Geophysical Journal International

Geophys. J. Int. (2013) **192,** 676–680 doi: 10.1093/gji/ggs057

EXPRESS LETTER

Resolution tests revisited: the power of random numbers

Jeannot Trampert,[1] Andreas Fichtner[1] and Jeroen Ritsema[2]

[1]*Department of Earth Sciences, Utrecht University, P.O. Box* 80021, *NL*-3508 *TA, Utrecht, The Netherlands. E-mail: jeannot@geo.uu.nl*
[2]*Department of Earth and Environmental Sciences, University of Michigan,* 2534 *C. C. Little Building, 1100 North University Ave., Ann Arbor, MI,* 48109-1005*, USA*

Accepted 2012 November 5. Received 2012 October 29; in original form 2012 September 6

SUMMARY

We propose a simple method where the inversion of synthetic data, corresponding to a zero-mean random input vector, is used to infer the average horizontal and vertical resolution lengths of tomographic models. The method works well if the resolution operator has a diagonally dominant structure. This assumption, although often verified in seismic tomography, can be tested by simply cross-correlating the input with the output of the synthetic simulation. The method is as efficient as a single checkerboard test, but reveals more easily interpretable information.

Key words: Inverse theory; Seismic tomography.

1 INTRODUCTION

Most geophysical inverse problems rely on iterative linear model updates, where iterations proceed until a suitable convergence criterion is satisfied. While many inverse methods exist (e.g. Parker 1994; Tarantola 2005), usually little attention is devoted to the assessment of the quality of the solution. This assessment requires us to evaluate the contributions from two competing terms. One of these terms involves the resolution which is a clearly defined operator for a strictly linear problem (solved iteratively or not), and a formal analysis allows us to write (e.g. Snieder & Trampert 1999) that

$$\tilde{\mathbf{m}} - \mathbf{m} = (\mathbf{I} - \mathbf{R})(\mathbf{m_0} - \mathbf{m}) + \mathbf{L}\epsilon\,, \qquad (1)$$

where $\mathbf{G}$ is the linear forward operator relating the data $\mathbf{d}$ to the target (or true) model $\mathbf{m}$ via $\mathbf{d} = \mathbf{Gm} + \epsilon$, ϵ representing unmodelled signal. $\tilde{\mathbf{m}} = \mathbf{m_0} + \mathbf{L}(\mathbf{d} - \mathbf{Gm_0})$ is the estimated model obtained by applying a linear inverse operator $\mathbf{L}$ to the data, and $\mathbf{m_0}$ is some optional starting model. The resolution operator is defined by $\mathbf{R} = \mathbf{LG}$ and $\mathbf{I}$ is the identity matrix. Expression (1) shows indeed that the estimated model deviates from the target model via two terms: a contribution from an imperfect resolution (deviation from the identity matrix) and a contribution from unexplained data due to an imperfect modelling or data uncertainties.

The concept of resolution cannot easily be extended to non-linear problems. However, in case that an iterative Gauss–Newton method is used (e.g. eq. 3.51 in Tarantola 2005), at convergence, say step n, the estimated model $\mathbf{m}_n$ is linearly close to the target model $\mathbf{m}$ similar to the expression in eq. (1). An interesting consequence is that for Newton-type schemes, only the linearized resolution at step n needs to be known for the assessment stage. For other non-linear inverse schemes, the concept of linearized resolution might not be so useful. For instance in conjugate gradient-based methods, the Hessian should be used instead (e.g. Fichtner & Trampert 2011).

In most realistic inverse problems, the evaluation of the resolution operator is numerically more demanding than the error propagation term. We propose here a method which efficiently provides its essential characteristics rather than the complete matrix.

The calculation of the resolution matrix corresponding to small linear systems is computationally simple. The difficulty arises when the system becomes large [$O(10^6)$ data and unknowns]. Although Soldati *et al.* (2006) showed that supercomputers can handle such calculations, it remains a challenge to visualize and interpret a matrix with more than a million-square entries. Usually, the interpretation is done rowwise or columnwise. A row of the resolution matrix informs us on how an estimated parameter is linearly related to the target parameters. This is known as the averaging kernel and was first introduced by Backus & Gilbert (1968). The column of the resolution matrix describes the blurring of an input delta-function by the inverse operator. In optics, this is known as the point-spread function and in seismic tomography as the spike-test (Spakman 1991). While it does not matter of course if one analyses the resolution matrix by row or by column, for high dimensional problems, it is still a challenge to inspect as many rows or columns as there are model parameters. Therefore synthetic resolution tests are very popular. The idea is to invert synthetic data corresponding to a few known input structures and compare the results to these input models. The goal is to learn something about resolution, but Lévêque *et al.* (1993) have shown that it is not easy to interpret linear combinations of the columns.

