used our wavelet scheme to deconstruct two tomographic Earth models and evaluated those, both for their sparsity and to study the distribution of mantle structure as a function of scale, depth, and geographical location (s). While we have argued that we can learn much from such exercises, we have only partially reached our end goal, which is to harness the power and performance of spherical wavelet bases to build *new* seismic tomographic models, directly from the data, and which are expected to be sparse in such bases. In other words, we have not solved any inverse problems yet. In this section we explain how the new wavelets can be used to do that, too.

Wavespeed models are constructed from seismic data. With respect to a reasonably sized global model parametrization these data are incomplete, as seismic stations are mostly concentrated in a limited number of regions around the globe — that is, until the oceanic arrays of the future (Simons et al. 2009; Vincent & Simons 2011). As usual we shall assume that a background velocity model is known, and that our goal is to solve the data for a perturbation $m(\mathbf{x})$ to that reference model. We may approximate the seismic observations

$$\int_{\oplus} K(\mathbf{x})\, m(\mathbf{x})\, d^3\mathbf{x} = d, \tag{8}$$

which are of the most general kind described by such integral equations and with K any of a veritable plethora of possible kernel functions (Nolet 2008), by the discretization on the grid defined in Section 2. This leads to an inverse problem in matrix form,

$$\mathbf{K} \cdot \mathbf{m} = \mathbf{d}, \tag{9}$$

where the aim is to reconstruct the model values $\mathbf{m}$ from the data vector $\mathbf{d}$. The elements of $\mathbf{m}$ are the values of the model inside of each voxel and the elements of every row of $\mathbf{K}$ will be the numerical values of the integral of the kernel $K(\mathbf{x})$ over those voxels.

Eq. (9) remains beholden to the usual assumption of linearity

in linking the model perturbation $\mathbf{m}$ to the data $\mathbf{d}$. Acknowledging that the data may be contaminated by (Gaussian) noise $\mathbf{n}$, the inverse problem is defined as requiring us to find the best choice of $\mathbf{m}$ by which to reduce the data misfit: the squared ℓ_2 norm $\|\mathbf{K}\cdot\mathbf{m}-\mathbf{d}\|_2^2$, to the noise level, $\|\mathbf{n}\|_2^2$. Because the data are incomplete, the problem is ill-posed and infinitely many such models exist. Additional conditions need to be imposed to arrive at a unique and physically acceptable solution. This is often done by adding a penalty term $\mathcal{P}(\mathbf{m})$ to the data misfit, which leads to the functional

$$\mathcal{F}(\mathbf{m}) = \|\mathbf{K}\cdot\mathbf{m}-\mathbf{d}\|_2^2 + \mathcal{P}(\mathbf{m}), \tag{10}$$

which is to be minimized. The role of the penalty term is to ensure that $\mathcal{F}$ has a unique and acceptable minimizer. The trade-off between data fit and *a priori* information is encoded in the penalty $\mathcal{P}$. A convenient and often advocated choice for $\mathcal{P}(\mathbf{m})$ is a multiple of the norm-squared of the Laplacian of the model, $\mathcal{P}(\mathbf{m}) = \lambda\|\nabla^2\mathbf{m}\|_2^2$, which favors smoothness in the solutions (see, e.g., Yanovskaya & Ditmar 1990; VanDecar & Snieder 1994). The equations for the minimum of $\mathcal{F}(\mathbf{m})$ remain linear,

$$\mathbf{K}^{\mathrm{T}}\cdot\mathbf{K}\cdot\mathbf{m} + \lambda(\nabla^2)^{\mathrm{T}}\nabla^2\mathbf{m} = \mathbf{K}^{\mathrm{T}}\cdot\mathbf{d}, \tag{11}$$

and can thus be handled by standard algorithms. The trade-off parameter λ needs to be carefully chosen (Hansen 1992).

The novelty now is that we should be able to use model *sparsity* rather than smoothness as prior information, though we should qualify this statement by the admission that the choice of wavelet basis itself (D4, CDF 4–2, and so on) introduces a certain level of smoothness in built into the representation itself. As discussed in Section 3 seismic tomographic models may be very well represented by a sparse wavelet expansion. Incorporating this knowledge from the start may therefore lead to important benefits to the behavior of the inversion scheme. In the following we shall assume that we have chosen a particular family of wavelet and scaling basis functions, see Section 2, to represent and build the unknown model. Below we describe an algorithm that chooses, based on the data, the number, position and magnitude of the nonzero coefficients with respect to this basis as it iterates (i.e. during the inversion, not beforehand). Once the family of wavelet functions is chosen, the selection of the non-zero coefficients is therefore part of the inversion algorithm itself. The choice of the nonzero coefficients is not a pre-processing step: the representation of the unknown model is both parsimonious and data-driven, much like in partition-modelling approaches (Bodin et al. 2009; Bodin & Sambridge 2009). On the other hand, the choice of basis is made up front, and a basis of smooth functions will introduce some smoothness in the model (e.g. when, due to absence of data, a whole region can be well approximated using a single large-scale smooth wavelet function).

With the model $m(\mathbf{x})$ expanded in our wavelet basis via the transform A as in the notation of Section 3, and the individual basis functions collected in the columns of a matrix $\mathbf{S}$, the synthesis map, the pixel-basis model vector $\mathbf{m}$ is

$$\mathbf{m} = S(\mathbf{w}) = \mathbf{S}\cdot\mathbf{w}, \tag{12}$$

with $\mathbf{w}$ the vector of expansion coefficients in this basis. In having previously defined our construction in terms of a discrete wavelet transform we do not need to devise a separate form of discretization for each of the many choices of wavelet bases that are available to us. In this flexible approach we define the grid size of the cubed sphere at the outset and we are thus able to switch between the various wavelet bases without much additional effort. As we shall remark later on $\mathbf{S}$ will usually be provided as a (fast) software algorithm and not as a matrix *per se*. We shall also see that the seismic inversions only require application of $\mathbf{S}$ and its transpose $\mathbf{S}^{\mathrm{T}}$. The inverse $\mathbf{S}^{-1}$, the analysis map, is not required to be known — or even exist, as is the case for a redundant set of basis functions.

The sparsity of the model parameters $\mathbf{w}$ can now be encouraged by choosing the penalty $\mathcal{P}$ to be proportional to the number of nonzero entries in $\mathbf{w}$, which we write as $\|\mathbf{w}\|_0$ for short. The functional to be minimized then becomes

$$\mathcal{F}_0(\mathbf{w}) = \|\mathbf{K}\cdot\mathbf{m}-\mathbf{d}\|_2^2 + \lambda\|\mathbf{w}\|_0. \tag{13}$$

We define the solution to the inverse problem as

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \left(\|\mathbf{K}\cdot\mathbf{S}\cdot\mathbf{w}-\mathbf{d}\|_2^2 + \lambda\|\mathbf{w}\|_0 \right), \tag{14}$$

and the reconstructed model is

$$\hat{\mathbf{m}} = \mathbf{S}\cdot\hat{\mathbf{w}}. \tag{15}$$

The functional in eq. (14) however is not convex: there exist local minima which makes the minimization much less feasible than solving a system of linear equations. Despite this an iterative algorithm based on hard thresholding exists (Blumensath & Davies 2008, 2009), as briefly discussed by Loris et al. (2010).

An alternative, and computationally much more tractable, method for imposing model sparsity in a given basis is to use an ℓ_1 norm penalty (Donoho 2006; Daubechies et al. 2004; Bruckstein et al. 2009). By identifying $\|\mathbf{w}\|_1 = \sum_i |w_i|$, and choosing $\mathcal{P} = 2\lambda\|\mathbf{w}\|_1$ for the penalty function, then

$$\mathcal{F}_1(\mathbf{w}) = \|\mathbf{K}\cdot\mathbf{S}\cdot\mathbf{w}-\mathbf{d}\|_2^2 + 2\lambda\|\mathbf{w}\|_1 \tag{16}$$

is to be minimized. This functional is convex: a local minimum is therefore automatically a global minimum (Loris et al. 2007). This minimum is not necessarily identical to global minimum in eq. (14), but mathematical conditions on the matrix $\mathbf{K}$ exist that guarantee this (Donoho 2006). The important point is that minimizing eq. (16) is atractable way of obtaining a model with many coefficients *exactly* equal to zero.

