Geophys. J. Int. (2004) **157,** 595–606 doi: 10.1111/j.1365-246X.2004.02114.x

A deterministic algorithm for experimental design applied to tomographic and microseismic monitoring surveys

Andrew Curtis,[1,2] Alberto Michelini,[3] David Leslie[1] and Anthony Lomax[4]

[1] *Schlumberger Cambridge Research, High Cross, Madingley Road, Cambridge* CB3 0EL, *UK. E-mail: curtis@cambridge.oilfield.slb.com*
[2] *University of Edinburgh, Department of Geology and Geophysics, Grant Institute, West Mains Road, Edinburgh, UK*
[3] *Instituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS), Borgo Grotta Gigante* 42/c*, Sgonico* 34010*, Trieste, Italy*
[4] *Scientific Software, Mouans-Sartoux, France. E-mail: anthony@alomax.net, www.alomax.net*

Accepted 2003 August 18. Received 2003 July 8; in original form 2002 August 20

SUMMARY

Most general experimental design algorithms are either: (i) stochastic and hence give different designs each time they are run with finite computing power, or (ii) deterministic but converge to results that depend on an initial or reference design, taking little or no account of the range of all other possible designs. In this paper we introduce an approximation to standard measures of experimental design quality that enables a new algorithm to be used. The algorithm is simple, deterministic and the resulting experimental design is influenced by the full range of possible designs, thus addressing problems (i) and (ii) above. Although the designs produced are not guaranteed to be globally optimal, they significantly increase the magnitude of small eigenvalues in the model–data relationship (without requiring that these eigenvalues be calculated). This reduces the model uncertainties expected post-experiment. We illustrate the method on simple tomographic and microseismic location examples with varying degrees of seismic attenuation.

Key words: inversion, microseismicity, tomography.

1 INTRODUCTION

When data are difficult, time consuming or otherwise expensive to acquire, the ability to decide in advance which data are likely to add most information and hence to be most valuable becomes important (Curtis 2000). There may be a trade-off between data value and cost of data acquisition in order to decide which data will ultimately be collected. The field of experimental design covers techniques that accomplish this valuation and design process (e.g. Silvey 1980; Atkinson & Donev 1992).

In geophysics, experimental design techniques have been applied to seismic tomography problems (Curtis 1999a,b), electromagnetic experiments (Maurer & Boerner 1998; Maurer *et al.* 2000) and seismic location (Kijko 1977; Rabinowitz & Steinberg 1990; Steinberg *et al.* 1995). In all of these applications, and many others, the measure of data importance is defined in terms of the amount of information each potential data set (each experimental design) is expected to provide about specific model parameters of interest when the model–data relationship is assumed to be approximately linear. However, this measure is related in a highly non-linear way to the design. This led all of the above authors to implement optimization algorithms, which either use initial designs that are gradually modified into improved designs (e.g. Kijko 1977; Rabinowitz & Steinberg 1990; Steinberg *et al.* 1995, all using the heuristic method of Mitchell 1974), or use stochastic optimization methods (e.g. genetic algorithm or simulated annealing) to find sets of good designs (e.g. Smith *et al.* 1992; Maurer & Boerner 1998; Curtis 1999a,b). The disadvantage of the first approach is that the best design found always depends on the starting design used, and this starting design contains no information about the range of all possible designs. The disadvantage of the second approach is that their stochastic nature implies that they do not necessarily achieve the same result each time they are run; indeed, there is no guarantee that they will perform well at all given limitations of computing power and hence on the number of designs that can be tested.

In this paper we present a new algorithm that can be used to design any experiment deterministically when: (a) the number of possible designs is finite, and (b) the model–data relationship is approximately linear. The final design is influenced by the full range of all possible experimental designs. We achieve this by approximating the usual measure of data importance: this approximation allows a deterministic algorithm to be used to create a final design. This approach confers advantages of (i) the repeatability of the design process, and (ii) the guarantee of obtaining an experimental design within a finite length of time for a problem of finite size. Although the final design is not guaranteed to be globally optimal (the best out of all possible designs), in all of the examples examined in this paper the designs created performed well. Similar performance was also observed when the algorithm was used to design optimal surveys to elicit knowledge from experts (Curtis & Wood 2004).

Below we describe the new methodology and illustrate it by designing two simple tomographic experiments and two simple microseismic monitoring surveys. This paper is the culmination of several pieces of work: it is based on a suggestion to improve inverse

problem conditioning made by Sabatier (1977), further investigated by Michelini and Lomax (unpublished manuscript) and developed into the current experimental design method by Curtis and Leslie (Schlumberger confidential report).

2 METHODOLOGY

For any given model vector $\mathbf{m}$ of dimension P within model space M, let the forward problem of estimating the corresponding data vector $\mathbf{d}$ of dimension N in data space D be accomplished by the matrix operator $\mathbf{A}$, i.e.

$$\mathbf{d} = \mathbf{A}\mathbf{m}, \tag{1}$$

where the row i and column j of $\mathbf{A}$ contain the element

$$A_{ij} = \frac{\partial d_i}{\partial m_j} \tag{2}$$

where d_i and m_j are elements of vectors $\mathbf{d}$ and $\mathbf{m}$, respectively. In linearized problems both $\mathbf{m}$ and $\mathbf{d}$ may be perturbations around reference values (e.g. Tarantola 1987). Then, for a given data vector $\mathbf{d}_0$, we often wish to find the model vector $\mathbf{m}_0 \in M$ such that $|\mathbf{d}_0 - \mathbf{A}\mathbf{m}_0|^2$ is minimized. Theoretically, this is accomplished by pre-multiplying eq. (1) by $\mathbf{A}^T$ (where T is the matrix transpose) and taking a matrix inverse:

$$\mathbf{m_0} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{d_0}. \tag{3}$$

Instability in the solution arises because the $N \times N$ square matrix,

$$\mathbf{L} = \mathbf{A}^T\mathbf{A}, \tag{4}$$

is often near singular, i.e. some of its eigenvectors, $\{\mathbf{e}_i : i = 1, \ldots, N\}$, for example, have extremely small eigenvalues $\{\lambda_i : i = 1, \ldots, N\}$. Measurement errors in the data space D propagate into solution $\mathbf{m}_0$ parallel to each eigenvector $\mathbf{e}_i$ with an amplification $1/\lambda_i$; hence, when small eigenvalues exist the solution becomes unstable and unreliable, and the inverse problem is said to be ill-conditioned (Menke 1989).

The algorithm for survey design introduced herein is based on an observation made by Sabatier (1977). Sabatier noted that ill-conditioning in matrix $\mathbf{A}$ is caused because many data add little or no additional information to other data collected. In other words, there are some model–datum relationships (described by rows of matrix $\mathbf{A}$) that are linear combinations of all of the other model–datum relationships (other rows of matrix $\mathbf{A}$).