The essential information of the resolution matrix of large seismic tomography studies, is found along diagonals (e.g. Soldati *et al.*

2006) rather than rows and columns. The main diagonal indicates how well we recover the amplitude of the individual model parameters. The next diagonal informs us on how a model parameter is linearly related to its adjacent and so on. These secondary diagonals are usually small, if data coverage is good, unless we arrive at the diagonal representing the same parameter but at a different depth level. We suggest to apply matrix probing techniques to infer the characteristics of the essential diagonals of the resolution matrix. We will show below that the main diagonal informs us on the average horizontal resolution, whereas the other diagonals on the average vertical resolution.

Matrix probing is a collective term for randomized algorithms designed to analyse matrices. They are most useful when explicit representations of the matrix are unknown or too expensive to be computed, but the application of the matrix to a vector is relatively straightforward. A low rank approximation of a matrix can efficiently be computed by applying it to random vectors (Halko *et al.* 2011). These techniques share many ideas with compressive sampling algorithms (e.g. Candès & Wakin 2008). If a matrix (or more interestingly its inverse) can be approximated by a linear combination of a small number of given basis matrices, the corresponding coefficients can easily be found by multiplying the matrix with a few random vectors (Chiu & Demanet 2012; Demanet *et al.* 2012). An (2012) used a similar idea and parametrized the resolution matrix rowwise as Gaussians. The application to several random vectors efficiently provides the coefficients of the Gaussians. The diagonal of an unknown matrix can also be found by applying it to successive random vectors (Bekas *et al.* 2007). This technique has successfully been used by MacCarthy *et al.* (2011) to find the diagonal of the resolution matrix and is a generalization of the method to estimate the trace of an unknown matrix (Hutchinson 1990; Avron & Toledo 2011). We will use a variant of the latter to find the trace of the resolution matrix with one random probe. Too often authors cite computational limitations as the reason for not performing a resolution analysis. Our emphasis is therefore on a single matrix probe for reasons of computational efficiency. Of course if more probes are used, a better estimate will be found (Avron & Toledo 2011).

2 METHOD

The trace of a matrix $\mathbf{R}$ may be found by calculating the average of successive evaluations of the quadratic form $\mathbf{x}^t\mathbf{R}\mathbf{x}$, where $\mathbf{x}$ are zero-mean random vectors:

$$\mathbf{x}^t\mathbf{R}\mathbf{x} = \sum_i R_{ii}x_i^2 + \sum_i \sum_{j\neq i} x_i R_{ij} x_j. \tag{2}$$

Hutchinson (1990) showed that by applying independent and identically distributed Rademacher vectors (i.e. vectors whose entries are independent random signs) and averaging the corresponding quadratic forms, the second term on the right-hand side of (2) will vanish. Convergence can be slow though and Avron & Toledo (2011) determined the number of probes needed to obtain an estimation of the trace within a specified error. Clearly, for a general matrix, a single probe will not be sufficient, although x_i and x_j are two independent random numbers drawn from a zero-mean distribution. In the case of seismic tomography, the resolution matrix usually has fast decaying off-diagonal elements (e.g. Fig. 1 and figures in Soldati *et al.* 2006). Bekas *et al.* (2007) showed that, in that case, much less probes are needed. In the ideal case, where $R_{ij} = \delta_{ij}$, the second term on the right-hand side of (2) is identically zero. In the non-ideal but tomography-relevant case, most of the off-diagonal values of R_{ij} are small (Fig. 1) and the summation of products of independent zero-mean random numbers x_i and x_j with R_{ij} will not add very constructively. We demonstrate below that it is sufficient to use one single zero-mean random vector for the normalized expression (2) to converge towards

$$\mathbf{x}^t\mathbf{R}\mathbf{x}/\mathbf{x}^t\mathbf{x} \approx \sum_{i=1}^{M} R_{ii}x_i^2 / \sum_{i=1}^{M} x_i^2. \tag{3}$$