The functional (16) is not differentiable but because $\mathcal{F}_1$ is the sum of a differentiable and a separable non-differentiable part, convex optimization techniques can find $\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathcal{F}(\mathbf{w})$ and the corresponding model (15) with reasonable efficiency. Indeed the iteration

$$\mathbf{w}_{n+1} = \mathcal{U}[\mathbf{w}_n + \beta_n(\mathbf{w}_n - \mathbf{w}_{n-1})], \tag{17a}$$

$$\mathcal{U}(\mathbf{w}) = \mathcal{T}_{\alpha\lambda}\left[\mathbf{w} + \alpha\,\mathbf{S}^{\mathrm{T}}\cdot\mathbf{K}^{\mathrm{T}}\cdot(\mathbf{d}-\mathbf{K}\cdot\mathbf{S}\cdot\mathbf{w})\right], \tag{17b}$$

converges to the minimizer of (16), as shown by Beck & Teboulle (2009). Hereby $\mathcal{T}_{\alpha\lambda}$ now stands for "soft" thresholding (Mallat 2008) of the coefficients on a component-by-component basis, which is to say $\mathcal{T}_\tau(\mathbf{w}) = 0$ for $|\mathbf{w}| \le \tau$, and $\mathcal{T}_\tau(\mathbf{w}) = \mathbf{w} - \tau\,\mathrm{sgn}(\mathbf{w})$ for $|\mathbf{w}| > \tau$. This is a non-linear operation. The parameter α in eq. (17b) can be chosen as the reciprocal of the largest eigenvalue of $\mathbf{S}^{\mathrm{T}}\cdot\mathbf{K}^{\mathrm{T}}\cdot\mathbf{K}\cdot\mathbf{S}$. We choose $t_0 = 1$ and

$$\beta_n = (t_n - 1)/t_{n+1}, \tag{17c}$$

$$t_{n+1} = (1+\sqrt{1+4t_n^2})/2. \tag{17d}$$

A non-iterative direct algorithm also exists (Efron et al. 2004; Loris 2008), but because of the large problem sizes typically encountered in seismic tomography, we focus here on this so-called Fast Iterative Soft Thresholding Algorithm (FISTA). It has an $1/n^2$ rate of convergence to $\mathcal{F}_1(\hat{\mathbf{w}})$, which is in a sense optimal. The algorithm (17) was used by Loris et al. (2010) on a 3D toy tomographic model. There is however a typographical error in that work, which missed the factor 4 under the square root in eq. (17d).

The iterative algorithm (17) requires only two linear maps, and their transposes. First there is the linear map from model to data space, given by the matrix $\mathbf{K}$ in eq. (9). The second is the linear map $\mathbf{S}$ from model parameters to model space, eq. (12). This map is typically available in the form of a (fast) algorithm, *in casu* the inverse wavelet transform S, rather than explicitly in matrix form. Each iteration step of algorithm (17) requires one application of $\mathbf{K}$, $\mathbf{K}^{\mathrm{T}}$, $\mathbf{S}$ and $\mathbf{S}^{\mathrm{T}}$ each. Eqs (17a–17d) demonstrate that the iterative inversion algorithm does not require the inverse of the map $\mathbf{S}$, much as it does not require the inverse of $\mathbf{K}$. Moreover, neither $\mathbf{S}$ nor $\mathbf{K}$ need be invertible. As already mentioned, this means that a model may be represented by a sparse superposition of a redundant set of functions in which the expansion of the model is no longer unique. For example, redundant dual-tree wavelets were used in a synthetic tomography experiment by Loris et al. (2007).

In practice it turns out to be easier to keep $\mathbf{K}\cdot\mathbf{S}$ and $\mathbf{S}^{\mathrm{T}}\cdot\mathbf{K}^{\mathrm{T}}$ in eq. (17b) in factorized form. One can then easily switch bases by modifying $\mathbf{S}$ and rerunning the inversion algorithm. No new matrix $\mathbf{K}\cdot\mathbf{S}$ needs to be pre-computed, which is important given that $\mathbf{K}$ may have several hundreds of thousands of rows. This is particularly useful in the case of sparse reconstructions, where the choice of basis itself (e.g. D4, CDF 4–2, and so on, and this before the inversion determines which members of the basis set most usefully contribute to the solution) is one of the factors to be assessed by prior evaluation of the performance of synthetic model inversions or by inspection during the inversion. In other words, the iterative inversion procedure can be performed for a number of different choices of wavelet families. Section 3 made it clear that using model sparsity as *a priori* information depends on the details of the basis used. Setting up the inversion software in this manner is therefore forward-looking as new transforms can easily be incorporated later. Examples of emerging techniques that can be evaluated in this context are curvelets and shearlets (Candès et al. 2005; Labate et al. 2005; Easley et al. 2008), which offer better directional sensitivity than classical wavelet transforms but are redundant. The described flexibility of our approach was a major design requirement and will yield many dividends in future applications.

6 A SECOND CONSTRUCTION

In principle we are now ready to apply the first family of wavelet constructions on the cubed sphere that we introduced in Section 2 to the inverse problem in the manner outlined in Section 5. As shown in Section 3 we expect our solutions to be sparse, and as discussed in Section 4 we will be able to use this sparsity and the location- and scale-dependence of the results to make geophysical inference about the structure of seismic heterogeneity in the Earth.

As we recall, our First Construction entailed defining wavelet and scaling functions on a single chunk $\xi,\eta\in[-\pi/4,\pi/4]$ and then mapping them onto the sphere using eq. (1). By this definition the basis functions live on a single chunk. Without the modifications and preconditioning of the basis at the boundaries between the chunks that we introduced, sharp discontinuities in the behavior of the coefficients occur at the chunk edges; making the transforms edge-cognizant, as we did in the manner of Cohen et al. (1993), required special tailoring of the transforms. This is often cumbersome and in general harms our stated goal of keeping our procedure flexible enough to be able to switch from one wavelet family to another which might be more suitable with hindsight. In addition, the interval wavelet transforms that we used so far are not norm-preserving. Extensive experimentation with such bases revealed that despite their qualities in the representation of geophysical functions, i.e. in performing the forward mappings, when used for the inverse problem the solutions obtained using eq. (17) were plagued by unsightly artifacts at the seams between chunks. These we do not show as they appeared predominantly in synthetic tests with simple, isolated "phantoms" where their nature was immediately obvious; as the density of path coverage increases and the target structure becomes more Earth-like, their presence becomes harder to distinguish visually. Presumably the ℓ_1-thresholding could be adapted locally to counter this effect, but to be truly practical we should not have to resort to this. We thus desire a mechanism to map any localized basis function defined on a Cartesian grid to the sphere, with smoothness even across chunk boundaries. Here we present a straightforward, universal method that accomplishes this.

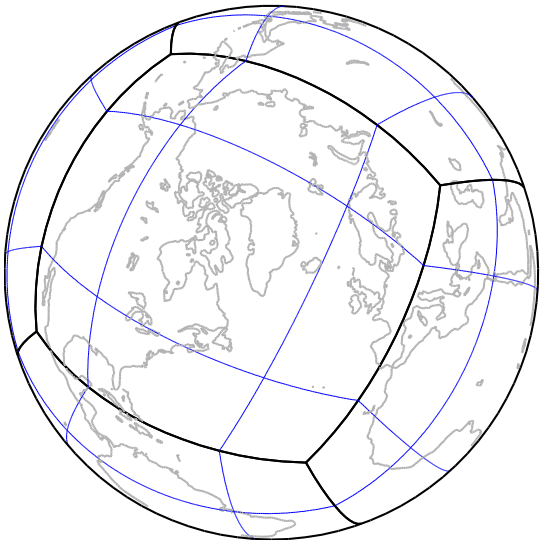

Figure 10. Aerial view showing our second adaptation of the cubed sphere of Ronchi et al. (1996). The black lines identify the boundaries of the six chunks that were apparent also in Fig. 1. The blue lines correspond to the boundaries of the overlapping "superchunks" as discussed in the text.