Sabatier (1977) suggested that inverse problem conditioning could be improved as follows: the data should be ordered by the P-dimensional angle between the row of matrix $\mathbf{A}$ corresponding to each datum and the space spanned by all other rows. This angle is a measure of the linear independence of each row relative to all others. Rows for which this angle is large contribute information about the model; rows for which the angle is small are effectively merely consistency conditions within the noise level of the data. The problem with these consistency conditions is that they are largely responsible for the effect that small variations in data resulting from noise can lead to large variations in estimated model parameters (ill-conditioning) in under-determined problems. Therefore, Sabatier (1977) suggested that only data corresponding to rows for which the associated angle is greater than some noise-dependent threshold should be used for the inversion: this has the effect of improving the conditioning of matrix $\mathbf{L}$.

This idea was developed by Michelini and Lomax (unpublished manuscript). They ordered the rows of $\mathbf{A}$ according to the following penalty function for each row i:

$$\alpha_i = \sum_{j=1, j\neq i}^{N} \left[\frac{\mathbf{a}^{(i)} \cdot \mathbf{a}^{(j)}}{\left\|\mathbf{a}^{(i)}\right\| \left\|\mathbf{a}^{(j)}\right\|} \right]^2 \quad (i = 1, 2, \ldots, N), \tag{5}$$

where $\mathbf{a}^{(i)}$ is the ith row of matrix $\mathbf{A}$. Geometrically, each α_i is the sum of squares of the cosines of the angles in an M-dimensional space between vector $\mathbf{a}^{(i)}$ and all other row vectors and hence, is a measure of the angle described by (Sabatier (1977), see above). In order to improve the conditioning and reduce the size of tomographic inverse problems for which a profusion of data exists, Michelini and Lomax preserved datum d_i only if α_i was greater than some threshold.

We will use a similar method to design surveys. The situation is slightly more complicated than that considered by Sabatier (1977), or by Michelini and Lomax because we wish to begin with all possible data that could be recorded with the available equipment (in the seismic case, for example, corresponding to all possible seismic arrivals recorded at all possible locations at which seismic sources/receivers of a range of possible types could be placed): we will preserve only that equipment for which the corresponding data is expected to contain information that is as independent as possible of all other preserved data. From here on we refer to the equipment used as receivers.

In order to do this we must change the penalty function used above to account for the fact that each receiver may contribute multiple data (e.g. traveltimes from several seismic arrivals) and, hence, may contribute several rows to matrix $\mathbf{A}$. In addition, we would like to be able to specify differences in expected data uncertainties (for example, as a result of differences in expected traveltime, picking accuracies for different arrival types, or different types of equipment) and which model parameters we would like the penalty to be particularly sensitive to and, hence, focused on in the resulting designs (see Curtis 1999b).

2.1 The importance of data

In order to define which receivers should be included in our experimental design we define a new measure of receiver quality. This, in turn, must depend on the importance of the data that a specified receiver is expected to record. We define the quality of receiver k to be given by the function,

$$\gamma_k = \sum_{j=1}^{n_k} \sum_{i=1}^{N_r} \left[1 - \frac{\left|\mathbf{a}^{(l_j^k)} \odot \mathbf{a}^{(i)}\right|}{\left\|\mathbf{a}^{(l_j^k)}\right\| \left\|\mathbf{a}^{(i)}\right\|} \frac{\sigma_d^{(l_j^k)} \sigma_d^{(i)}}{\sigma_{dx}^2 w_{mx}^2} \right]^{\epsilon} \delta^{(l_j^k)} \delta^{(i)} \tag{6}$$

$(k = 1, 2, \ldots, S)$, where S is the total number of receivers, $l_j^k : j = 1, \ldots, n_k$ is the set of row indices that defines the set of all rows $\mathbf{a}^{(l_j^k)} : j = 1, \ldots, n_k$ associated with receiver number k, $\delta^{(i)}$ equals one or zero depending on whether row i is on (used) or off (unused), respectively, $\mathbf{A}$ contains derivatives corresponding to all possible data that could be recorded given available equipment and physical constraints [the ith datum corresponding to row $\mathbf{a}^{(i)}$], N_r equals the total number of such data, $\sigma_d^{(i)}$ is the standard deviation expected for datum i (the maximum such standard deviation squared across the data set being σ_{dx}^2), w_{mx}^2 is the maximum squared model parameter weighting factor across all model parameters, and ϵ is a positive exponent. For any two vectors $\mathbf{a}$ and $\mathbf{b}$ the weighted scalar product $\odot$ is defined to be

$$\mathbf{a} \odot \mathbf{b} = \sum_{l=1}^{P} \frac{a_l b_l}{w_m^{(l)} w_m^{(l)}}, \tag{7}$$

where $w_m^{(l)}$ is the weighting factor associated with model parameter l and norms in the denominator of equation (6) are calculated according to the weighted scalar product in equation (7), i.e. $\|\mathbf{a}\|^2 = \mathbf{a} \odot \mathbf{a}$. Notice that because each $\mathbf{a}^{(i)}$ is normalized in eq. (6), each term within the summations is always positive in the range [0, 1].

We have chosen to use a quality function (receivers with highest quality function values provide the most information about the model parameters of interest) rather than a penalty function [vice versa; see, for example, the penalty function presented in eq. (5)]. This is to ensure that γ_k increases either as matrix rows become less linearly dependent (the scalar products tend towards zero), or as the number of data (n_k) that can be collected using receiver k increases (the reverse is not true for penalty functions).

Other than the change just described, eq. (6) is merely a generalization of eq. (5): this in turn is Michelini and Lomax's suggested measure of the data angle of Sabatier (1977). Eq. (6) generalizes to the case where the angle measurement is weighted both by the data uncertainties (through σ_d and normalized by σ_{dx}) and by focussing on desired subsets of model parameters (through w_m and normalized by w_{mx}). It also includes an extra summation (over n_k) to account for the possibility of multiple matrix rows (multiple data) being recorded per receiver.

The experimental design algorithm (below) requires that we can rapidly recalculate the quality of all receivers when any particular receiver is switched off. For example, switching off receiver number q with the effect of removing all associated rows $\mathbf{a}^{(l_j^q)}$, $j = 1, \ldots, n_q$ of matrix $\mathbf{A}$. Consider the effect on the quality of a different receiver (number s) of switching off any single one of these rows, row $\mathbf{a}^{(r)}$, for example. Receiver s is associated with rows $\mathbf{a}^{(l_j^s)}$, $j = 1, \ldots, n_s$. For each row $\mathbf{a}^{(l_j^s)}$, the contribution to quality function γ_s is the sum of the terms in eq. (6) involving N weighted scalar products with all other rows. Hence, for any row $\mathbf{a}^{(l_j^s)}$ such that $l_j^s \neq r$, the only change in that row's contribution to γ_s is the term involving the weighted dot product between rows r and l_j^s (because row r will be switched off). When row r is switched off, the total change to quality function γ_s for all rows such that $l_j^s \neq r$ is therefore given by:

$$c_{1,s}^{(r)} = \sum_{j=1, l_j^s \neq r}^{n_s} \left[1 - \frac{\left| \mathbf{a}^{(l_j^s)} \odot \mathbf{a}^{(r)} \right|}{\left\| \mathbf{a}^{(l_j^s)} \right\| \left\| \mathbf{a}^{(r)} \right\|} \frac{\sigma_d^{(l_j^s)} \sigma_d^{(r)}}{\sigma_{dx}^2 w_{mx}^2} \right]^{\epsilon} \delta^{(l_j^s)} \delta^{(r)}, \tag{8}$$

where the δ terms on the right-hand side have the same values as before receiver q was switched off.