This is the weighted average (or mathematical expectation) of the elements of the diagonal of $\mathbf{R}$. Multiplying this average by the dimension M of the matrix is then a good approximation of the trace of $\mathbf{R}$ or

$$\mathrm{tr}(\mathbf{R}) \approx M \cdot \mathbf{x}^t\mathbf{R}\mathbf{x}/\mathbf{x}^t\mathbf{x}. \tag{4}$$

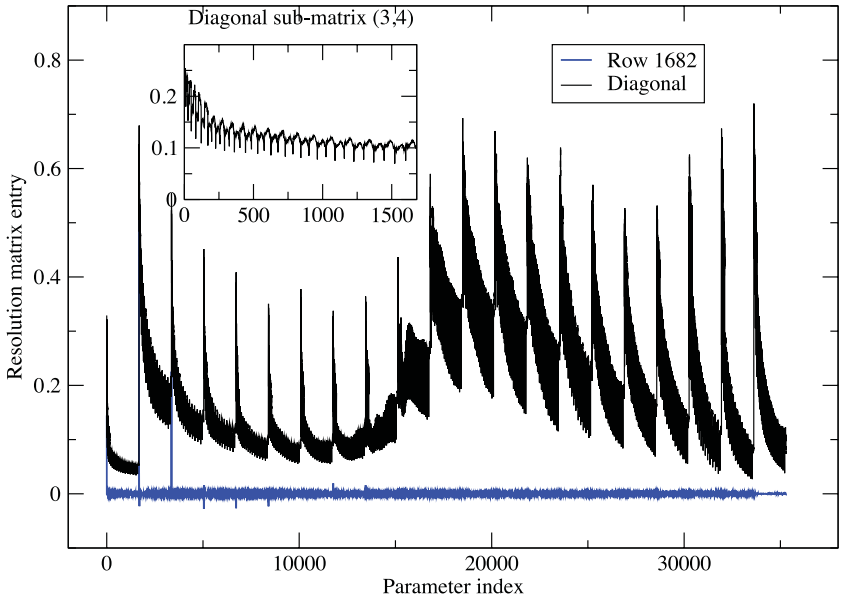

Figure 1. Diagonal and row 1682 of the resolution matrix corresponding to model S40RTS (Ritsema *et al.* 2011). The inset shows the diagonal of the corresponding submatrix $\mathbf{R}^{34}$.

In seismic tomography, the Earth is parametrized by N horizontal and K vertical basis functions, where $N \cdot K = M$. Because the vertical smearing usually dominates over the horizontal one, the resolution matrix appears to have K dominant diagonals, where usually $K \ll N$. The principal diagonal indicates how well we recover the amplitude of a given parameter. After N horizontal entries along a row, the next diagonal indicates how much a parameter at a given horizontal position is correlated with the one at the same horizontal position but at the following depth index and so on. Fig. 1 shows an example for the seismic tomography model S40RTS (Ritsema *et al.* 2011) which is parametrized horizontally with spherical harmonics up to degree 40 ($N = 1681$ coefficients) and vertically using $K = 21$ cubic splines. The dominant entries are along the main diagonal and decay with increasing spherical harmonic degree. This is due to the applied horizontal smoothing constraint. For a specific row of the resolution matrix, the off-diagonal terms are small, except for the same horizontal parameter at different depths. A similar structure of the resolution matrix was found by Soldati *et al.* (2006) for a different tomographic model and parametrization. This structure suggests that we can partition the random input vector $\mathbf{x}$ into K subvectors and $\mathbf{R}$ into $K \cdot K$ submatrices. The vector product of an input subvector $\mathbf{x^l}$ with an output subvector $\mathbf{(Rx)^m}$, where superscripts ^{l} and ^{m} are subvector indexes, will be a sum of bilinear forms

$$\sum_{n=1}^{K} (\mathbf{x^l})^\mathbf{t}(\mathbf{R^{mn}})(\mathbf{x^n}) = \sum_n \sum_i (R^{mn})_{ii}(x^l)_i(x^n)_i + \sum_n \sum_i \sum_{j \neq i} (x^l)_i (R^{mn})_{ij}(x^n)_j. \quad (5)$$

