

A method of spherical harmonic analysis in the geosciences via hierarchical Bayesian inference

J.B. Muir^{1,2} and H. Tkalčić²

¹*Seismological Laboratory, California Institute of Technology, Pasadena, CA 91125, USA. E-mail: jmuir@gps.caltech.edu*

²*Research School of Earth Sciences, Australian National University, Canberra ACT 0200, Australia*

Accepted 2015 August 27. Received 2015 August 26; in original form 2015 August 25

SUMMARY

The problem of decomposing irregular data on the sphere into a set of spherical harmonics is common in many fields of geosciences where it is necessary to build a quantitative understanding of a globally varying field. For example, in global seismology, a compressional or shear wave speed that emerges from tomographic images is used to interpret current state and composition of the mantle, and in geomagnetism, secular variation of magnetic field intensity measured at the surface is studied to better understand the changes in the Earth's core. Optimization methods are widely used for spherical harmonic analysis of irregular data, but they typically do not treat the dependence of the uncertainty estimates on the imposed regularization. This can cause significant difficulties in interpretation, especially when the best-fit model requires more variables as a result of underestimating data noise. Here, with the above limitations in mind, the problem of spherical harmonic expansion of irregular data is treated within the hierarchical Bayesian framework. The hierarchical approach significantly simplifies the problem by removing the need for regularization terms and user-supplied noise estimates. The use of the corrected Akaike Information Criterion for picking the optimal maximum degree of spherical harmonic expansion and the resulting spherical harmonic analyses are first illustrated on a noisy synthetic data set. Subsequently, the method is applied to two global data sets sensitive to the Earth's inner core and lowermost mantle, consisting of PKPab-df and PcP-P differential traveltime residuals relative to a spherically symmetric Earth model. The posterior probability distributions for each spherical harmonic coefficient are calculated via Markov Chain Monte Carlo sampling; the uncertainty obtained for the coefficients thus reflects the noise present in the real data and the imperfections in the spherical harmonic expansion.

Key words: Fourier analysis; Inverse theory; Spatial analysis; Probability distributions.

1 INTRODUCTION

A common problem in geophysics is the representation of data on the sphere in terms of spherical harmonics. The essence of this problem is that of the estimation of a continuous field, from which the data are sampled with error, using a finite basis function expansion to represent the field. Where data are regularly gridded, a spherical harmonic transform (analogous to the Fourier transform) exists. If data are sufficiently dense, irregular data may be interpolated to a form suitable for such techniques (Healy *et al.* 2003). However, for sparse data with a measurement density that is non-uniform, this can introduce significant errors, or restrict the analysis to low orders. Additionally, as geophysical data are invariably noisy, it is useful to determine bounds on the credible intervals of the spherical harmonic coefficients. The spherical harmonic transform is a linear operator, hence the classical method by which to invert for spherical harmonics is by one of the methods of regularized

least squares (Aster *et al.* 2011), the most common of which uses a single regularization parameter λ . The regularization parameter creates a trade-off between goodness of fit and model complexity. In the context of spherical harmonic analysis, large λ will penalize large spherical harmonic coefficients, generating a smoother, less complex, final spatial model of the data. The methods by which the regularization parameter(s) can be efficiently chosen to best balance model complexity and goodness of fit have been intensely studied, and there have been significant developments of the least-squares method to improve regularization and avoid spectral leakage due to model basis set truncation (Snieder & Trampert 1999). Alternatively, the spherical harmonic analysis problem is handled naturally and simply within the Bayesian framework, eliminating the need for regularization terms that complicate norm-minimization-based approaches. Here, we discuss a Markov Chain Monte Carlo (MCMC) implementation of hierarchical Bayesian inversion for the spherical harmonic coefficients (model parameters) and data noise

(a hyperparameter). This type of sampling is said to be ‘hierarchical’ as the standard deviation of an assumed Gaussian noise for each data point is represented by a free parameter. Treating the data noise as a hyperparameter in this fashion supplements the resulting model with its uncertainty.

Importantly, the fit to the data improves with increasing the number of spherical harmonic coefficients (i.e. in our case, too high degree expansion will cause overfitting of data and overestimated uncertainties), but the fit alone is not a sufficient criterion for the selection of the maximum degree of spherical harmonic expansion. Therefore, a criterion for an optimal model dimension selection, which here depends on the maximum degree of spherical harmonic expansion, has to be established. A transdimensional type of sampling has recently been applied in many inverse problems to determine optimal model selection, despite the high computational burden (Malinverno 2002; Sambridge *et al.* 2006; Dettmer *et al.* 2010; Tkalčić *et al.* 2013). Alternatively, an information criterion (IC) can be used at lower computational cost (Schwarz 1978; Pachhai *et al.* 2014). Here, we utilize the corrected Akaike information criterion (AICc) to choose an appropriate maximum degree of expansion.

2 THEORY

2.1 Spherical harmonic expansion

The spherical harmonics form an often used basis set for the description of physical data in geophysics. Spherical harmonics Y_l^m are the solutions to Laplace’s equation $\nabla^2 \Phi(\theta, \phi) = 0$ on the sphere:

$$Y_l^m = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos \theta) e^{im\phi}, \quad (1)$$

where P_l^m is an associated Legendre polynomial of appropriate degree (Press 2007, pp. 292–297) (there are many possible definitions for the spherical harmonics; we choose the one used by Press due to the ready availability of algorithms to handle this definition). The harmonic degree l and order m express the nodal structure of the spherical harmonic Y_l^m ; in particular a spherical harmonic Y_l^m has l nodal great circles for which the Y_l^m is zero. Any twice differentiable scalar function f on the sphere may be written in terms of spherical harmonics (Courant & Hilbert 1966, pp. 510–521). For some coefficients c_l^m

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l c_l^m Y_l^m(\theta, \phi). \quad (2)$$