As opposed to the geometry of the Ronchi et al. (1996) cubed sphere shown in Fig. 1, we now cover the sphere with six larger chunks, by extending the coordinates by 50% on each chunk, to $\tilde{\xi},\tilde{\eta}\in[-3\pi/8,3\pi/8]$, see Fig. 10. We shall refer to these partially overlapping domains as "superchunks". In $(\tilde{\xi},\tilde{\eta})$ coordinates they are simply six large squares (rather: cubes if we take the radial direction into account also), with the "original" chunks at their centers. Functions defined on this central part can now smoothly cross into the outer part, that is, they are allowed to spill over into another chunk while staying in the same superchunk. Fig. 11 shows a selection of examples where this is the case. The smoothness of the functions across the boundaries is apparent, though we note that if we were to plot them in the manner in which Fig. 2 was presented, they would appear to have kinks in them; this is simply because the coordinate transform of eq. (1) itself is non-smooth.

To map a function defined on a single superchunk $\tilde{\kappa}=1\rightarrow 6$ to the corresponding chunk and its neighbors, one loops over all the voxels in this central chunk and its four neighbors. The center (ξ,η,r,κ) of each such voxel is mapped to (x,y,z) coordinates using formula (1). In the same way as eq. (2) we then calculate

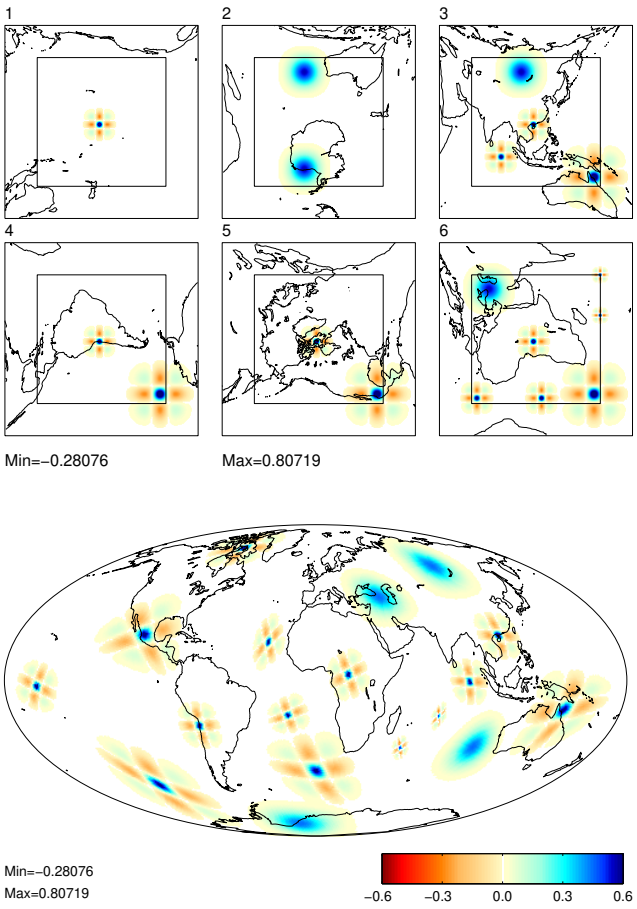

Figure 11. Top: Six superchunks, as defined in Fig. 10 with a number of wavelet and scaling functions defined on them. Bottom: The same functions mapped to the sphere by the procedure described in Section 6. They are smooth everywhere.

$$(\tilde{\xi},\tilde{\eta}) = \begin{cases} [\,\mathrm{atan}(z/y), \mathrm{atan}(-x/y)\,] & \text{if} \quad \tilde{\kappa}=1, \\ [\,\mathrm{atan}(y/x), \mathrm{atan}(-z/x)\,] & \text{if} \quad \tilde{\kappa}=2, \\ [\,\mathrm{atan}(-y/z), \mathrm{atan}(x/z)\,] & \text{if} \quad \tilde{\kappa}=3, \\ [\,\mathrm{atan}(x/z), \mathrm{atan}(-y/z)\,] & \text{if} \quad \tilde{\kappa}=4, \\ [\,\mathrm{atan}(-z/x), \mathrm{atan}(y/x)\,] & \text{if} \quad \tilde{\kappa}=5, \\ [\,\mathrm{atan}(-x/y), \mathrm{atan}(z/y)\,] & \text{if} \quad \tilde{\kappa}=6, \end{cases} \tag{18}$$

to convert these (x, y, z) to the $(\tilde{\xi}, \tilde{\eta}, \tilde{r} = r)$ coordinates in the superchunk $\tilde{\kappa}$, limited to $-3\pi/8 \le \tilde{\xi}, \tilde{\eta} \le 3\pi/8$. This then determines which voxel in the superchunk is mapped to the voxel in the original chunk. The index of the voxel in each superchunk $\tilde{\kappa}$ is

$$i = 1 + \left\lfloor \left(\tilde{\xi} + \frac{3\pi}{8}\right) \frac{2N}{\pi} \right\rfloor, \quad j = 1 + \left\lfloor \left(\tilde{\eta} + \frac{3\pi}{8}\right) \frac{2N}{\pi} \right\rfloor, \tag{19}$$

where N is the number of voxels in the ξ and η directions of a chunk and $\lfloor\,\rfloor$ indicates rounding down. Voxel indices run from $1 \to N$ in an original chunk and from $1 \to 3N/2$ in a superchunk. The central part of a superchunk is a copy of the original chunk, whereas the voxels outside the center of a superchunk are mapped to neighboring chunks. As the superchunks partially overlap, a chunk voxel on the sphere may receive contributions from up to three superchunks: a voxel near a chunk corner may receive three contributions, a voxel near a chunk edge may receive two, and voxels near chunk centers only one. The identifications are most easily made by table look-up.

In Fig. 11 we show a number of wavelet functions from this Second Construction at a variety of locations. These now map smoothly to the sphere. The wavelets shown here are from the Cohen et al. (1992) CDF 4–2 wavelet family, as in Figure 4. These are mirror-symmetric in the (ξ, η) domain, but they are no longer orthogonal. As in Section 2 the wavelets at a fixed scale are not rotations of each other on the sphere, but rather translates in the superchunk $(\tilde{\xi}, \tilde{\eta})$ domain. This effect is most noticeable for the wavelet and scaling functions that are located on or near chunk edges, specifically near the corners. Basis functions that have the same norm in the superchunk domain may not have the same norm in the chunk domain.

7 NUMERICAL EXPERIMENTS

We consider a set of great-circle paths that is a global collection of 2469 earthquakes and 199 stations yielding 8490 surface-wave paths, a situation based on, if not identical to, the ray path coverage in the models of Rayleigh-wave phase speeds at 80 s period obtained by Trampert & Woodhouse (1995, 1996, 2001). For simplicity we convert this path coverage to the ray-theoretical values of arrival-time sensitivity expressed in our model domain. The image in Fig. 12 (top panel) renders all rays in this data set of realistically heterogeneous global seismic sensitivity. For synthetic input model we chose a single interval of the Montelli et al. (2006) model centered on 722 km depth, shown in Fig. 12 (second from the top). In addition, and this is an admitted departure from realism, we select four circular regions of null structure. Their purpose is to test the inversion algorithm and the choice of basis when sharp wavespeed contrasts are known to be present in the true model. We calculate the travel-time perturbations over these 8490 ray paths and add Gaussian noise to them with an rms value that is 10% of that of the rms of the data. The variance of this noise is denoted σ^2.