Each submatrix $\mathbf{R^{mn}}$ also has a decaying diagonal because of the horizontal smoothing constraint (i.e. an increasing index i corresponds to a higher spherical harmonic degree which is increasingly damped) and small off-diagonal elements (Fig. 1). Terms involving off-diagonal elements of each subresolution matrix will not contribute for the same reason as those in eq. (2) for the resolution matrix as a whole. $(x^l)_i$ and $(x^n)_i$ are different zero-mean random numbers, and because all diagonal elements of the submatrix $(R^{mn})_{ii}$ have the same sign, the only significant contribution to the sum will be from terms where $n = l$. We are therefore left with

$$(\mathbf{x^l})^\mathbf{t}(\mathbf{Rx})^\mathbf{m} / (\mathbf{x^l})^\mathbf{t}(\mathbf{x^l}) \approx \sum_{i=1}^{N} (R^{ml})_{ii}(x^l)_i^2 / \sum_{i=1}^{N} (x^l)_i^2. \quad (6)$$

The right-hand side of (6) is simply the weighted average of the diagonal elements $\mathbf{R^{ml}}$, and multiplied by the number of horizontal elements N, a good approximation of the trace of this submatrix or

$$\mathrm{tr}(\mathbf{R^{ml}}) \approx N \cdot (\mathbf{x^l})^\mathbf{t}(\mathbf{Rx})^\mathbf{m} / (\mathbf{x^l})^\mathbf{t}(\mathbf{x^l}). \quad (7)$$

We emphasize that we don't have to know the resolution matrix explicitly to use the proposed analysis. If the matrix has a few dominant diagonals, it is sufficient to generate one random zero-mean vector and calculate the corresponding synthetic data. The inversion of the latter with the same regularization parameters as used for the real data, will directly give $\mathbf{Rx}$, and vector product (3) or (6) will provide the desired traces for resolution analysis.

3 RESULTS

Model S40RTS is built with 8000 independent parameters. To test if the resolution matrix is diagonally dominant, we evaluate the full cross-correlation between a random input vector $\mathbf{x}$, uniformly drawn between -1 and 1, and the corresponding output vector $\mathbf{Rx}$ (Fig. 2). The cross-correlation at position 1 corresponds to expression (3) and is scaled to give the total trace of the resolution matrix. We find 8009, whereas the exact trace for this particular inversion is 8007 (we calculated the exact resolution matrix explicitly for this model). The cross-correlation at position 2 corresponds to the sum of the next diagonal and so on. They are small, indicating that the matrix is indeed diagonally dominant. The next significant diagonal is at position 1682 telling us that there is a correlation with the following depth layer. Eventually the modulo-1681 diagonals also disappear in the noise (i.e. background oscillations in the cross-correlation) of the cross-correlation. The background noise has a root-mean-square value of 70 which we take to be the overall precision with which we can evaluate the sum of a particular diagonal of the full resolution matrix. The number of consecutive modulo-1681 sums above the

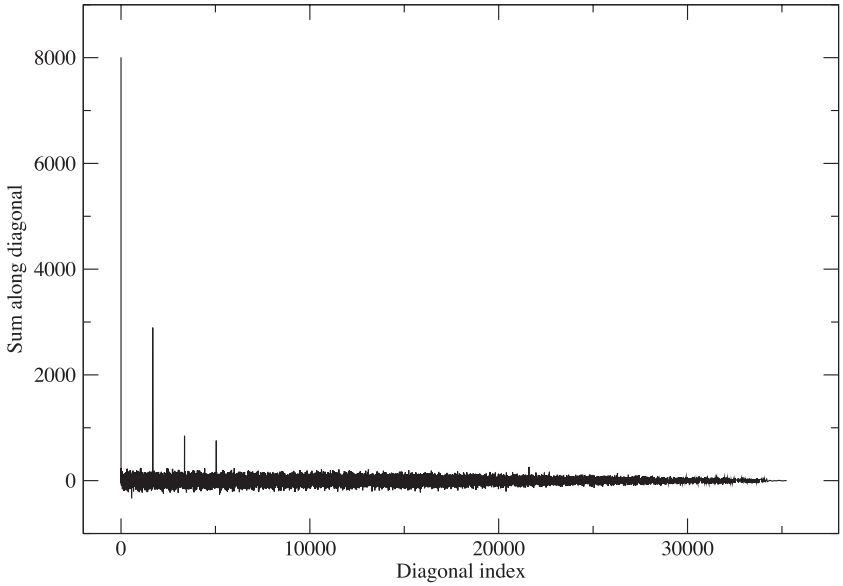

Figure 2. Cross-correlation of the total random input vector with the total output vector.

noise, renders a good indication of the average vertical resolution (4 spline knots in this case).