It is possible that $l_j^s = r$ for some j if both receivers q and s are required to record the single datum number r (e.g. both a seismic source and seismic receiver are required to measure a single travel-time datum). In this case, the entire set of scalar products of row r with all others will be lost when this row is switched off. Hence, the total change to quality function γ_s will be given by:

$$c_{2,s}^{(r)} = \sum_{i=1}^{N} \left[1 - \frac{\left| \mathbf{a}^{(r)} \odot \mathbf{a}^{(i)} \right|}{\left\| \mathbf{a}^{(r)} \right\| \left\| \mathbf{a}^{(i)} \right\|} \frac{\sigma_d^{(r)} \sigma_d^{(i)}}{\sigma_{dx}^2 w_{mx}^2} \right]^{\epsilon} \delta^{(r)} \delta^{(i)}. \tag{9}$$

Finally, the total correction to quality function γ_s is given by the sum of these two correction terms summed again over all rows r that were associated with the receiver q when it was switched off:

$$c_s^{(q)} = \sum_{i=1}^{n_q} c_{1,s}^{l_r^q} + c_{2,s}^{l_r^q}. \tag{10}$$

2.2 Algorithm

The algorithm used here carries out the following steps:

(i) Calculate all S receiver qualities using eq. (6).

(ii) Find the receiver (for example, number q) with the lowest quality [this will be switched off in step (iv)].

(iii) Using eqs (8), (9) and (10) calculate corrections $\mathbf{c}^{(l_j^q)_s}$ to the qualities of all other receivers s caused by the removal of rows $l_j^q : j = 1, \ldots, n_q$ associated with receiver q. Update all receiver qualities with these corrections.

(iv) Switch off the receiver with the lowest quality by setting $\delta^{(l_j^q)} = 0$, $j = 1, \ldots, n_q$.

(v) Repeat the process from step (ii) until data acquisition using the receivers remaining switched on can be implemented at an acceptable cost.

(vi) The remaining receivers constitute the final survey design.

There are several points of interest concerning this algorithm. First, notice that in this algorithm the subjective, noise-dependent threshold required by Sabatier (1977) and Michelini and Lomax (unpublished manuscript) has been replaced by the objective cost threshold in step (v).

Secondly, the algorithm begins with all possible receivers switched on. Because the quality measures considered at each iteration depend on all active receivers, this is the sense in which the algorithm, and thus the final design, is influenced by all possible receivers and, hence, by all possible experimental designs.

Thirdly, the algorithm does not guarantee convergence to the globally optimal design because at each iteration it chooses the best design from those that can be formed by switching off another receiver without switching any back on. Thus, the design chosen suffers from the legacy of receivers switched off previously. This aspect of the algorithm could be relaxed by allowing one or more receivers to be switched on at each iteration. However, to achieve global optimality, either all combinations of receivers would have to be explored, or a stochastic approach could be adopted where, for example, a metropolis or simulated annealing algorithm was used to switch receivers on and off. The former approach would lead to a combinatoric expansion in required computation, the latter would guarantee global optimality only in the limit of infinite computation (even for a finite problem). Neither approach confers the advantages of the current algorithm, which is simple, deterministic and converges efficiently to a final design.

Fourthly, instead of beginning with the full design (all receivers on) and iteratively removing receivers, one could begin with no receivers and add them one-by-one using correction terms eqs (8) and (9) in the opposite sense. Notice, though, that in the case of zero parameter weighting and equal data uncertainty, the first receiver added would provide equal information whichever receiver was added. Hence, to converge to a unique solution, S experimental designs would have to be created, each found by seeding the algorithm with a different initial receiver. Whether this alternative algorithm or the one presented above would be more efficient depends on both the total possible number of receivers S and the number of receivers required in the final design.

Fifthly, some previously published algorithms allow many designs to be found, each with approximately the same quality. This can be exploited later when further cost considerations (financial, time, effort) make the acquisition of some designs more attractive than others. Such cost considerations can easily be incorporated within the algorithm described above by including a cost factor associated with each receiver in eq. (6). Thus, if at a given iteration several receivers provide equally little information to the design, this ambiguity can be exploited by switching off the most costly receiver.

Finally, this algorithm relies on eq. (6) providing a reliable measure of experimental design quality. We note again that while most previous algorithms mentioned in the Introduction explicitly optimize measures of the quality of the eigenvalue spectrum itself, eq. (6) is only an approximation to such a measure. However, despite all of the shortcomings of our algorithm described above, in the examples below we demonstrate that the designs found seem to perform close to optimally in the cases tested.

3 EIGENVALUE ANALYSIS OF RESULTING DESIGNS

Fig. 1 shows the geometries of the maximum allowed set of ray paths in two simple tomographic experiments. Traveltimes can be measured along each of the ray paths and the aim of the experiment is to invert these data for the slowness of each of the nine square cells, assuming the slowness is constant within each cell. Each ray path has a source at one end and a receiver at the other, however, consistent with the vocabulary used above, we refer to both source and receiver as simply receivers. Hence, for example, the experiment depicted on the left of Fig. 1 has a maximum number of 20 receivers. The aim of the experimental design procedure is to allow the slowness of each cell to be inferred with the minimum number of receivers.

Matrix $\mathbf{A}$ has elements A_{ij}, which are the derivatives of traveltime i with respect to slowness j. We assume that all traveltimes have equal uncertainty so $\sigma_d^{(i)} = 1$ for all i. We also apply no weighting to the different model parameters so $w_m^{(l)} = 1$ for all l. This ensures that all information relevant to designing the experiment is contained within matrix $\mathbf{A}$ alone, so that eigenvalue analysis of $\mathbf{A}^T\mathbf{A}$ provides full information about the quality of the final design.

The algorithm above was run to remove receivers sequentially from the maximum set shown on the left of Fig. 1. The algorithm removed receivers in the following order:

13, 14, 16, 15, 3, 4, 7, 8, 6, 5, 10, 9, 1, 2, 12, 11, 18, 17, 20, 19.