The reconstruction is by the algorithm (17) using the four-level CDF 4–2 wavelets under the Second Construction by which smooth chunk crossings were enabled, as shown in Fig. 11. In keeping with the description of Section 5 the dual aim is to satisfy the noisy data in the traditional ℓ_2 sense while favoring a model that is sparse in the wavelet basis by minimizing the ℓ_1 norm of the coefficients. Fig. 12 (third from the top) shows the obtained solution. Of the $6 \times 128^2 = 98304$ degrees of freedom in this parameterization the algorithm terminates on a model with 1670 nonzeroes. Due to lack of data, the relative output error is high: 33.5%.

The behavior of the solution through the 1000 iterations is shown in Fig. 12 (bottom panel), which plots the ℓ_1 norm of the wavelet coefficients against a measure of the evolving misfit calculated as the "reduced chi-square"

$$\frac{\chi^2}{N} = \frac{\|\mathbf{d} - \hat{\mathbf{d}}\|_2^2}{\sigma^2 \|\mathbf{d}\|_0} = \frac{\|\mathbf{d} - \mathbf{K}\cdot\hat{\mathbf{m}}\|_2^2}{\sigma^2 \|\mathbf{d}\|_0}, \tag{20}$$

in other words, the squared ℓ_2 norm of the data misfit normalized by the noise variance and the total number of data constraints, for which it is reasonable to assume (Nolet 2008) that it is distributed as a χ_1^2 variable. The starting point of the iteration is marked by the filled white circle, and the final solution by the filled white square, which is arrived at when the χ^2/N variable in eq. (20) reaches its expectation 1. Every one of the 1000 solutions in the sequence is marked by a black cross. As we notice the algorithm (17) rapidly reduces the data misfit in the first few steps, slowing down after that, at the same time increasing the sum of the absolute values of the wavelet coefficients. After "turning a corner" in this space, the remainder of the time spent is in reducing the ℓ_1 norm of the

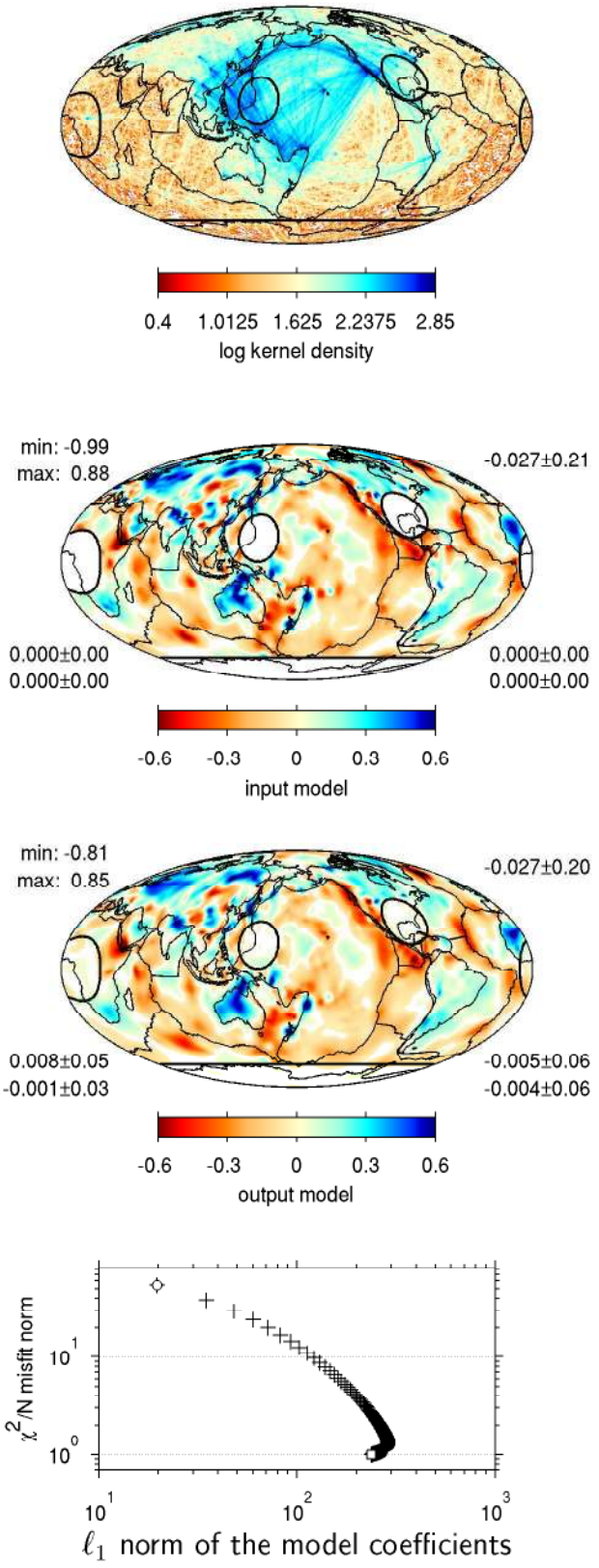

Figure 12. Synthetic experiment under realistic conditions with a twist, illustrating the recovery of a seismic tomographic model with artificially introduced "blank spots" from noisy data, using the Second Construction discussed in Section 6 and the iterative algorithm of Section 5. The solution after one iteration is represented by the filled white circle in the bottom panel; the solution after 1000 iterations by the filled white square.

coefficients of the solution while slowly converging to the target reduced chi-square of $\chi^2/N = 1$.

The solution is very good; the input model is well matched and the leakage of the solution into the areas where no structure should be recovered is relatively minor. The global map views in Fig. (12) had all values that fall below a threshold of $1/20^{\text{th}}$ of their maximum absolute value rendered white for visual guidance, and they receive several additional annotations for us to be able to judge the quality of the solution quantitatively. The minimum and maximum values of the models are quoted in the top left corner, and in the top right corner we show their mean and the root mean squared values. The four sets of numbers in the bottom left and right hand corners quote these same metrics for the areas contained inside of the circular areas. As the input model has no structure there, all of these are zeroes. This is no longer the case for the output which serves as our way to evaluate the leakage of the solution into those areas. As we can see the comparison is very favorable.

While we have conducted numerous synthetic tests with a multitude of synthetic input models (including checkerboard tests, Gaussian shaped anomalies positioned at various locations, and using a variety of ray path coverages), only one of these tests is reported here. As noted by Loris et al. (2007, 2010) there are more algorithms available to us than the one described in eq. (17), and as we have argued in this paper there is a plethora of wavelet constructions that can be brought to bear on the inverse problem of global seismic tomography. All of these alternatives remain in principle candidates to be implemented using our First or Second Construction for wavelets on the sphere. A more detailed comparison of their relative performance is well underway and will be reported in forthcoming work. It is there also that we will fully integrate the third dimension into our formalism. Conceptually, there is no difficulty in doing this: we have transformed the ball of the Earth into six independent or partially overlapping Cartesian model domains with three separable coordinates. Taking into account the depth dimension merely involves applying a third wavelet transform to the result of the transform in the two angular coordinates, but as there remain choices to be made, a thorough discussion remains outside the scope of the current paper.

8 CONCLUSIONS

Until now, seismic wavespeed models of the Earth have been routinely parameterized in terms of spherical harmonics, networks of tetrahedral nodes, rectangular voxels, or spherical splines. However, there were few approaches to Earth model parameterization by wavelets on the three-dimensional ball. To the rich field of wavelets on the ball or its surface, the sphere, we have contributed two new flexible constructions that are eminently suited to solve seismological tomographic inverse problems.

To form the numerical grid we considered a surface tesselation known as the "cubed sphere", popular in fluid dynamics and computational seismology, which can be combined with a (semi-regular) radial subdivision. This mapping transforms the entire volume of the mantle into six portions. In the new variables, these "chunks" correspond to rectangular boxes with Cartesian coordinates. Standard algorithms can then be used to perform the wavelet transformation (or any other) in each of the six bounded volumes. We developed two possible classes of discrete wavelet transforms in the angular dimension of the cubed sphere. One relies on preconditioning and special boundary filters to account for the edges separating the chunks; another one broadens the definition of the

cubed sphere to include chunks that partially overlap, on which we implement standard wavelet transforms.