When f is real, it is more convenient to work with real spherical harmonics, so that the spherical harmonic coefficients c_l^m are also real. We define these as follows:

$$Y_{lm} = \begin{cases} \frac{i}{\sqrt{2}}(Y_l^m - (-1)^m Y_l^{-m}) & \text{if } m < 0 \\ Y_l^m & \text{if } m = 0 \\ \frac{1}{\sqrt{2}}(Y_l^m + (-1)^m Y_l^{-m}) & \text{if } m > 0 \end{cases}. \quad (3)$$

In this case, f is expanded similarly for real c_{lm} :

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l c_{lm} Y_{lm}(\theta, \phi). \quad (4)$$

For a finite maximum degree l' , the spherical harmonic analysis of degree l' is defined as the set of coefficients c_{lm} that minimizes

$$\left\| f(\theta, \phi) - \sum_{l=0}^{l'} \sum_{m=-l}^l c_{lm} Y_{lm}(\theta, \phi) \right\| \quad (5)$$

in the L_2 norm (Schultz & Zhang 1994). The L_2 norm is appropriate in this sense if the data errors in f can be adequately described by a covariance matrix (Tarantola 2005). When f is defined or may be easily interpolated to a discrete, regular grid, then spherical harmonic analysis may be solved via a 2-D fast Fourier transform (FFT) approach (Healy *et al.* 2003). Where f is a function measured only at irregularly distributed discrete points (θ_i, ϕ_i) , then instead spherical harmonic analysis becomes a linear inverse problem—this is often the case in geophysics. We define the data vector $\mathbf{d}$, the resolution kernel G and the parameter vector $\mathbf{q}$ by

$$\mathbf{d} = \begin{pmatrix} f(\theta_1, \phi_1) \\ f(\theta_2, \phi_2) \\ \vdots \\ f(\theta_n, \phi_n) \end{pmatrix},$$

$$G = \begin{pmatrix} Y_{00}(\theta_1, \phi_1) & Y_{-11}(\theta_1, \phi_1) & \dots & Y_{l'l'}(\theta_1, \phi_1) \\ Y_{00}(\theta_2, \phi_2) & Y_{-11}(\theta_2, \phi_2) & \dots & Y_{l'l'}(\theta_2, \phi_2) \\ \vdots & \vdots & \ddots & \vdots \\ Y_{00}(\theta_n, \phi_n) & Y_{-11}(\theta_n, \phi_n) & \dots & Y_{l'l'}(\theta_n, \phi_n) \end{pmatrix},$$

$$\mathbf{q} = \begin{pmatrix} c_{00} \\ c_{-11} \\ \vdots \\ c_{l'l'} \end{pmatrix}. \quad (6)$$

Then the spherical harmonic analysis is the $\mathbf{q}$ that minimizes $\|G\mathbf{q} - \mathbf{d}\|$. Even if data are defined on a regular grid, the inverse problem formulation has the advantage over FFT-based approaches that it does not assume that f can be exactly represented by the spherical harmonics up to degree l' ; it permits the introduction of variance terms to prevent overfitting noisy data. Previous authors have solved this problem via various regularized or weighted least-squares techniques (Schultz & Zhang 1994; Sneeuw 1994), but these methods become increasingly intractable when data noise relationships become significant. A hierarchical Bayesian approach allows us to integrate out data noise as a hyperparameter, in order to find the correct posteriors for the spherical harmonics (which are not necessarily Gaussian even if the posteriors for fixed data noise are).

2.2 Hierarchical Bayesian formulation

Here, we use a nonlinear approach within the Bayesian framework to resolve the issues present in the classical approaches. We sample from the posterior probability distribution $P(\mathbf{q}|\mathbf{d})$ using the normal likelihood function (corresponding to a least-squares fit with a constant diagonal covariance matrix)

$$P(\mathbf{d}|\mathbf{q}) = \frac{1}{(\sqrt{2\pi}\sigma^2)^n} e^{-\frac{(G\mathbf{q}-\mathbf{d})^2}{2\sigma^2}}, \quad (7)$$

where n is the number of data points in $\mathbf{d}$. If the noise statistics in the data are known not to be normal, the likelihood function may be changed appropriately. Although we can often assume a normal distribution for the noise statistics of our data from maximum

uncertainty arguments (Malinverno & Parker 2006), it is in general impossible to confidently assign a specific value to the variance σ^2 of the data. However, the data variance is obviously a parameter of interest, as the size of the variance dictates the strength of the conclusions we are able to draw from the data. Therefore, we use a hierarchical Bayesian model in which σ is treated as a hyperparameter and is also sampled. Treating σ in this way allows the data noise to be properly accounted for, without requiring the user to make arbitrary assumptions as to the scale of the noise. This technique has been employed with success to perform regional tomography (Bodin *et al.* 2012a), calculate receiver functions from *P*-wave arrivals (Bodin *et al.* 2012b), deconvolution of receiver functions (Kolb & Lekić 2014), sampling of autoregression parameters (Dettmer *et al.* 2012) and identification of ultralow velocity zones (Pachhai *et al.* 2014). The ability to extract the posterior probability distribution of the unknown data error gives the Bayesian MCMC approach presented in this paper a significant advantage over existing classical methods of spherical harmonic analysis; it also makes the problem nonlinear in the σ parameter, motivating our use of MCMC sampling.

Assuming that the prior probability distributions for $\mathbf{q}$ and σ are independent, the full joint posterior distribution is

$$P(\mathbf{q}, \sigma | \mathbf{d}) \propto P(\mathbf{d} | \mathbf{q}, \sigma) P(\mathbf{q}) P(\sigma), \quad (8)$$

where $P(\mathbf{q})$ and $P(\sigma)$ are the prior probability distributions for $\mathbf{q}$ and σ , respectively. We assume for the priors that the c_{lm} are independently uniformly distributed on some ranges (a_{lm} , b_{lm}) so that

$$P(\mathbf{q}) = \prod_{l=0}^{\infty} \prod_{m=-l}^l \left(\frac{1}{b_{lm} - a_{lm}} \right). \quad (9)$$

For the data variance parameter σ , we assume a non-informative Jeffrey's prior (Sivia & Skilling 1996, p. 107)

$$P(\sigma) \propto \frac{1}{\sigma}. \quad (10)$$

In practice, this is implemented by assuming that $P(\log_{10}(\sigma))$ is uniformly distributed on some large range (a_σ , b_σ). The posterior is sampled by via 'MCMC'. To improve the convergence properties of the chain, an adaptive Metropolis–Hastings step method was used to tune the chain parameters to their optimal values—strictly speaking, this renders the chain non-Markovian but the chain has the correct mixing properties and converges to the posterior (Haario *et al.* 2001).