Comparing this order with Fig. 1 we see that ray paths that are effectively duplicated (with respect to their sensitivity to model parameters) are removed first. The four plots in Fig. 2 show the eigenvalue spectra after 1, 6, 11 and 16 receivers, respectively, were removed from the maximum set. We compare these eigenspectra to those obtained by removing a random set of the same number of receivers from the maximum set, producing a random design. Each plot in Fig. 2 shows 20 realizations of such random designs (many of these spectra overlie each other) and the dashed line shows the average of these realizations in each case.

The design algorithm was then run on the experiment depicted on the right of Fig. 1, which resembles a coarsely discretized version of a cross-well experiment with two vertical wells following the left and right edges of the model. The order of removal of receivers was:

8, 10, 6, 2, 7, 3, 9, 1, 11, 5, 4.

Comparing this order with Fig. 1 we see that the first three receivers removed (6, 8 and 10) create a symmetrical receiver geometry in the left and right wells. Receivers removed thereafter are the central receivers (2, 3, 7 and 9), thus maintaining the maximum angular coverage with the remaining receivers (1, 4, 5 and 11). The four plots in Fig. 3 show the eigenvalue spectra after 1, 3, 5 and 7 receivers, respectively, were removed from the maximum set. These are compared to spectra obtained by removing a random set of the same number of receivers. Each plot in Fig. 3 shows 20 realizations of spectra from such random designs.

Every plot in Figs 2 and 3 shows that the design produced by the new algorithm maximizes the magnitude of the small eigenvalues. This behaviour is observed in all examples tested. Sometimes this is achieved at the expense of the large eigenvalues and this characteristic will be observed as a direct trade-off in the microseismic examples presented later. However, in most plots the larger eigenvalues are also higher than, or at least close to, those obtained on average for a random design.

This behaviour makes sense intuitively. The algorithm effectively only removes data for which their sensitivities to the model parameters are almost or completely linearly dependent on other data from the experiment. Such data increase the magnitude of the large eigenvalues of $\mathbf{A}^T\mathbf{A}$ so the algorithm effectively only seeks to remove large eigenvalues, leaving smaller eigenvalues intact.

This behaviour is desirable in an experimental design method provided that data is expected to be of sufficiently high quality that we aim to include as much information in the small (non-zero) eigenvalues as possible. Noise in the data will be projected into the model space parallel to each eigenvector with an amplification of the inverse of the corresponding eigenvalue (see earlier). To avoid instability in the inverse problem solution when noise is significant, either eigenvectors associated with small eigenvalues must be removed from the model space, or small eigenvalues must be boosted by regularization. Hence, if data is expected to be of poor quality then in some cases it may be better to use an experimental design method that improves the large eigenvalues at the expense of the

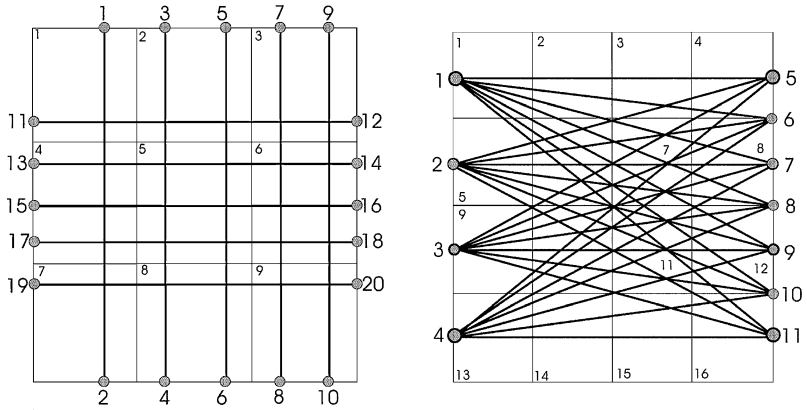

Figure 1. Geometries of two simple tomography problems: squares represent discretization of the medium into cells of constant slowness numbered 1 to 9 (left) and 1 to 16 (right). Bold lines are ray paths along which traveltimes were measured. Each path has a source at one end and a receiver at the other numbered 1 to 20 (left) and 1 to 11 (right).

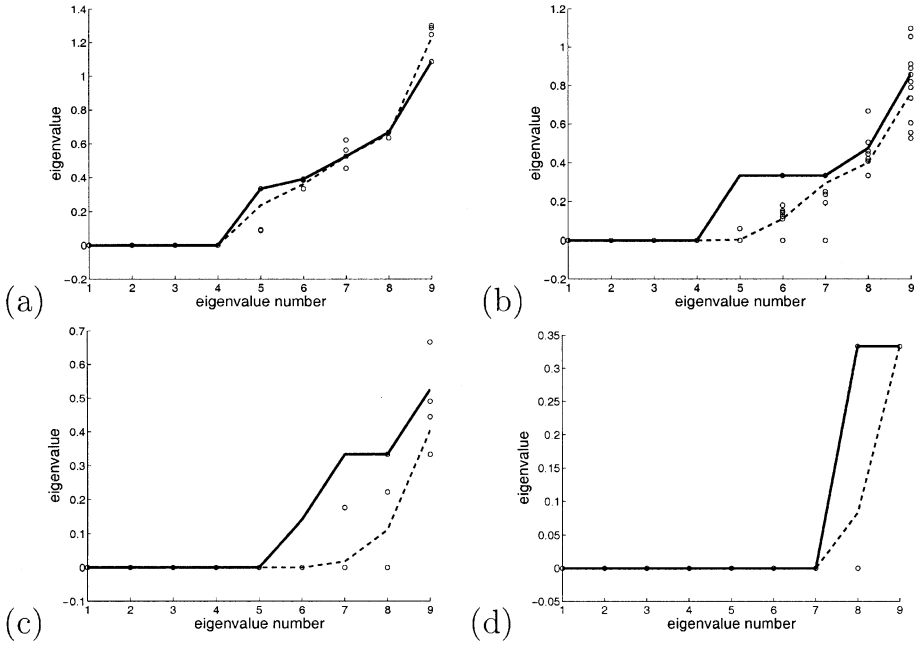

Figure 2. Plots (a), (b), (c) and (d), respectively, show eigenvalue spectra after 1, 6, 11 and 16 receivers are removed from the tomography problem on the left of Fig. 1. Circles show 20 realizations of spectra after a random set of 1, 6, 11 and 16 receivers were removed, respectively (many eigenvalues are equal so circles overlap). Dashed lines show the average of these spectra. Bold lines show spectra after the same numbers of receivers were removed in the order defined by the design algorithm.