Much has been gained by our design of procedures that efficiently parameterize the seismological inverse problem. First, the multiresolution character of a wavelet basis allows for the models to be represented with an effective spatial resolution that varies as a function of position within the Earth. Second, inversion schemes that are formulated in terms of wavelets can exploit recent theoretical and numerical advances by which the most sparse solution vector, in wavelet space, is found through iterative minimization of a combination of the ℓ_2 (to fit the data) and ℓ_1 norms (to promote sparsity in wavelet space).

In preparation of the continuing increase in high-quality seismic data that is expected in the decades to come, our focus has also been on numerical efficiency and the ability to use parallel computing in constructing the model. We have shown how our seismic model representation behaves under progressive thresholding of the wavelet coefficients, and how the geographically distributed power of published seismic models varies over the scale lengths that can be independently resolved. Synthetic tests under realistic conditions validates the approach that we advocate for the future of seismic tomography, which shows the ability to explain heterogeneous, massive data sets under the constraint that the best-fitting models should also be sparse in the wavelet bases used.

ACKNOWLEDGEMENTS

We thank Huub Douma and Massimo Fornasier for valuable discussions throughout the past several years, and the Editor-in-Chief, Jeannot Trampert, one anonymous reviewer and Malcolm Sambridge for advice which helped improve the manuscript. Comments by Sujoy Mukhopadhyay and Chris Harig inspired aspects of two of the figures. FJS was supported by Princeton University account 195-2142 (thanks to Debbie Fahey!) while enjoying the hospitality of the Vrije Universiteit Brussel and the Katholieke Universiteit Leuven in the Summer of 2010, where the writing-up started. IL is "chercheur qualifié" of the F.R.S.-FNRS (Belgium). Portions of this research were supported by VUB-GOA grant 062 to ICD and IL, the FWO-Vlaanderen grant G.0564.09N to ICD and IL, and by NSF grant CMG-0530865 to ICD, GN and F. A. Dahlen. GN and JC received support from the ERC (Advanced grant 226837) and a Marie Curie Re-integration Grant (project 223799). Computer code is freely available from `homepages.ulb.ac.be/~igloris/` and `www.frederik.net`.

REFERENCES

Aki, K., Christoffersson, A. & Husebye, E. S., 1977. Determination of the three-dimensional seismic structure of the lithosphere, *J. Geophys. Res.*, **82**(2), 277–296.

Amirbekyan, A. & Michel, V., 2008. Splines on the 3-dimensional ball and their application to seismic body wave tomography, *Inverse Problems*, **24**, 015022, doi: 10.1088/0266–5611/24/1/015022.

Amirbekyan, A., Michel, V. & Simons, F. J., 2008. Parameterizing surface-wave tomopgraphic models with harmonic spherical splines, *Geophys. J. Int.*, **174**(2), 617–628, doi: 10.1111/j.1365–246X.2008.03809.x.

Antoine, J.-P. & Vandergheynst, P., 1999. Wavelets on the 2-sphere: A group-theoretical approach, *Appl. Comput. Harmon. Anal.*, **7**, 262–291.

Antoine, J.-P., Demanet, L., Jacques, L. & Vandergheynst, P., 2002. Wavelets on the sphere: implementation and approximations, *Appl. Comput. Harmon. Anal.*, **13**, 177–200.

Bauer, F. & Gutting, M., 2011. Spherical fast multiscale approximation by locally compact orthogonal wavelets, *Intern. J. Geomath.*, **2**(1), 69–85, doi: 10.1007/s13137–011–0015–0.

Beck, A. & Teboulle, M., 2009. A fast iterative shrinkage-threshold algorithm for linear inverse problems, *SIAM J. Imag. Sci.*, **2**, 183–202, doi: 10.1137/080716542.

Becker, T. W. & Boschi, L., 2002. A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosys.*, **3**, 1003.

Becker, T. W., Chevrot, S., Schulte-Pelkum, V. & Blackman, D. K., 2006. Statistical properties of seismic anisotropy predicted by upper mantle geodynamic models, *J. Geophys. Res.*, **111**(B10), B08309, doi:10.1029/2005JB004095.

Becker, T. W., Browaeys, J. T. & Jordan, T. H., 2007. Stochastic analysis of shear-wave splitting length scales, *Earth Planet. Sci. Lett.*, **259**(3–4), 526–540.

Bergeron, S. Y., 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**(15), 2311–2314.

Blumensath, T. & Davies, M. E., 2008. Iterative thresholding for sparse approximations, *J. Fourier Anal. Appl.*, **14**(5, doi: 10.1007/s00041-008-9035-z), 629–654.

Blumensath, T. & Davies, M. E., 2009. Iterative hard thresholding for compressed sensing, *Appl. Comput. Harmon. Anal.*, **27**(3), 265–274.

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, **178**(3), 1411–1436, doi: 10.1111/j.1365–246X.2009.04226.x.

Bodin, T., Sambridge, M. & Gallagher, K., 2009. A self-parametrizing partition model approach to tomographic inverse problems, *Inverse Problems*, **25**, 055009, doi: 10.1088/0266–5611/25/5/055009.

Boschi, L. & Dziewoński, A. M., 1999. High- and low-resolution images of the Earth's mantle. Implications of different approaches to tomographic modeling, *J. Geophys. Res.*, **104**(B11), 25567–25594.

Boschi, L., Ekström, G. & Kustowksi, B., 2004. Multiple resolution surface wave tomography: the Mediterranean basin, *Geophys. J. Int.*, **157**, 293–304, doi: 10.1111/j.1365–246X.2004.02194.x.

Bozdağ, E. & Trampert, J., 2010. Assessment of tomographic mantle models using spectral element seismograms, *Geophys. J. Int.*, **180**(3), 1187–1199.

Bruckstein, A. M., Donoho, D. L. & Elad, M., 2009. From sparse solutions of systems of equations to sparse modeling of signals and images, *SIAM Rev.*, **51**(1), 34–81, doi: 10.1137/060657704.

Bull, A. L., McNamara, A. K. & Ritsema, J., 2009. Synthetic tomography of plume clusters and thermochemical piles, *Earth Planet. Sci. Lett.*, **278**(3–4), 152–162.

Cammarano, F., Goes, S., Deuss, A. & Giardini, D., 2005. Is a pyrolitic adiabatic mantle compatible with seismic data?, *Earth Planet. Sci. Lett.*, **232**(3–4), 227–243, doi: 10.1016/j.epsl.2005.01.03.

Candès, E. J., Demanet, L., Donoho, D. L. & Ying, L., 2005. Fast discrete curvelet transforms, *Multisc. Model. Simul.*, **5**, 861–899.

Candès, E. J., Romberg, J. K. & Tao, T., 2006. Stable signal recovery from incomplete and inaccurate measurements, *Comm. Pure Appl. Math.*, **59**(8), 1207–1223.

Capdeville, Y., Gung, Y. & Romanowicz, B., 2005. Towards global earth tomography using the spectral element method: a technique based on source stacking, *Geophys. J. Int.*, **162**(2), 541–554.

Chambolle, A. & Lions, P.-L., 1997. Image recovery via total variation minimization and related problems, *Numer. Math.*, **76**(2), 167–188.

Chevrot, S. & Zhao, L., 2007. Multiscale finite-frequency Rayleigh wave tomography of the Kaapvaal craton, *Geophys. J. Int.*, **169**(1), 201–215, doi: 10.1111/j.1365–246X.2006.03289.x.

Chevrot, S., Montagner, J.-P. & Snieder, R. K., 1998. The spectrum of tomographic Earth models, *Geophys. J. Int.*, **133**, 783–788.