Once the MCMC chain has converged to the posterior, the marginal probability distributions for the c_{lm} coefficients are directly obtained as the histograms of the values that each coefficient takes on in the post-convergence chain. The marginal probability distributions of other auxiliary variables of interest (that are functions of $\mathbf{q}$), such as the power spectrum (Davies *et al.* 1992), defined as

$$Q_l = \frac{1}{2l+1} \sum_{m=-l}^l |c_{lm}|^2 \quad (11)$$

can be found simply by computing their value at each point in the chain and then constructing the resulting histograms.

2.3 Model space selection via the AICc

The final difficulty in representing data in terms of spherical harmonics is determining the appropriate maximum degree l' of the expansion. Truncation of the vector of model parameters to a particular degree imposes smoothness on the final spatial model, and so

determines what kind of physical inferences we can draw from the inverted model. In order to facilitate the choice of maximum degree, without arbitrary user input, we make use of the AICc. The AICc is a functional of the posterior probability distribution $P(\mathbf{q}, \sigma | \mathbf{d})$, which is defined by

$$\text{AICc}[P(\mathbf{d} | \mathbf{q}, \sigma)] = -\ln(\max[P(\mathbf{d} | \mathbf{q}, \sigma)]) + 2k + \frac{2k(k+1)}{n-k-1}, \quad (12)$$

where k is the number of free parameters in the model (i.e. the length of $\mathbf{q}$ plus 1) and the maximum is taken over all values of $\mathbf{q}$ and σ . The maximum expansion degree l' which best balances explanatory power with simplicity is the l' with the smallest AICc; the AICc can also be used to weight the contributions of inverted spatial models of different maximum degrees to produce a final 'multi-maximum-degree' map, although this is not explored in this paper. Burnham & Anderson (2004) recommend using the AICc over other information criteria such as the Bayesian Information Criterion (Schwarz 1978) for model selection with finite data sets where the true data generating process is not among the candidate models. This is in particular true for regression problems based on series expansion, which is the exact problem we are solving in spherical harmonic analysis (Yang 2005). As the AICc definition includes a maximum log likelihood term, it is not directly computable from the MCMC output. For this study, we take the mean values of the parameters from the MCMC as an estimate of the maximum likelihood parameters, and then use conjugate-gradient optimization to improve this estimate to obtain our final maximum log likelihood. Other criteria, such as the Deviance Information Criterion (DIC) and Bayesian Predictive Information Criterion (BPIC) are directly computable from the MCMC chain, however the DIC has problems with overfitting the data and the BPIC is an unnecessarily complicated calculation for this simple problem (Ando 2007).

At first impression, transdimensional MCMC would appear to be a good candidate for determining the maximum degree l' of spherical harmonic analysis, and it has the advantage of determining model complexity in a single inversion, rather than requiring multiple inversions for different spaces of models as is the case for IC approaches. However, transdimensional MCMC fails to perform adequately when changes in complexity cannot be easily handled by adding or deleting a single extra variable, as the likelihood of step acceptance when multiple variables are added or removed is extremely low. As adding an extra degree l to the expansion increases the number of variables by $2l+1$, spherical harmonic expansion is not a good candidate for transdimensional MCMC. In general, model parameterizations that are spatially localized, and can have members added and subtracted in a logical fashion (Voronoi cells, wavelets, etc.), can be treated using transdimensional MCMC, but global basis sets such as spherical harmonics behave poorly.

3 RESULTS

3.1 Synthetic test

We implemented a Bayesian spherical harmonic analysis program using the publicly available PyMC v2.3 (Python Monte Carlo) module (Patil *et al.* 2010). This module contains several built in convergence checking functions that make assessing the progress of the MCMC chain easy.

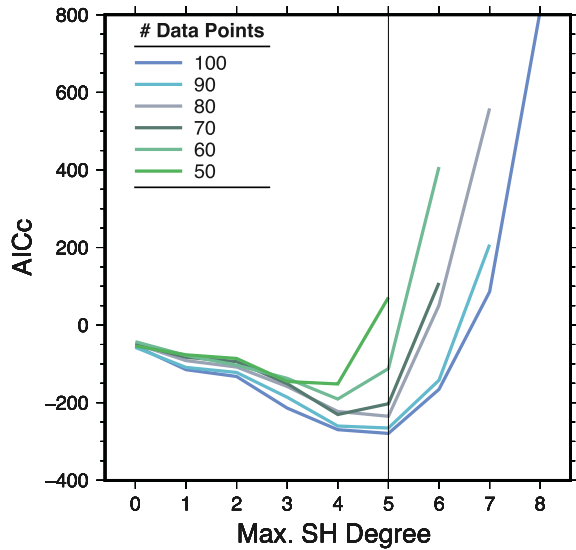


Figure 1. The AICcs of the inversions of the test model in Fig. 2 for different maximum degrees of expansion and number of data points are shown. A degree 5 expansion is clearly the minimum AICc when the number of data points is greater than 80.