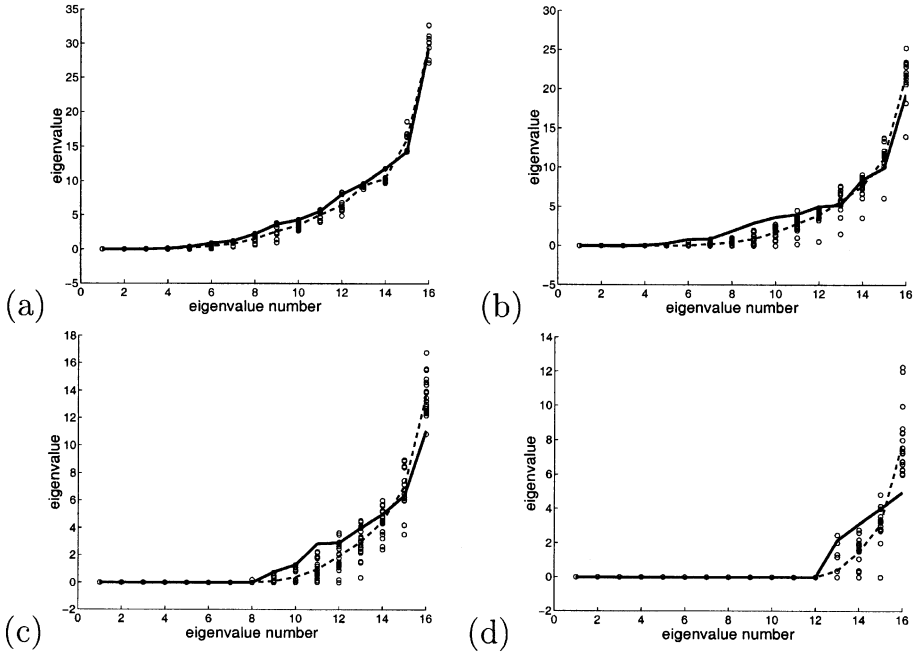

Figure 3. Plots (a), (b), (c) and (d), respectively, show eigenvalue spectra after 1, 3, 5 and 7 receivers are removed from the tomography problem on the right of Fig. 1. Circles show 20 realizations of spectra after a random set of 1, 3, 5 and 7 receivers were removed, respectively. Dashed lines show the average of these spectra. Bold lines show spectra after the same numbers of receivers were removed in the order defined by the design algorithm.

small ones (Curtis & Snieder 1997 and Curtis 1999a, present such a method). When data is expected to be of high quality (low noise) then the current algorithm may provide best results.

4 MICROSEISMIC MONITORING EXAMPLES

Fig. 4 shows the geometry of a simplified (2-D) microseismic monitoring experiment for which we will design an optimal receiver geometry using the above algorithm. A vertical well on the left intersects a reservoir formation at a depth of 1500 m. It is often the case that when large amounts of fluid are pumped into or out of the reservoir through the well, fracturing occurs in the formation as a consequence of resulting stress changes. Such fracturing is expected to occur close to the reservoir layer and to be concentrated close to the well. We have represented the expected distribution of fractures with seven discrete fracture locations at horizontal distances 10×2^n m, $n = 0, \ldots, 6$ from the well, marked by asterisks.

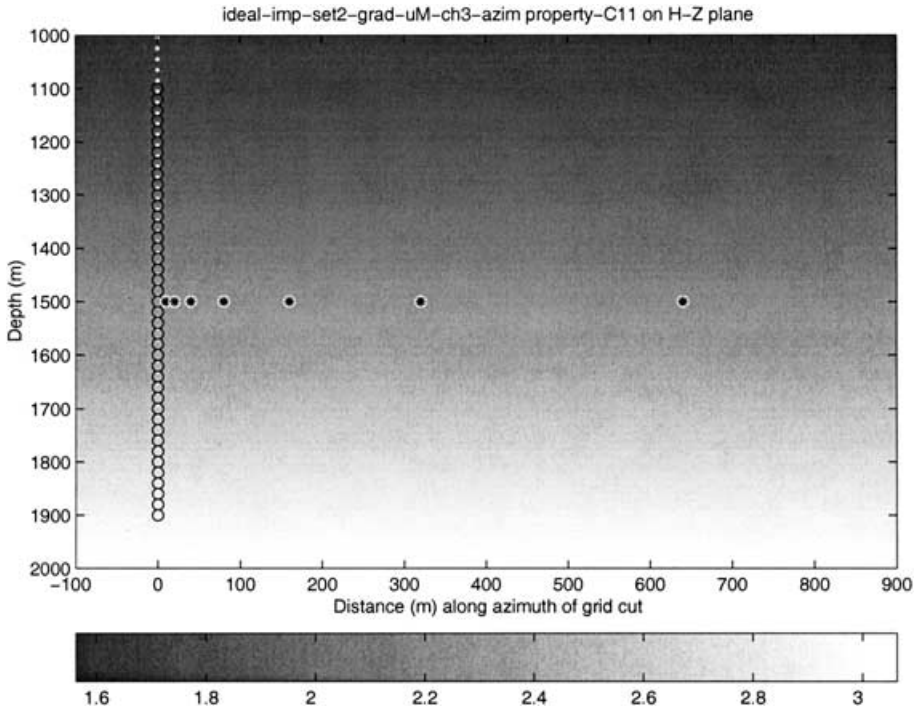

Figure 4. Velocity model (shading), receiver locations (black circles) and microseismic event locations (asterisks) used in experimental design examples.

When a fracture occurs, seismic P and S waves travel out into the formation at their respective seismic velocities. In Fig. 4 the P-wave velocities increase linearly with depth and are represented by the shading (in km s^{-1}). S-wave velocities are assumed to follow $V_p/V_s = 1.6$ where V_p and V_s are P- and S-wave velocities, respectively. As the seismic energy travels it is attenuated by geometrical and anelastic effects: energy attenuation increases as anelasticity increases and as energy travels further. Some of the remaining energy passes through the well and can be detected by seismic receivers placed there. By detecting the P and S wave arrival times, t_p and t_s, respectively, of energy traversing essentially the same path through the formation, the set of data $t_s^i - t_p^i : i = 1, \ldots, N$ (arrival time differences at each of N receivers) is related to the event location and is not dependent on its time of occurrence (see e.g. Tarantola & Valette 1982). The uncertainty σ_d^i of each of $t_s^i - t_p^i$ datum is dependent on the signal to noise ratio at each receiver and in this example we assume that this depends on energy attenuation only. We use the following relationship between attenuation and expected data uncertainty σ_d^i that was derived in another (confidential) study, although any other relevant relationship could also be used:

$$\sigma_d^i = \sigma_{\text{ref}} \exp\left[-gB(t - t_{\text{ref}})\right] \left(\frac{t_{\text{ref}}}{t}\right)^{BC}. \tag{11}$$

Here t is the traveltime, $g = \pi f/Q$ where f is the dominant frequency and Q is the formation quality factor (the inverse of attenuation) that is assumed constant over the background medium. t_{ref} and σ_{ref} are estimated traveltimes and uncertainties at some reference receiver (we use the receiver at 1500 m depth). Constants B and C control the overall form of the relationship and we will use $C = 1$ throughout this paper. Data uncertainty is, therefore, expected to increase as attenuation increases or as the receiver is placed further away from the expected event location (towards the top or bottom of the well) because this increases the time of travel t of the seismic energy. We represent the set of all possible receiver locations (and, hence, the set of possible experimental designs) by the 41 locations marked as circles within the well in Fig. 4.