Chevrot, S., Montagner, J.-P. & Snieder, R. K., 1998. The spectrum of tomographic Earth models: Correction, *Geophys. J. Int.*, **135**, 311.

Chiao, L. & Liang, W. T., 2003. Multiresolution parameterization for geophysical inverse problems, *Geophysics*, **68**, 199.

Chiao, L.-Y. & Kuo, B.-Y., 2001. Multiscale seismic tomography, *Geophys. J. Int.*, **145**, 517–527, 10.1046/j.0956–540x.2001.01403.x.

Cohen, A., Daubechies, I. & Feauveau, J., 1992. Biorthogonal bases of compactly supported wavelets, *Comm. Pure Appl. Math.*, **45**, 485–560, doi: 10.1002/cpa.3160450502.

Cohen, A., Daubechies, I. & Vial, P., 1993. Wavelets on the interval and fast wavelet transforms, *Appl. Comput. Harmon. Anal.*, **1**, 54–81.

Constable, S. C., Parker, R. L. & Constable, C. G., 1987. Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, **52**(3), 289–300.

Daubechies, I., 1988. Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math.*, **41**, 909–996.

Daubechies, I., 1992. *Ten Lectures on Wavelets*, vol. 61 of **CBMS-NSF Regional Conference Series in Applied Mathematics**, Society for Industrial & Applied Mathematics, Philadelphia, Penn.

Daubechies, I., Defrise, M. & de Mol, C., 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math.*, **57**(11), 1413–1457, doi: 10.1002/cpa.20042.

Davies, J. H., Gudmundsson, O. & Clayton, R. W., 1992. Spectra of mantle shear wave velocity structure, *Geophys. J. Int.*, **108**(3), 865–882, doi: 10.1111/j.1365–246X.1992.tb03476.x.

Debayle, E. & Sambridge, M., 2004. Inversion of massive surface wave data sets: Model construction and resolution assessment, *J. Geophys. Res.*, **109**, B02316, doi: 10.1029/2003JB002652.

Denison, D. G. T., Holmes, C. C., Mallick, B. K. & Smith, A. F. M., 2002. *Bayesian methods for nonlinear classification and regression*, John Wiley.

Deschamps, F. & Trampert, J., 2003. Mantle tomography and its relation to temperature and composition, *Phys. Earth Planet. Inter.*, **140**(4), 277–291, doi: 10.1016/j.pepi.2003.09.004.

Dobson, D. C. & Santosa, F., 1996. Recovery of blocky images from noisy and blurred data, *SIAM J. Appl. Math.*, **56**(4), 1181–1198.

Donoho, D. L., 2006. For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution, *Comm. Pure Appl. Math.*, **59**(6), 797–829, doi: 10.1002/cpa.20132.

Donoho, D. L. & Johnstone, I. M., 1994. Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, **81**(3), 425–455.

Donoho, D. L. & Johnstone, I. M., 1995. Adapting to unknown smoothness via wavelet shrinkage, *J. Acoust. Soc. Am.*, **90**(432).

Dziewoński, A. M., 1984. Mapping the lower mantle: Determination of lateral heterogeneity in *P* velocity up to degree and order 6, *J. Geophys. Res.*, **89**(B7), 5929–5952.

Easley, G., Lim, W. & Labate, D., 2008. Sparse directional image representations using the discrete shearlet transform, *Appl. Comput. Harmon. Anal.*, **25**, 25–46.

Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R., 2004. Least angle regression, *Ann. Statist.*, **32**(2), 407–499, doi: 10.1214/009053604000000067.

Ekström, G., Tromp, J. & Larson, E. W. F., 1997. Measurements and global models of surface wave propagation, *J. Geophys. Res.*, **102**(B4), 8137–8157.

Fernández, N. L. & Prestin, J., 2006. Interpolatory band-limited wavelet bases on the sphere, *Constr. Approx.*, **23**, 79–101, doi: 10.1007/s00365–005–0601–1.

Foufoula-Georgiou, E. & Kumar, P., eds., 1994. *Wavelets in Geophysics*, Academic Press, San Diego, Calif.

Freeden, W. & Michel, V., 1999. Constructive approximation and numerical methods in geodetic research today – an attempt at a categorization based on an uncertainty principle, *J. Geodesy*, **73**(9), 452–465.

Freeden, W. & Michel, V., 2004. Orthogonal zonal, tesseral and sectorial wavelets on the sphere for the analysis of satellite data, *Adv. Comput. Math.*, **21**(1–2), 181–217.

Freeden, W. & Michel, V., 2004. *Multiscale potential theory*, Birkhäuser, Boston, Mass.

Garcia, R. F., Chevrot, S. & Calvet, M., 2009. Statistical study of seismic heterogeneities at the base of the mantle from *PKP* differential traveltimes, *Geophys. J. Int.*, **179**(3), 1607–1616.

Gauch, H. G., 2003. *Scientific method in practice*, Cambridge Univ. Press, Cambridge, UK.

Gholami, A. & Siahkoohi, H. R., 2010. Regularization of linear and non-linear geophysical ill-posed problems with joint sparsity constraints, *Geophys. J. Int.*, **180**(2), 871–882, doi: 10.1111/j.1365–246X.2009.04453.x.

González, A., 2010. Measurement of areas on a sphere using Fibonacci and latitude-longitude lattices, *Math. Geosc.*, **42**(1), 49–64.

Gudmundsson, O., Davies, J. H. & Clayton, R. W., 1990. Stochastic analysis of global traveltime data: mantle heterogeneity and random errors in the ISC data, *Geophys. J. Int.*, **102**(1), 25–43.

Gurnis, M., 1986. Quantitative bounds on the size spectrum of isotopic heterogeneity within the mantle, *Nature*, **323**, 317–320, doi:10.1038/323317a0.

Hansen, P. C., 1992. Analysis of discrete ill-posed problems by means of the L-curve, *SIAM Rev.*, **34**(4), 561–580, doi: 10.1137/1034115.

Hedlin, M. A. H. & Shearer, P. M., 2000. An analysis of large-scale variations in small-scale mantle heterogeneity using Global Seismographic Network recordings of precursors to *PKP*, *J. Geophys. Res.*, **105**(B6), 13655.

Hemmat, A. A., Dehghan, M. A. & Skopina, M., 2005. Ridge wavelets on the ball, *J. Approx. Theory*, **136**(2), 129–139.

Hernlund, J. W. & Houser, C., 2008. On the statistical distribution of seismic velocities in Earth's deep mantle, *Earth Planet. Sci. Lett.*, **265**(3–4), 423–437.

Holschneider, M., Chambodut, A. & Mandea, M., 2003. From global to regional analysis of the magnetic field on the sphere using wavelet frames, *Phys. Earth Planet. Inter.*, **135**, 107–124.

Houser, C. & Williams, Q., 2009. The relative wavelengths of fast and slow velocity anomalies in the lower mantle: Contrary to the expectations of dynamics?, *Phys. Earth Planet. Inter.*, **176**(3–04), 187–197.

Hung, S.-H., Chen, W.-P., Chiao, L.-Y. & Tseng, T.-L., 2010. First multi-scale, finite-frequency tomography illuminates 3D anatomy of the Tibetan Plateau, *Geophys. Res. Lett.*, **37**, L06304, doi:10.1029/2009GL041875.

Jawerth, B. & Sweldens, W., 1994. An overview of wavelet-based multiresolution analyses, *SIAM Rev.*, **36**(3), 377–412.

Jensen, A. & la Cour-Harbo, A., 2001. *Ripples in Mathematics*, Springer, Berlin.

Jordan, T. H., Puster, P., Glatzmaier, G. A. & Tackley, P. J., 1993. Comparisons between seismic Earth structures and mantle flow models based on radial correlation functions, *Science*, **261**(5127), 1427–1431.