We first tested the algorithm on a Gaussian random field, with a power spectral density defined by

$$Q_l = \frac{1}{(l+1)\sqrt{2\pi}} e^{\frac{(\ln(l+1)-1)^2}{2}}. \quad (13)$$

As a test of the AICc as a technique for determining the optimal complexity of the inversion (i.e. the dimension of the model), we generated data up to spherical harmonic degree 5. We sampled the input field at the locations of randomly chosen long-period global seismic network seismometers, predominantly distributed in the northern hemisphere, and added 5 per cent Gaussian noise. For the inversion, we assumed that all c_{lm} fell in the range $(-10, 10)$ and that $\log_{10}(\sigma)$ fell in the range $(-6, 0)$. The MCMC chain was run for 2.001×10^6 samples. The first 10^6 were discarded as a burn-in and the resulting chain sampled every 10^3 iterations. This resulted in 10^3 samples of the posterior distributions of the c_{lm} , which were confirmed to be statistically independent via autocorrelation plots. The runs were performed for maximum expansion degrees 0–8, and numbers of data points ranging from 50–100; the resulting plot of the AICc is shown in Fig. 1. The degree of the input field (degree 5) is clearly recovered as the minimum of the AICc when the number of data points is greater than 80, whereas a less complex model is chosen when the data are fewer. Therefore, the degree 5 is best supported by the data whereas for fewer data points, the AICc supports a less complex model, as the data does not produce as stringent requirements.

Fig. 2 shows the comparison between two different recovery tests for the input test function f from eq. (13) and 100 data points that correspond to the locations of long period seismometers (Fig. 2a). The recovery from a regularized least-squares inversion is shown in Fig. 2(b), whereas that of the Bayesian spherical harmonic analysis program is shown using the inverted mean map (Fig. 2c) and standard deviation map (Fig. 2d) constructed via MCMC chain at each geographic location. The inversion from a regularized least-squares inversion is performed using regularization parameter found via l-curve curvature maximization (Hansen 2000). In this approach the l-curve, showing the trade-off between model fit $\|Gq - d\|$ and model size $\|q\|$ is formed on a log–log plot, parameterized by the

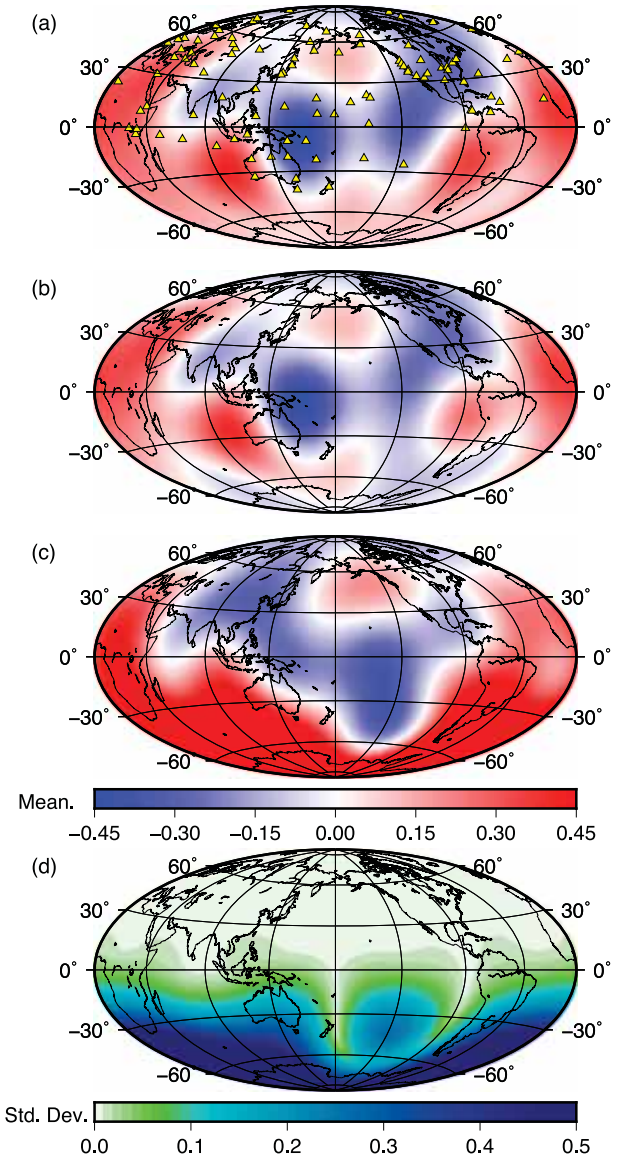


Figure 2. (a) The distribution of the sampling points of the input test model is shown by yellow triangles; these correspond to the locations of 100 long-period seismometers. The input test model (a random Gaussian field of maximum degree 5) is underlaid. (b) The reconstruction of the model via regularized least-squares inversion. (c) The reconstruction of the model via the Bayesian inversion technique. (d) Standard deviation of the posterior distribution of the model parameters as a function of coordinates; this reflects the uncertainty in the inversion, taking into account correlations between different model parameters.

regularization parameter λ . The λ chosen for the final inversion is taken from the corner of this l-curve—the corner may be found by maximizing the l-curve curvature. Comparison between Figs 2(a) and (b) shows that the pattern of the input model is generally well recovered in the northern hemisphere but somewhat misrepresents the input model in the southern hemisphere where data is scarce.

The model recovered through the hierarchical Bayesian inversion (Fig. 2c) is a mean over all post-burn solutions and thus the input velocity heterogeneity pattern is somewhat smoothed out, however the solution is accompanied by the standard deviation map (Fig. 2d). In addition, the correspondence between the true Figs 2(a) and (c) is significantly better (in at least a qualitative sense) in the southern hemisphere. The true value of the Bayesian approach of our method