We wish to select an optimal set of seismic receiver locations such that the representative set of microseismic events will be located as well as possible in the following sense. If we define the model vector **m** to be the set of all event location parameters (two coordinates for each of seven events, hence **m** has 14 elements) then we wish the uncertainty on **m** to be minimized. In these examples we set all $w_m^{(l)} = 1$ so that no model space weighting is used. We calculate elements of matrix **A** in eq. (2) by:

(i) calculating traveltimes between all potential event and receiver locations by solving the eikonal equation;
(ii) perturbing event locations in turn by small amounts horizontally then vertically;
(iii) recalculating traveltimes for the perturbed locations;
(iv) calculating derivatives in matrix **A** using finite difference approximations.

The above definitions, relationships and calculations provide all information necessary to apply the design algorithm presented earlier. We first design an experiment with constant background velocity (no velocity gradient) of $V_p = 3000$ m s^{-1}, $V_s = 1875$ m s^{-1}. In this case, energy travels along straight paths between events and receivers. Each $t_s^i - t_p^i$ datum for any event provides an estimate of the distance l_i between that event and receiver i because

$$l_i = \left(t_s^i - t_p^i\right)\left(\frac{1}{V_s} - \frac{1}{V_p}\right)^{-1}. \tag{12}$$

Hence, the location of each event is obtained from the set of such distances by triangulation.

The design algorithm is applied exactly as described earlier. The entire array of all possible receiver locations is used as a starting design and during each iteration of the algorithm one receiver is removed (switched off). We continue to iterate until only one receiver remains. The order in which receivers are removed defines a ranking

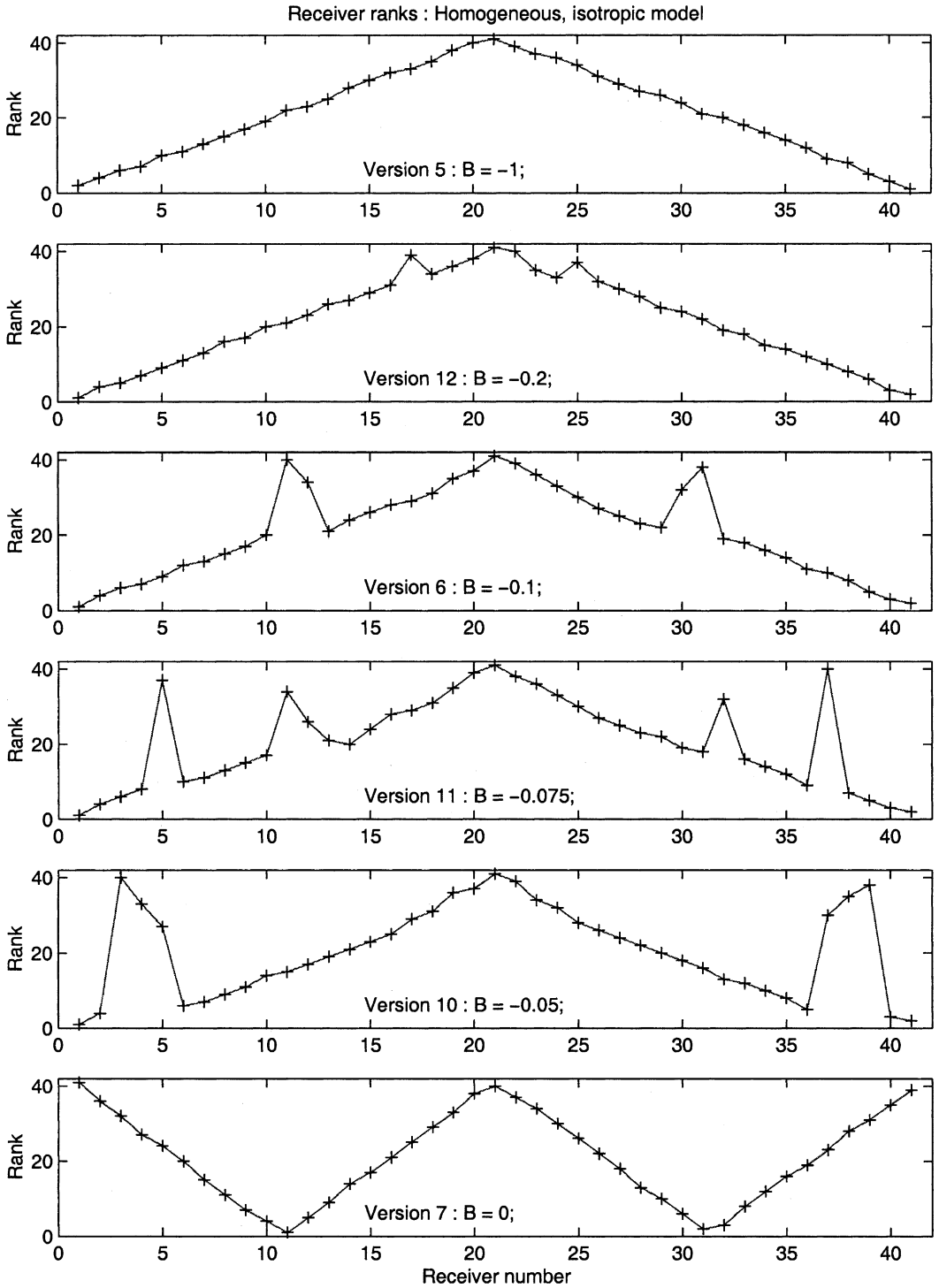

Figure 5. Results obtained using a homogeneous background velocity model and increasing effects of attenuation (from bottom to top). Values of parameter B are shown on the figure.

of receivers: rank 1 is attributed to the first receiver removed, rank 41 is attributed to the final receiver remaining. We can then plot these ranks as shown in Fig. 5.

Fig. 5 shows six independent sets of results each obtained using a different value of B to generate data uncertainties ($B = -1, -0.2, -0.1, -0.075, -0.05$ and 0 in eq. 11). The final (lowest) plot shows results for a completely elastic medium in which data uncertainty does not increase with distance from the event. In practice, these plots are used as follows: select the set of ranks associated with what is believed to be the correct relationship between attenuation and data uncertainty, then add receivers of consecutive, decreasing rank to the experimental design starting at rank 41. Receivers are added until either sufficient information is expected to be obtained from those already selected, or until some threshold of cost of carrying out the experiment has been superseded.