Kárason, H. & van der Hilst, R. D., 2000, Constraints on mantle convection from seismic tomography, in *The History and Dynamics of Global Plate Motions*, edited by M. A. Richards, R. G. Gordon, & R. D. van der Hilst, vol. 121 of **Geophysical Monograph**, Amer. Geophys. Union, Washington, D. C.

Klees, R. & Haagmans, R. H. N., eds., 2000. *Wavelets in the Geosciences*, vol. 90 of **Lecture Notes in Earth Sciences**, Springer, Berlin.

Komatitsch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation — I. Validation, *Geophys. J. Int.*, **149**, 390–412.

Labate, D., Lim, W.-Q., Kutyniok, G. & Weiss, G., 2005, Sparse multidimensional representation using shearlets, in *Wavelets XI*, edited by M. Papadakis, A. F. Laine, & M. A. Unser, vol. 5914, pp. 254–262, doi: 10.1117/12.613494, SPIE.

Lauritzen, P. H., Nair, R. D. & Ullrich, P. A., 2010. A conservative semi-Lagrangian multi-tracer transport scheme (CSLAM) on the cubed-sphere grid, *J. Comput. Phys.*, **229**(5), 1401–1424.

Lekić, V. & Romanowicz, B., 2011. Inferring upper-mantle structure by full waveform tomography with the spectral element method, *Geophys. J. Int.*, **185**(2), 799–831, doi: 10.1111/j.1365–246X.2011.04969.x.

Lessig, C. & Fiume, E., 2008. SOHO: Orthogonal and symmetric Haar wavelets on the sphere, *ACM Trans. Graph.*, **27**(1), 4, doi: 10.1145/1330511.1330515.

Loris, I., 2008. L1Packv2: A Mathematica package for minimizing an ℓ_1-penalized functional, *Comput. Phys. Comm.*, **179**, 895–902, doi: 10.1016/j.cpc.2008.07.010.

Loris, I., 2009. On the performance of algorithms for the minimization of ℓ_1-penalized functionals, *Inv. Probl.*, **25**, 035008, doi:10.1088/0266–5611/25/3/035008.

Loris, I., Nolet, G., Daubechies, I. & Dahlen, F. A., 2007. Tomographic inversion using ℓ_1-norm regularization of wavelet coefficients, *Geophys. J. Int.*, **170**(1), 359–370, doi: 10.1111/j.1365–246X.2007.03409.x.

Loris, I., Douma, H., Nolet, G., Daubechies, I. & Regone, C., 2010. Nonlinear regularization techniques for seismic tomography, *J. Comput. Phys.*, **229**(3), 890–905, doi: 10.1016/j.jcp.2009.10.020.

Mallat, S., 2008. *A Wavelet Tour of Signal Processing, The Sparse Way*, Academic Press, San Diego, Calif., 3rd edn.

Mallat, S. G., 1989. Multiresolution approximations and wavelet orthonormal bases of L2(R), *Trans. Am. Math. Soc.*, **315**(1), 69–87.

Margerin, L. & Nolet, G., 2003. Multiple scattering of high-frequency seismic waves in the deep Earth: *PKP* precursor analysis and inversion for mantle granularity, *J. Geophys. Res.*, **108**, 2514.

Masters, G., Laske, G., Bolton, H. & Dziewonski, A. M., 2000, The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: implications for chemical and thermal structure, in *Earth's deep interior. Mineral Physics and tomography from the atomic to the global scale*, edited by S. Karato, A. Forte, R. Liebermann, G. Masters, & L. Stixrude, vol. 117 of **Geophysical Monograph**, pp. 63–87, Amer. Geophys. Union, Washington, D. C.

McEwen, J. D., Hobson, M. P., Mortlock, D. J. & Lasenby, A. N., 2007. Fast directional continuous spherical wavelet transform algorithms, *IEEE Trans. Signal Process.*, **55**(2), 520–529.

Mégnin, 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**(11), 1299–1302, doi: 10.1029/97GL01256.

Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory*, vol. 45 of **International Geophysics Series**, Academic Press, San Diego, Calif., Rev. edn.

Montelli, R., Nolet, G., Dahlen, F. A., Masters, G., Engdahl, E. R. & Hung, S.-H., 2004. Global *P* and *PP* traveltime tomography: rays versus waves, *Geophys. J. Int.*, **158**, 637–654.

Montelli, R., Nolet, G., Dahlen, F. A. & Masters, G., 2006. A catalogue of deep mantle plumes: new results from finite-frequency tomography, *Geochem. Geophys. Geosys.*, **7**, Q11007, doi: 10.1029/2006GC001248.

Narcowich, F. J. & Ward, J. D., 1996. Nonstationary wavelets on the m-sphere for scattered data, *Appl. Comput. Harmon. Anal.*, **3**, 324–336.

Ni, S. & Helmberger, D. V., 2003. Seismological constraints on the South African superplume; could be the oldest distinct structure on Earth, *EPSL*, **206**(1–2), 119–131.

Nolet, G., ed., 1987. *Seismic Tomography*, Reidel, Hingham, MA.

Nolet, G., 2008. *A Breviary for Seismic Tomography*, Cambridge Univ. Press, Cambridge, UK.

Nolet, G. & Montelli, R., 2005. Optimal parametrization of tomographic models, *Geophys. J. Int.*, **161**(2), 365–372, doi: 10.1111/j.1365–246X.2005.02596.x.

Oliver, M. A., 2009. Special issue on applications of wavelets in the geosciences, *Math. Geosc.*, **41**(6), 609–610.

Passier, M. L. & Snieder, R. K., 1995. On the presence of intermediate-scale heterogeneities in the upper mantle, *Geophys. J. Int.*, **123**, 817–837.

Piazzoni, A. S., Steinle-Neumann, G., Bunge, H.-P. & Dolejš, D., 2007. A mineralogical model for density and elasticity of the Earths mantle, *Geochem. Geophys. Geosys.*, **8**, Q11010, doi: 10.1029/2007GC001697.

Piromallo, C., Vincent, A. P., Yuen, D. A. & Morelli, A., 2001. Dynamics of the transition zone under Europe inferred from wavelet cross-spectra of seismic tomography, *Phys. Earth Planet. Inter.*, **125**, 125–139.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P., 1992. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, Cambridge Univ. Press, New York, 2nd edn.

Pulliam, R. J., Vasco, D. W. & Johnson, L. R., 1993. Tomographic inversions for mantle P wave velocity structure based on the minimization of l2 and l1 norms of International Seismological Centre travel time residuals, *J. Geophys. Res.*, **98**, 699–734.

Puster, P., Jordan, T. H. & Hager, B. H., 1995. Characterization of mantle convection experiments using two-point correlation functions, *J. Geophys. Res.*, **100**(B4), 6351–6365.

Qin, Y., Capdeville, Y., Montagner, J. P., Boschi, L. & Becker, T. W., 2009. Reliability of mantle tomography models assessed by spectral element simulation, *Geophys. J. Int.*, **177**(1), 125–144.

Ritsema, J., 2005. Global seismic structure maps, *Geol. Soc. Am. Spec. Paper*, **388**, 11–18, doi: 10.1130/2005.2388(02).

Ritsema, J., van Heijst, H.-J. & Woodhouse, J. H., 1999. Complex shear wave velocity structure imaged beneath Africa and Iceland, *Science*, **286**, 1925–1928.

Ritsema, J., van Heijst, H. J. & Woodhouse, J. H., 2004. Global transition zone tomography, *J. Geophys. Res.*, **109**(B2), B02302, doi: 10.1029/2003JB002610.

Ritsema, J., McNamara, A. K. & Bull, A. L., 2007. Tomographic filtering of geodynamic models: Implications for model interpretation and large-scale mantle structure, *J. Geophys. Res.*, **112**, B01303, doi: 10.1029/2006JB004566.

Ritsema, J., A. Deuss, A., van Heijst, H. J. & Woodhouse, J. H., 2010. S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophys. J. Int.*, pp. doi: 10.1111/j.1365–246X.2010.04884.x.