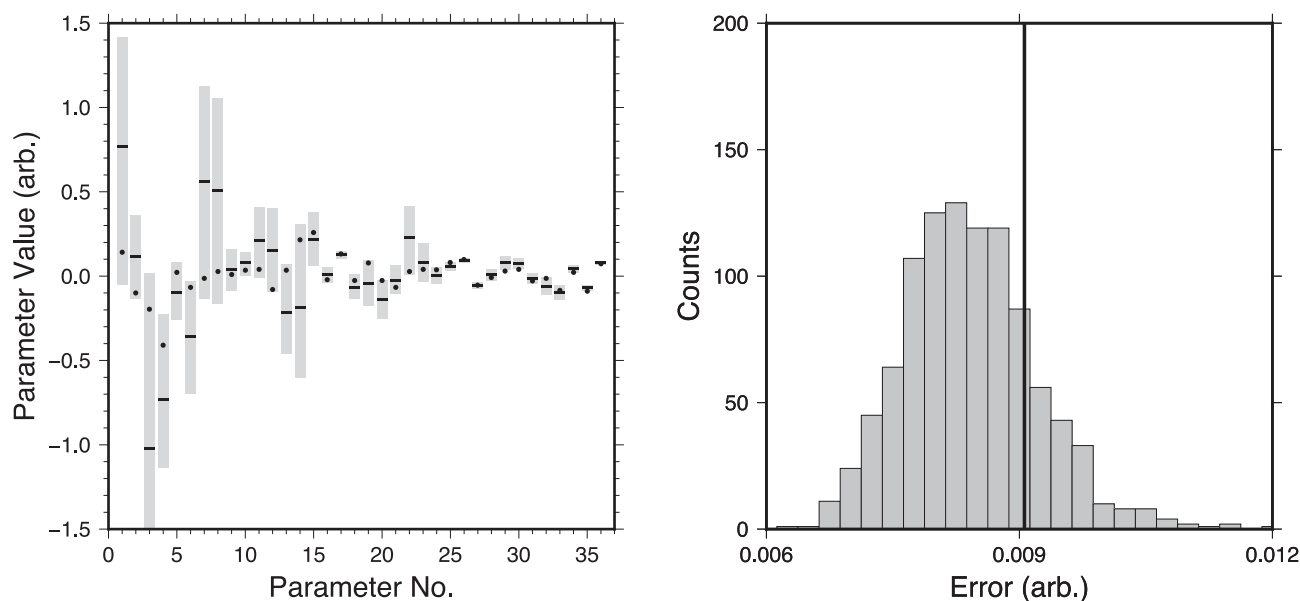


Figure 3. The means (black bars) and 95 per cent credible intervals (light grey boxes) are shown at left for the 36 c_{lm} parameters corresponding to a degree 5 spherical harmonic expansion for the synthetic experiment shown in Fig. 2. The true values of the c_{lm} corresponding to the test model are shown by black circles. At right, the histogram (trending towards the unnormalized posterior probability distribution) of the data error σ , as found by the hierarchical Bayesian inversion algorithm, is shown. The actual value of the input data error σ (5 per cent Gaussian noise) is shown by the vertical black line.

compared to approaches such as the regularized least-squares inversion is that, in addition to simply inverting for the c_{lm} parameters, the full posterior probability distributions for these parameters, and the data noise, are computed. A summary of this information showing the means and 95 per cent credible intervals is shown in Fig. 3. It is evident from this plot that the lower degree parameters are more poorly constrained by the available data, but despite this in all cases the mean lies close to the true value of the parameter.

3.2 Application to global data sets sensitive to the lowermost mantle and inner core

All data analysis begins by visual inspection. In the spatial domain, decomposition of the data into a basis set representation such as spherical harmonics is useful to build intuition about the behaviour of the data, before expensive procedures such as seismic tomography are taken out. As an example, we perform a spherical harmonic Bayesian hierarchical inversion of PKPab-df and PcP-P differential traveltime residuals using the AICc model dimension selection criterion (Fig. 4). We calculate differential traveltime residuals as $\Delta t = \text{observed} - \text{theoretical differential traveltimes}$, where the theoretical differential traveltime is calculated using spherically symmetric ak135 reference model of the Earth (Kennett *et al.* 1995).

The residuals are projected on to the bottoming point of the PKPdf and PcP rays respectively (Fig. 5, top row). Projecting to the bottoming point implicitly assumes that the traveltime anomaly is concentrated in the inner core for PKPab-df (as the PKPdf leg bottoms in the inner core) and the lowermost mantle for PcP-P (as PcP reflects off the core–mantle boundary); we have chosen to make this assumption for the raw traveltime data to remove the ambiguity in assigning anomalies along the ray paths for the sake of demonstrating how the technique performs. These data sets are a part of larger data sets that were used to constrain P -wave velocity in the lowermost mantle, in conjunction with other deep earth differential phases such as PKPbc-df (Young *et al.* 2013).

In practice, this assumption would be inadequate for PKPab-df differential traveltime data due to significant differences between the ray paths of PKPdf and PKPab in the mantle. Nonetheless, it is interesting to see what this assumption yields for inner core structure.

In our example case, the optimal maximum expansion degree l' is not known a priori. Therefore, we conduct an initial assessment using the AICc. For $0 \leq l' \leq 12$, the AICc is calculated by running the MCMC chain for 10^6 samples with a burn in of 2×10^5 samples, saving the chain every 200 samples. The maximum likelihood parameters from the resulting chains are taken as an estimate of the global maximum likelihood parameters, and the AICcs are calculated. Fig. 4 shows the results of this analysis; $l' = 10$ and $l' = 7$ are revealed as the maximum expansion degrees best supported by the data for the PKPab-df and PcP-P data sets—that is, $l' = 10$ and 7 offer the best trade-off between the complexity of the expansion and the data fit. We subsequently perform a final run with the optimal l' , with 2.001×10^6 total samples, a 10^6 sample burn in period, saving the results every 10^3 samples. The means of the parameters are used to reconstruct the final expansion in Fig. 5. The final expansion for each data set is shown for the strongest harmonic degrees (1, 2 and 3), as the bottom of Fig. 5 indicates.