For an elastic, homogeneous medium with $B = 0$, following this approach results in an experimental design in which receivers are clustered around the top, middle and bottom of the well with twice

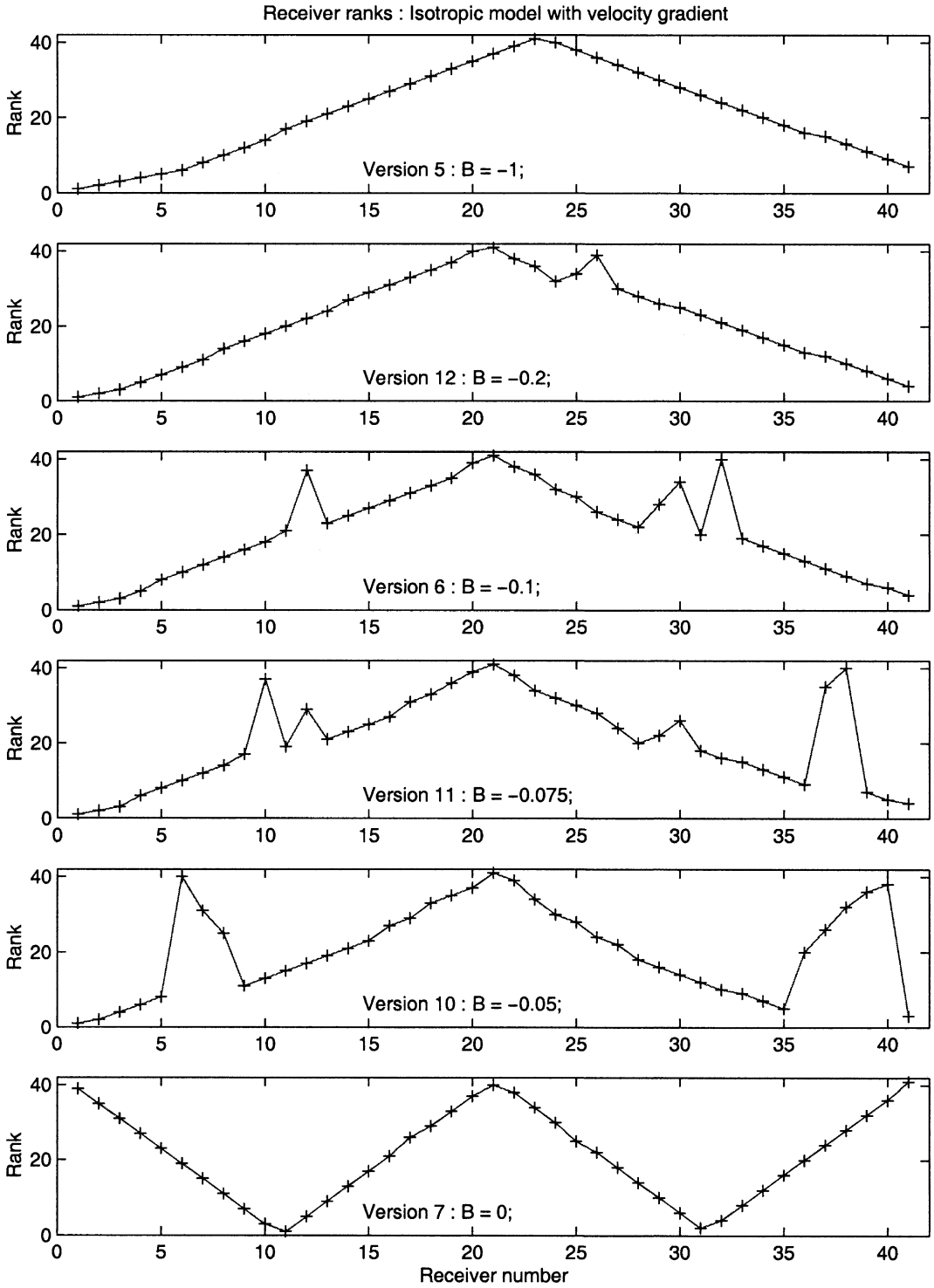

Figure 6. Results obtained using the linear background velocity model shown in Fig. 4 and increasing effects of attenuation (from bottom to top). Values of parameter B are shown on the figure.

as many receivers in the middle than at the top or bottom of the well. This makes sense intuitively: with no anelasticity the seismic energy is expected to be detected at all receiver locations with equal uncertainty; the best source location estimates can be obtained by using as wide a fan of event-receiver paths as possible for triangulation. As B increases in magnitude (B is negative), receivers near the top or bottom of the well still produce a wide fan of event-receiver paths, however, this trades off with increasing uncertainty with distance from the reservoir. Hence, as B decreases towards -1 the optimal fan becomes narrower and narrower until ($B = -1$) we obtain the best experimental design by simply choosing receiver locations as close as possible to the reservoir around 1500 m depth.

Fig. 6 presents results obtained when the velocity gradient shown in Fig. 4 is used in place of the homogenous velocity field used to obtain results in Fig. 5. The results show similar patterns to those in Fig. 5, except that the depths at which receivers should be placed are