Ronchi, C., Iacono, R. & Paolucci, P. S., 1996. The "Cubed Sphere": A new method for the solution of partial differential equations in spherical geometry, *J. Comput. Phys.*, **124**, 93–114, doi: 10.1006/jcph.1996.0047.

Rudin, L. I., Osher, S. & Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms, *Physica D*, **60**(1–4), 259–268, doi: 10.1016/0167–2789(92)90242–F.

Saltzer, R. L., van der Hilst, R. D. & Kárason, H., 2001. Comparing *p* and *s* wave heterogeneity in the mantle, *Geophys. Res. Lett.*, **28**(7), 1335–1338.

Sambridge, M. & Rawlinson, N., 2005, Seismic tomography with irregular meshes, in *Seismic Earth: Array Analysis of Broadband Seismograms*, edited by A. Levander & G. Nolet, vol. 157 of **Geophysical Monograph**, pp. 49–65, Amer. Geophys. Union.

Sambridge, M., Beghein, C., Simons, F. J. & Snieder, R., 2006. How do we understand and visualize uncertainty?, *The Leading Edge*, **25**(5), 542–546.

Schmidt, M., Han, S.-C., Kusche, J., Sanchez, L. & Shum, C. K., 2006. Regional high-resolution spatiotemporal gravity modeling from GRACE data using spherical wavelets, *Geophys. Res. Lett.*, **33**(8), L0840, doi: 10.1029/2005GL025509.

Schröder, P. & Sweldens, W., 1995. Spherical wavelets: Efficiently representing functions on the sphere, *Computer Graphics Proceedings (SIGGRAPH 95)*, pp. 161–172.

Schuberth, B. S. A., Bunge, H.-P. & Ritsema, J., 2009. Tomographic filtering of high-resolution mantle circulation models: Can seismic heterogeneity be explained by temperature alone?, *Geochem. Geophys. Geosys.*, **10**.

Shearer, P. M. & Earle, P. S., 2004. The global short-period wavefield modelled with a Monte Carlo seismic phonon method, *Geophys. J. Int.*, **158**(3), 1103–1117.

Simons, F. J., van der Hilst, R. D., Montagner, J.-P. & Zielhuis, A., 2002. Multimode Rayleigh wave inversion for heterogeneity and azimuthal anisotropy of the Australian upper mantle, *Geophys. J. Int.*, **151**(3), 738–754, doi: 10.1046/j.1365–246X.2002.01787.x.

Simons, F. J., Nolet, G., Georgief, P., Babcock, J. M., Regier, L. A. & Davis, R. E., 2009. On the potential of recording earthquakes for global seismic tomography by low-cost autonomous instruments in the oceans, *J. Geophys. Res.*, **114**, B05307, doi:10.1029/2008JB006088.

Spakman, W. & Bijwaard, H., 2001. Optimization of cell parameterization for tomographic inverse problems, *Pure Appl. Geophys.*, **158**, 1401–1423.

Starck, J. L., Moudden, Y., Abrial, P. & Nguyen, M., 2006. Wavelets, ridgelets and curvelets on the sphere, *Astron. Astroph.*, **446**, 1191–1204.

Strang, G. & Nguyen, T., 1997. *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, Mass., 2nd edn.

Swinbank, R. & Purser, R. J., 2006. Fibonacci grids: A novel approach to global modelling, *Quart .J. Royal Meteor. Soc.*, **132**(619), 1769–1793.

Tkalčić, H. & Romanowicz, B., 2002. Short scale heterogeneity in the lowermost mantle: insights from *pcp–p* and *scs-s*

data, *Earth Planet. Sci. Lett.*, **201**(1), 57–68, doi: 10.1016/S0012–821X(02)00657–X.

Trampert, J. & Snieder, R., 1996. Model estimations biased by truncated expansions: Possible artifacts in seismic tomography, *Science*, **271**(5253), 1257–1260, doi:10.1126/science.271.5253.1257.

Trampert, J. & van der Hilst, R. D., 2005, Towards a quantitative interpretation of global seismic tomography, in *Earth's Deep Mantle: Structure, Composition, and Evolution*, edited by R. D. van der Hilst, J. Bass, J. Matas, & J. Trampert, vol. 160 of **Geophysical Monograph**, pp. 47–62, Amer. Geophys. Union, Washington, D. C.

Trampert, J. & Woodhouse, J. H., 1995. Global phase-velocity maps of Love and Rayleigh-waves between 40 and 150 seconds, *Geophys. J. Int.*, **122**(2), 675–690.

Trampert, J. & Woodhouse, J. H., 1996. High resolution global phase velocity distributions, *Geophys. Res. Lett.*, **23**(1), 21–24.

Trampert, J. & Woodhouse, J. H., 2001. Assessment of global phase velocity models, *Geophys. J. Int.*, **144**(1), 165–174, doi: 10.1046/j.1365–246x.2001.00307.x.

van der Hilst, R. D. & Kárason, H., 1999. Compositional heterogeneity in the bottom 1000 kilometers of Earth's mantle: Toward a hybrid convection model, *Science*, **283**(5409), 1885–1888.

VanDecar, J. C. & Snieder, R., 1994. Obtaining smooth solutions to large, linear inverse problems, *Geophysics*, **59**(5), 818–829.

Vasco, D. W., Johnson, L. R., Pulliam, R. J. & Earle, P. S., 1994. Robust inversion of IASP91 travel time residuals for mantle *P* and *S* velocity structure, earthquake mislocations, and station corrections, *J. Geophys. Res.*, **99**, 13727–13755.

Vasco, D. W., Johnson, L. R. & Marques, O., 1999. Global Earth structure: Inference and assessment, *Geophys. J. Int.*, **137**(2), 381–407.

Vincent, H. T. & Simons, F. J., 2011. Son-O-Mermaid: A green instrument platform for the blue ocean, *Geophys. Res. Abstr.*, **13**, 12908.

Wang, Z. & Dahlen, F. A., 1995. Spherical-spline parameterization of three-dimensional Earth models, *Geophys. Res. Lett.*, **22**, 3099–3102.

Wang, Z., Tromp, J. & Ekström, G., 1998. Global and regional surface-wave inversions: A spherical-spline parameterization, *Geophys. Res. Lett.*, **25**(2), 207–210.

Wiaux, Y., Jacques, L. & Vandergheynst, P., 2005. Correspondence principle between spherical and Euclidean wavelets, *Astroph. J.*, **632**, 15–28, doi: 10.1086/432926.

Wiaux, Y., McEwen, J. D. & Vielva, P., 2007. Complex data processing: Fast wavelet analysis on the sphere, *J. Fourier Anal. Appl.*, **13**(4), 477–493, 10.1007/s00041–006–6917–9.

Woodhouse, J. H. & Dziewoński, A. M., 1984. Mapping the upper mantle: Three-dimensional modeling of Earth structure by inversion of seismic waveforms, *J. Geophys. Res.*, **89**(B7), 5953–5986.

Wysession, M. E., 1996. Large-scale structure at the core-mantle boundary from diffracted waves, *Nature*, **382**, 244–248.

Wysession, M. E., Langenhorst, A., Fouch, M. J., Fischer, K. M., Al-Eqabi, G. I., Shore, P. J. & Clarke, T. J., 1999. Lateral variations in compressional/shear velocities at the base of the mantle, *Science*, **284**, 120–125.

Yanovskaya, T. B. & Ditmar, P. G., 1990. Smoothness criteria in surface wave tomography, *Geophys. J. Int.*, **102**, 63–72.

Yuen, D. A., Vincent, A. P., Kido, M. & Vecsey, L., 2002. Geophysical applications of multidimensional filtering with wavelets, *Pure Appl. Geophys.*, **159**(10), 2285–2309.

Zhang, Y.-S. & Tanimoto, T., 1993. High-resolution global upper-mantle structure and plate-tectonics, *J. Geophys. Res.*, **98**(B6), 9793–9823.