For the PKPab-df data set, the recovered degree 1 is not strong (Fig. 5, left column), and the resulting velocity distribution is characterized by the north-south pattern. The recovered degrees 2 and 3 are strong (Fig. 5, left column). All this is different from the observed east–west dichotomy from the PKPbc-df data set, according to which the quasi-eastern hemisphere is faster than the quasi-western hemisphere (Tanaka & Hamaguchi 1997; Wen & Niu 2002). This likely confirms that the effects from the lowermost mantle are too strong for our assumption to be valid, i.e. that the PKPab-df data set without mantle correction cannot be used to recover inner core structure (Tkalčić *et al.* 2002).

For the PcP-P data set, degree 1–3 structure is strong (Fig. 5, right column). It is interesting to compare the resulting expansions with the lowermost mantle P -wave velocity model in which the

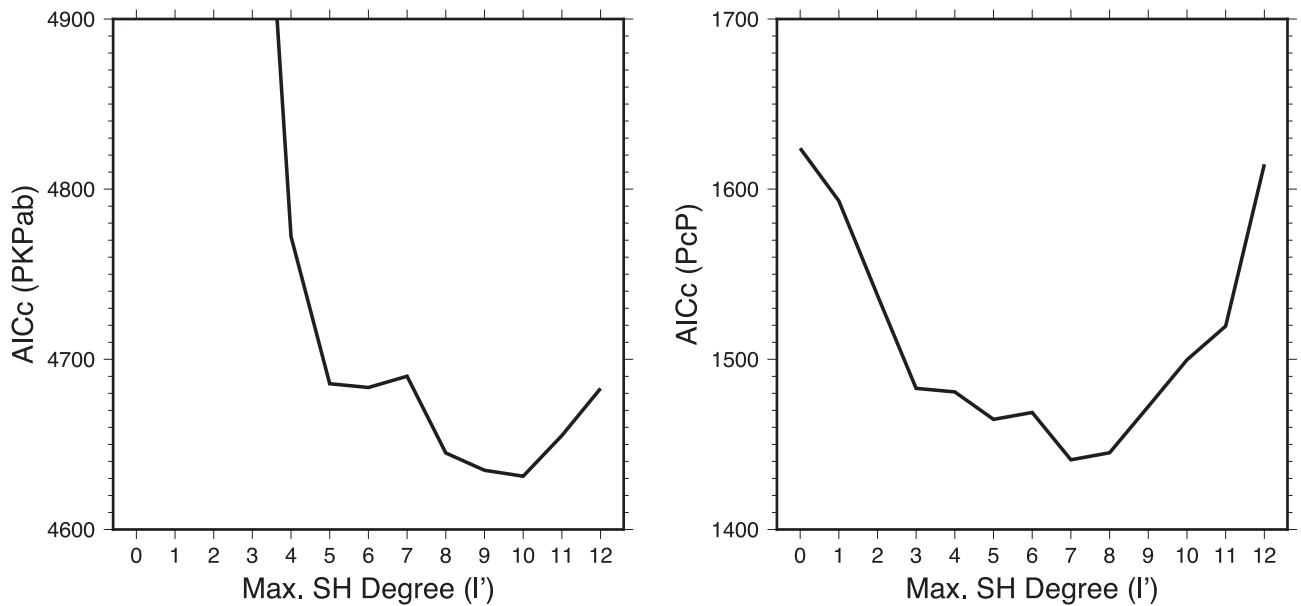


Figure 4. The corrected Akaike Information Criterion (AICc eq. 12) for expansions of maximum degree l' from 0 to 12 is shown for a data set consisting of PKPab-df (left) and PcP-P (right) differential traveltime residuals relative to the ak135 1-D seismic velocity model. $l' = 10$ is shown to be the model with the best support given the available data for the PKPab-df data set, and $l' = 7$ is the model with best support for the PcP-P data set.

same data set was used (fig. 9 of Young *et al.* 2013). It can be seen that the degree 2 of this data set significantly contributes to the obtained P -wave velocity map of the lowermost mantle, particularly in the northern hemisphere where the ray path coverage is abundant. Asia is dominated by fast velocity and the North Atlantic and North America are slow. That we recover predominantly degree 1–3 structure, despite allowing for much higher spherical harmonics, and without using damping, may be interpreted as an evidence of the strongly persistent low degree structure in the lowermost mantle (e.g. Dziewonski *et al.* 2010) keeping in mind that the PcP-P data set analysed here is a subset of a larger data set. We should be mindful however, with reference to the penultimate row of Fig. 5, that the uncertainty for these data sets is concentrated in the southern hemisphere as a result of low data coverage, which tempers the robustness of our inference of low degree structure in the southern hemisphere.

As the spatial distribution of data in our examples is not a result of a physical process that can be approximated by spherical harmonics, the values of individual coefficients have little meaning and are thus not computed. However, the length scale behaviour of the data set is useful to know. Hence the power spectra of the data are shown in the final row of Fig. 5. They are produced by calculating the P_l values for each iteration of the MCMC chain and finding the means and 95 per cent credible interval bounds. The power spectrum shows the strong degree 1–3 structure in agreement with Becker & Boschi (2002), here reflected in the raw traveltime data. The large amplitudes below Brazil (for PKPab-df) and the south Pacific (for PcP-P), where data coverage is sparse, show that for realistic data sets care must be taken in regions of poor coverage when making inferences based on spherical harmonic expansions; for example our inferences could be weighted by the variance of the computed model at each spatial location. This effect is a consequence of the global nature of the spherical harmonic basis set, as illustrated in the degree 1–3 components shown in Fig. 5. As shown in Fig. 5, it is easy, using the MCMC chains, to compute a spatial map of uncertainty that accompanies in the inversion results and allows us

to assess the accuracy of the inversion in regions of poor coverage. This is a significant advantage of this MCMC method for problems where the data coverage is uneven.