Rather than evaluating the cross-correlation of full random vectors, we take Expression (7) to estimate the traces of the $K \cdot K$ submatrices $\mathbf{R}^{\mathbf{ml}}$. Fig. 3 shows the traces of the corresponding submatrices. The average number of independent parameters resolved for a given depth index are along the diagonal. For instance at spline knot 12 (corresponding to depth of 951.3 km) the trace is 706. This corresponds to an average horizontal resolution of spherical harmonic degree $L = 25.6$ ($L = \sqrt{\mathrm{tr}} - 1$), or 783 km ($6371 \cdot \pi / L$). The peaks are not perfect delta functions but have a certain width. We measure the width of each peak where it drops below the noise level. For all submatrix estimations the root-mean-square of the noise is 15 (Fig. 4), which is about 70 for the full cross-correlation

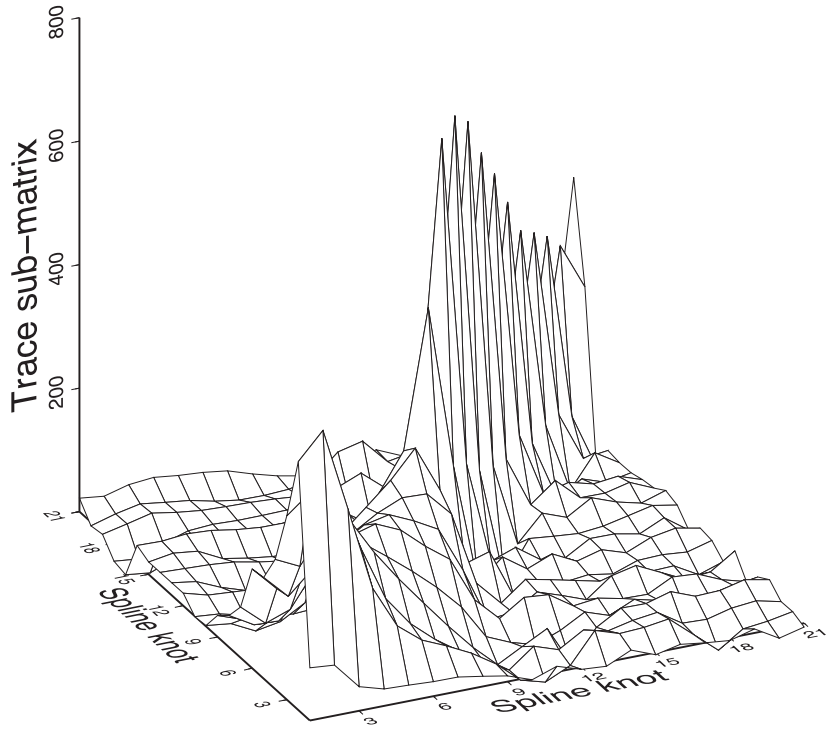

Figure 3. Traces of the 21 × 21 submatrices for model S40RTS.