4 CONCLUSION

We have applied the techniques of hierarchical Bayesian inference to the problem of spherical harmonic analysis of irregular data. This improves both the utility and ease of use of spherical harmonic analysis by removing the need to set regularization parameters via trial and error and allowing immediate calculation of parameter error estimates and ancillary variables such as degree power. We have illustrated the use of the AICc on realistic seismological data sets as an appropriate method for choosing between expansions of different maximum degree l' , as illustrated in our synthetic test. This method is a useful utility for a compact representation of global data sets where it has the advantage of clearly highlighting the length scales relevant to the data. It also significantly reduces the likelihood of user error when performing these analyses. For very large data sets, the performance of this method may be improved by the use of more advanced Monte Carlo techniques than adaptive Metropolis Hastings, such as Hamiltonian Monte Carlo (Neal 2011); likewise for small data sets, it may be more efficient to use a combination of the Cholesky decomposition to diagonalize the model covariance matrix and Gibbs sampling to sample from the posterior distribution (Bennett 1975; Neal 2011). However, for data sets on the order of a few thousand points, this method is simple and fast. This paper has dealt exclusively with the spherical harmonic analysis problem, and not other methods for global estimation of a spherical field such as kriging, however, elements of the discussion such as the hierarchical Bayesian formulation and use of the AICc are directly transferable to other basis-set-expansion estimation problems, including regional spatial analyses using wavelets and other specialized functions.

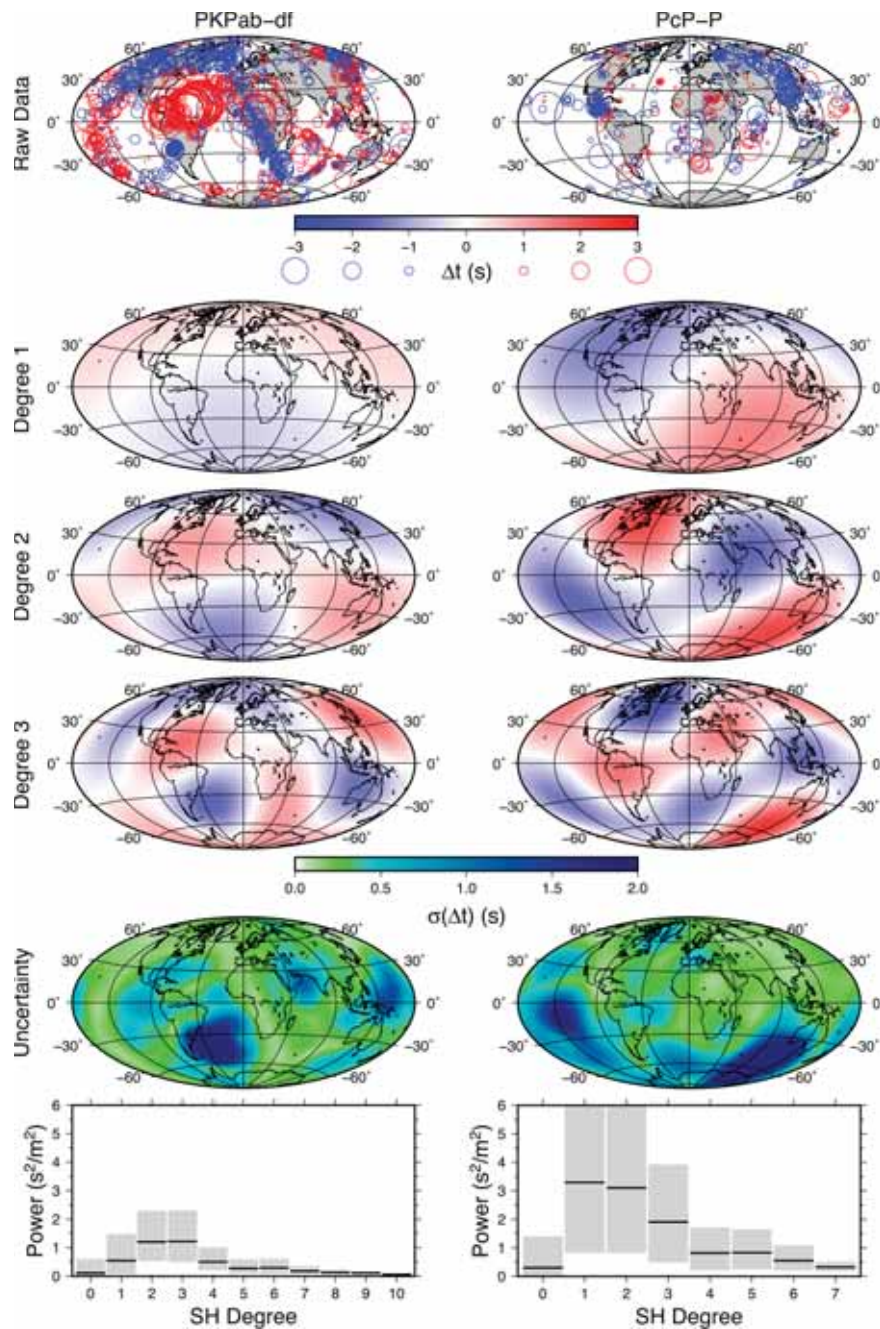


Figure 5. Measured residuals Δt (relative to the ak135 reference Earth model) are shown at top for PKPab-df and PcP-P differential traveltimes (Young *et al.* 2013). The circles representing the residuals are centred on the PKPab and PcP phase bottoming points. The data are expanded to the optimal degrees predicted by the AICs in Fig. 4, and the total uncertainty for the spherical harmonic analysis is shown. The expectation value of degree power P_l (black lines) and 95 per cent credible interval for the degree power (light grey boxes) have been calculated via the MCMC chain of the parameters obtained for each data set. The power in each degree is shown, up to $l' = 10$ for the PKPab-df data, and up to $l' = 7$ for the PcP-P data.

ACKNOWLEDGEMENTS

Jack Muir acknowledges the financial support of the Research School of Physics at the Australian National University, and the General Sir John Monash Foundation during this project.

REFERENCES

Ando, T., 2007. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models, *Biometrika*, **94**(2), 443–458.

Aster, R.C., Borchers, B. & Thurber, C.H., 2011. *Parameter Estimation and Inverse Problems*, Academic Press.

Becker, T.W. & Boschi, L., 2002. A comparison of tomographic and geodynamic mantle models, *Geochim. Geophys. Geosyst.*, **3**(1), 1003, doi:10.1029/2001GC000168.

Bennett, C.H., 1975. Mass tensor molecular dynamics, *J. Comput. Phys.*, **19**(3), 267–279.

Bodin, T., Sambridge, M., Rawlinson, N. & Arroucau, P., 2012a. Transdimensional tomography with unknown data noise, *Geophys. J. Int.*, **189**(3), 1536–1556.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012b. Transdimensional inversion of receiver

- functions and surface wave dispersion, *J. geophys. Res.*, **117**(B2), doi:10.1029/2011JB008560.
- Burnham, K.P. & Anderson, D.R., 2004. Multimodel inference understanding AIC and BIC in model selection, *Sociol. Method Res.*, **33**(2), 261–304.
- Courant, R. & Hilbert, D., 1966. *Methods of Mathematical Physics*, Vol. 1, Cambridge Univ. Press Archive.
- Davies, J.H., Gudmundsson, O. & Clayton, R., 1992. Spectra of mantle shear wave velocity structure, *Geophys. J. Int.*, **108**(3), 865–882.
- Dettmer, J., Dosso, S.E. & Osler, J.C., 2010. Bayesian evidence computation for model selection in non-linear geoacoustic inference problems, *J. acoust. Soc. Am.*, **128**(6), 3406–3415.
- Dettmer, J., Molnar, S., Steininger, G., Dosso, S.E. & Cassidy, J.F., 2012. Trans-dimensional inversion of microtremor array dispersion data with hierarchical autoregressive error models, *Geophys. J. Int.*, **188**(2), 719–734.
- Dziewonski, A.M., Lekić, V. & Romanowicz, B.A., 2010. Mantle Anchor Structure: An argument for bottom up tectonics, *Earth. planet. Sci. Lett.*, **299**(1), 69–79.
- Haario, H., Saksman, E. & Tamminen, J., 2001. An adaptive Metropolis algorithm, *Bernoulli*, **7**(2), 223–242.
- Hansen, P.C., 2000. The L-curve and its use in the numerical treatment of inverse problems, in *Computational Inverse Problems in Electrocardiology*, *Advances in Computational Bioengineering*, pp. 119–142, ed. Johnston, P., WIT Press.
- Healy, D., Jr, Rockmore, D.N., Kostelec, P.J. & Moore, S., 2003. FFTs for the 2-sphere—improvements and variations, *J. Fourier Anal. Appl.*, **9**(4), 341–385.
- Kennett, B.L.N., Engdahl, E. & Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.*, **122**(1), 108–124.
- Kolb, J.M. & Lekić, V., 2014. Receiver function deconvolution using trans-dimensional hierarchical Bayesian inference, *Geophys. J. Int.*, **197**(3), 1719–1735.
- Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.*, **151**(3), 675–688.
- Malinverno, A. & Parker, R.L., 2006. Two ways to quantify uncertainty in geophysical inverse problems, *Geophysics*, **71**(3), W15–W27.
- Neal, R., 2011. MCMC using Hamiltonian dynamics, in *Handbook of Markov Chain Monte Carlo*, Vol. 2, pp. 113–162, eds Brooks, S., Gelman, A., Jones, G. & Meng, X.-L., Chapman & Hall.
- Pachhai, S., Tkalčić, H. & Dettmer, J., 2014. Bayesian inference for ultralow velocity zones in the Earth's lowermost mantle: complex ULVZ beneath the east of the Philippines, *J. geophys. Res.*, **119**(11), 8346–8365.
- Patil, A., Huard, D. & Fonnesbeck, C.J., 2010. PyMC: Bayesian stochastic modelling in Python, *J. Stat. Soft.*, **35**(4), 1–81.
- Press, W.H., 2007. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, Cambridge Univ. Press.
- Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2006. Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.*, **167**, 528–542.
- Schultz, A. & Zhang, T.S., 1994. Regularized spherical harmonic analysis and the 3-D electromagnetic response of the Earth, *Geophys. J. Int.*, **116**(1), 141–156.
- Schwarz, G., 1978. Estimating the dimension of a model, *Ann. Stat.*, **6**(2), 461–464.
- Sivia, D. & Skilling, J., 1996. *Data Analysis: A Bayesian Tutorial*, Oxford University Press.
- Sneeuw, N., 1994. Global spherical harmonic analysis by least-squares and numerical quadrature methods in historical perspective, *Geophys. J. Int.*, **118**(3), 707–716.
- Snieder, R. & Trampert, J., 1999. *Inverse Problems in Geophysics*, Springer.
- Tanaka, S. & Hamaguchi, H., 1997. Degree one heterogeneity and hemispherical variation of anisotropy in the inner core from PKP (BC)–PKP (DF) times, *J. geophys. Res.*, **102**(B2), 2925–2938.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, 2nd edn, SIAM.
- Tkalčić, H., Romanowicz, B. & Houy, N., 2002. Constraints on D' structure using PKP (AB–DF), PKP (BC–DF) and PcP–P traveltime data from broad-band records, *Geophys. J. Int.*, **149**(3), 599–616.
- Tkalčić, H., Young, M., Bodin, T., Ngo, S. & Sambridge, M., 2013. The shuffling rotation of the Earth's inner core revealed by earthquake doublets, *Nature Geoscience*, **6**, 497–502.
- Wen, L. & Niu, F., 2002. Seismic velocity and attenuation structures in the top of the Earth's inner core, *J. geophys. Res.*, **107**(B11), ESE 2-1–ESE 2-13.
- Yang, Y., 2005. Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation, *Biometrika*, **92**(4), 937–950.
- Young, M., Tkalčić, H., Bodin, T. & Sambridge, M., 2013. Global P wave tomography of Earth's lowermost mantle from partition modeling, *J. geophys. Res.*, **118**(10), 5467–5486.