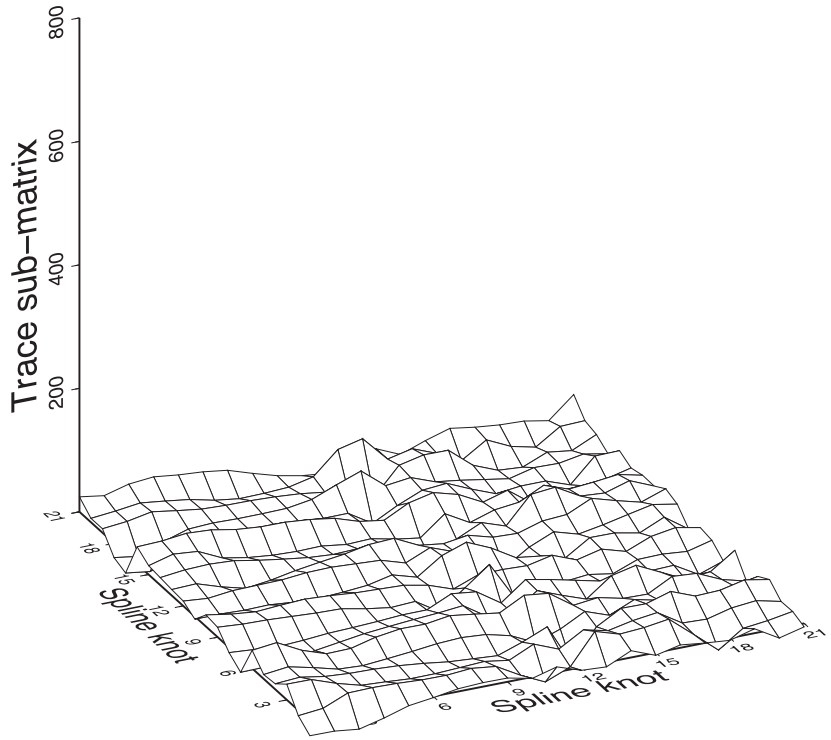

Figure 4. Difference between the traces of the exact resolution submatrices and those calculated using Expression (7).

Table 1. Average horizontal and vertical resolution for S40RTS, in brackets are values for increasing the weight of the surface waves by a factor of 10.

Spline knot	Depth (km)	Horizontal resolution (km)	Vertical resolution (km)
1	24.4	2112.3 (1543.1)	271.5 (610.0)
2	74.4	990.6 (732.3)	250.0 (315.0)
3	129.4	1207.2 (884.2)	248.5 (175.0)
4	189.9	1357.9 (1000.0)	284.5 (248.0)
5	256.4	1532.6 (1173.8)	312.0 (263.5)
6	329.7	1598.8 (1267.3)	338.5 (293.0)
7	410.2	1609.5 (1303.8)	357.0 (303.5)
8	498.8	1568.8 (1385.8)	331.5 (315.0)
9	596.3	1358.5 (1294.8)	312.5 (374.0)
10	703.6	990.8 (1052.9)	277.0 (392.0)
11	821.5	800.9 (870.4)	209.0 (239.5)
12	951.3	783.1 (861.4)	169.5 (186.0)
13	1094.1	804.8 (897.6)	185.5 (200.0)
14	1251.2	853.0 (967.7)	203.0 (217.0)
15	1424.0	896.3 (1029.0)	227.0 (239.0)
16	1614.1	971.5 (1136.3)	254.0 (263.5)
17	1823.2	1034.8 (1225.7)	278.0 (286.5)
18	2053.3	1082.2 (1290.9)	302.5 (309.5)
19	2306.3	1131.9 (1355.7)	314.5 (318.0)
20	2584.8	1243.3 (1495.1)	318.5 (316.0)
21	2891.0	1128.8 (1342.1)	281.0 (281.0)

divided by $\sqrt{K}$. At knot 12, we find an average depth resolution, defined as the half-width of the corresponding peak measured at twice the noise level, of 169.5 km. Table 1 gives the average horizontal and depth resolution for the other depths.

4 CONCLUDING REMARKS

Fig. 3 reveals that S40RTS has significantly lower vertical and horizontal resolution in the mantle transition zone despite the use of many high-quality overtone surface wave data. The reason for this is probably a specific choice for the weighting parameters. We produced another model with 8000 independent parameters, for which the weights of the surface waves compared to the body waves and normal mode data are higher by a factor of 10. The vertical and horizontal resolution is now much more uniform throughout the whole mantle (Table 1). Although the different data sets have slightly different misfits, the overall variance reduction is similar which makes the choice of a particular model subjective. It is not our purpose to discuss this new model, however our exercise serves as a reminder that even if millions of high-quality data enter a tomographic model construction, the details of resolution are still largely dependent on damping choices. The technique we proposed in this paper is very efficient to estimate the average horizontal and vertical resolution. we suggest that it can be used to identify desired properties in a model quickly and adjust the regularization accordingly.

In full waveform inversion, based on conjugate gradient methods, the Hessian plays the same role as the resolution in Newton-type algorithms (e.g. Fichtner & Trampert 2011). A similar analysis to that above for average horizontal and vertical resolution can thus be done using the Hessian instead of the resolution operator. Again we don't need to know the Hessian explicitly, but we only need to be able to estimate $\mathbf{Hx}$ efficiently. If $\mathbf{x}$ is a random zero-mean vector, expressions (3) or (6) can be used where $\mathbf{R}$ is replaced by $\mathbf{H}$.

In summary, the recipe is as follows. Draw a zero-mean random vector $\mathbf{x}$ and evaluate the corresponding data using the forward operator, giving $\mathbf{d} = \mathbf{Gx}$. Invert those data with the linear inverse operator to find $\tilde{\mathbf{x}} = \mathbf{Rx} = \mathbf{Ld}$. First evaluate the full cross-correlation between $\mathbf{x}$ and $\tilde{\mathbf{x}}$ which should look similar to that on Fig. 2. If the resolution operator is indeed dominated by a few diagonals only, the random vector and resolution can be partitioned, and Expression (6) can be used to evaluate the average horizontal and vertical resolution lengths. If the cross-correlation looks significantly different to that of Fig. 2, a more time-consuming analysis is needed, for instance based on the parametrization of the resolution operator (Fichtner & Trampert 2011; An 2012), matrix probing using more random vectors (MacCarthy *et al.* 2011) or a full calculation (Soldati et al. 2006).

REFERENCES

An, M., 2012. A simple method for determining the spatial resolution of a general inverse problem, *Geophys. J. Int.,* **191,** 849–864.

Avron, H. & Toledo, S., 2011. Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix, *J. ACM,* **58**(8), doi:10.1145/1944345.1944349.

Backus, G. & Gilbert, F., 1968. The resolving power of gross earth data, *Geophys. J. R. astr. Soc.,* **16,** 169–205.

Bekas, C., Kokiopoulou, E. & Saad, Y., 2007. An estimator for the diagonal of a matrix, *Appl. Numer. Math.,* **57,** 1214–1229.

Candès, E.M. & Wakin, M.B., 2008. An introduction to compressive sampling, *IEEE Signal Process. Mag.,* March, 21–30.

Chiu, J. & Demanet, L., 2012. Matrix probing and its conditioning, *SIAM J. Numer. Anal.,* **50,** 171–193.

Demanet, L., Létourneau, P.-D., Boumal, N., Calandra, H. & Chiu, J. S., S., 2012. Matrix probin: a randomized preconditioner for the wave-equation Hessian, *Appl. Comput. Harmon. Anal.,* **32,** 155–168.

Fichtner, A. & Trampert, J., 2011. Resolution analysis in full waveform inversion, *Geophys. J. Int.,* **187,** 1604–1664.

Halko, N., Martinsson, P. & Tropp, J., 2011. Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions, *SIAM Rev.,* **53,** 217–281.

Hutchinson, M., 1990. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines, *Commun. Stat. - Simul. Comput.,* **19,** 433–450.

Lévêque, J.J., Rivera, L. & Wittlinger, G., 1993. On the use of the checkerboard test to assess the resolution of tomographic inversions., *Geophys. J. Int.,* **115,** 313–318.

MacCarthy, J., Brochers, B. & Aster, R., 2011. Efficient stochastic estimation of the model resolution matrix diagonal and generalized cross-validation for large geophysical inverse problems, *J. geophys. Res.,* **116**(B10304), doi:10.1029/2011JB008234.

Parker, R.L., 1994. *Geophysical Inverse Theory,* Princeton University Press, Princeton.

Ritsema, J., Deuss, A., van Heijst, H.J. & Woodhouse, J.H., 2011. 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.,* **184,** 1223–1236.

Snieder, R. & Trampert, J., 1999. Inverse problems in geophysics, in *Wavefield Inversion,* pp. 119–190, ed. Wirgin, A., Springer Verlag, New York.

Soldati, G., Boschi, L. & Piersanti, A., 2006. Global seismic tomography and modern parallel computers, *Ann. Geophys.,* **49,** 977–986.

Spakman, W., 1991. Delay-time tomography of the upper mantle below Europe, the Mediterranean and Asia minor, *Geophys. J. Int.,* **107,** 309–332.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* Society for Industrial and Applied Mathematics, Philadelphia.