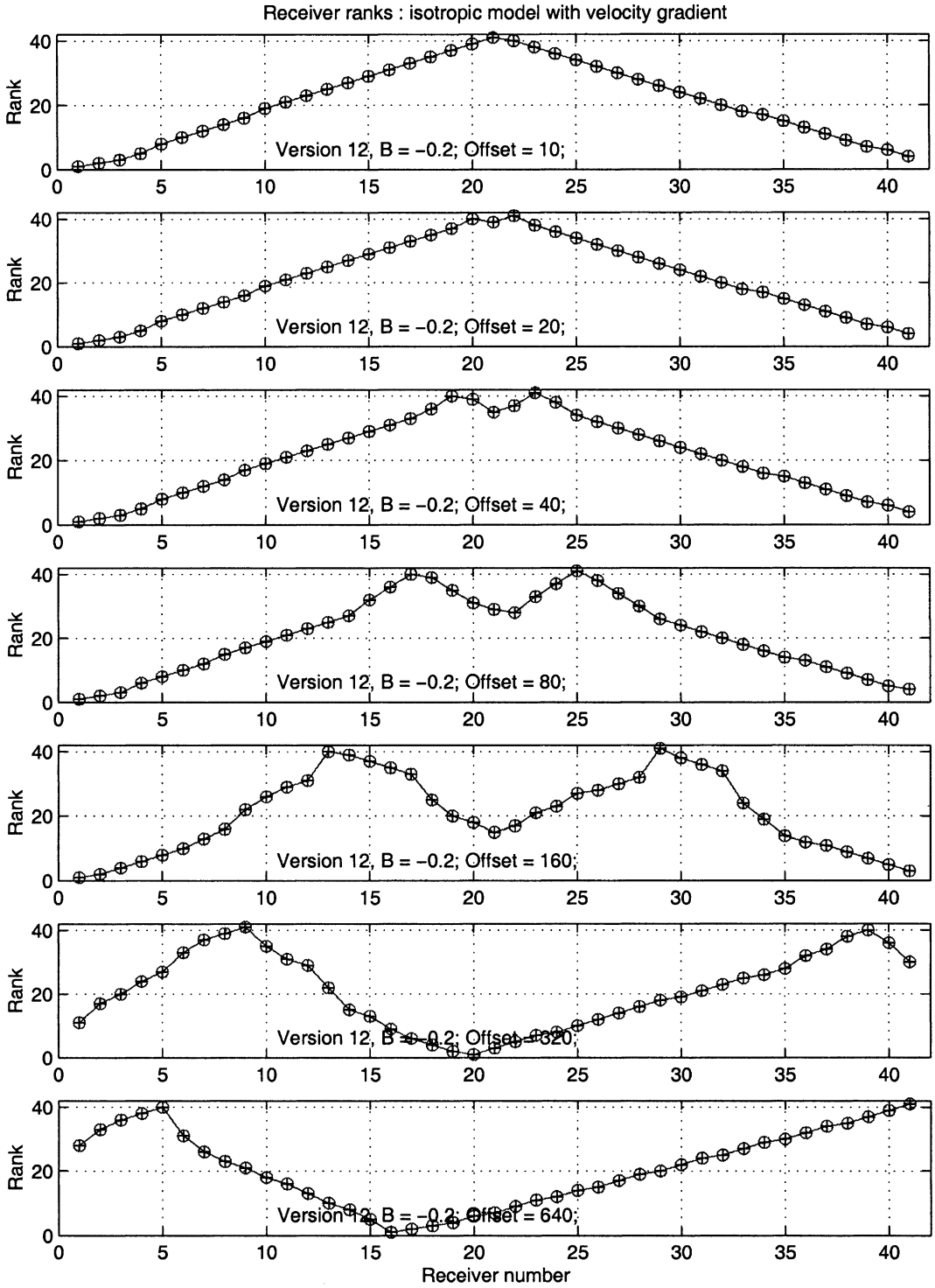

Figure 7. Results obtained for seven independent runs of the design algorithm using the linear (constant gradient) background velocity model shown in Fig. 4, fixed attenuation ($B = -0.2$), and using only a single event in each run. From top to bottom events were at 10 m, 20 m, 40 m, 80 m, 160 m, 320 m and 640 m from the well.

asymmetric with respect to the reservoir depth. Receivers should be placed preferentially at deeper locations than they would be in the homogenous velocity field. This makes sense because seismic energy arrives at deeper receivers more quickly as a result of the higher velocities at depth. Hence, seismic waves of each frequency oscillate through fewer cycles and attenuate less, resulting in decreased uncertainties relative to shallower receiver locations. In addition, the deeper receivers provide a relatively larger angular coverage of ray paths at the sources as a result of the curvature of rays in the gradient velocity field.

Fig. 7 shows the results obtained with a velocity gradient and $B = -0.2$ when each of the seven events is included individually in the design process. Consecutive plots in the series from top to bottom show optimal designs for locating events at increasing distance from

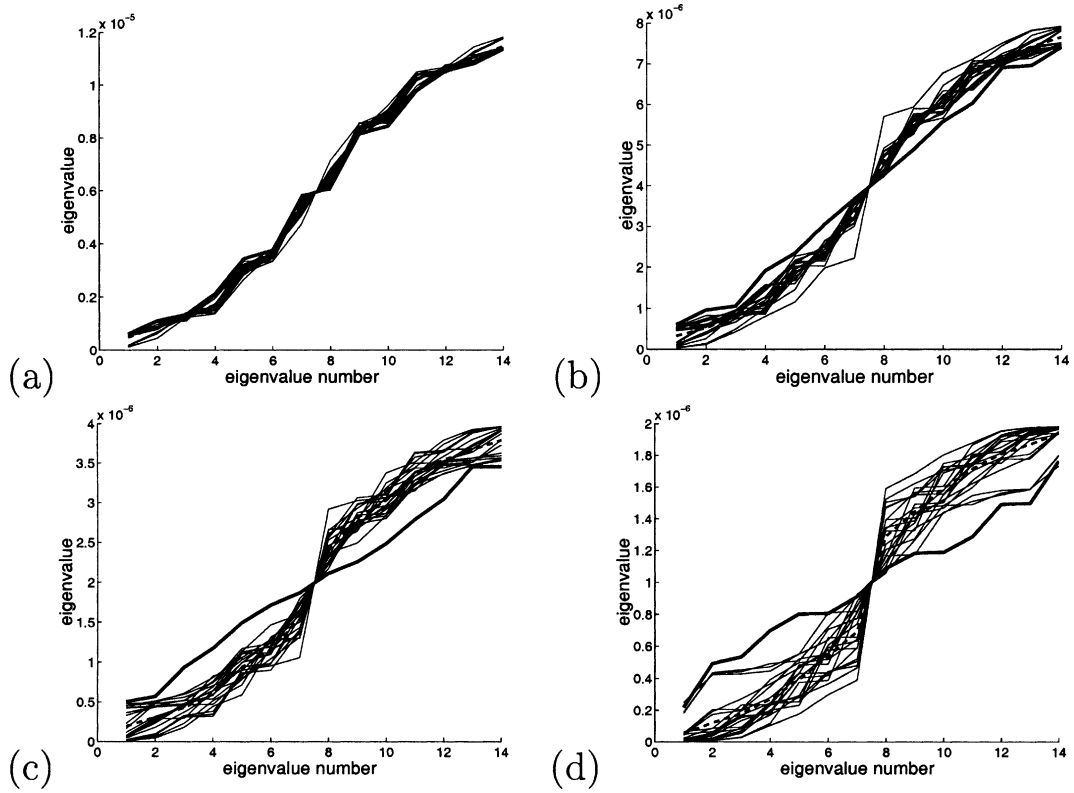

Figure 8. Plots (a), (b), (c) and (d), respectively, show eigenvalue spectra after 11, 21, 31 and 36 receivers are removed from the microseismic design shown in the lowermost plot in Fig. 5. Thin lines show 20 realizations of spectra after a random set of 11, 21, 31 and 36 receivers were removed, respectively. Dashed lines show the average of these spectra. Bold lines show spectra after the same numbers of receivers were removed in the order defined by the design in Fig. 5.

the well. For events at greater distance from the well (lower plots), we require receivers spaced further apart in order to generate the best possible fan of angles of departure of seismic energy travelling from events to the receivers. Although distantly spaced receivers would also produce the best fan for events close to the well, this trades off with the fact that receivers at greater distance are likely to obtain data with higher uncertainties. It is possible to obtain almost equally good fans using less distant receivers, which produce higher quality data. For events at only 10 m from the well (top plot) it is best to locate receivers as close as possible to the reservoir depth.

To compare these results with those of the tomography example presented earlier, Fig. 8 presents eigenvalue spectra of $\mathbf{A}^T\mathbf{A}$ for four of the designs from the lowermost plot of Fig. 5. The designs analysed in each of the four plots consist of the receivers remaining after 11, 21, 31 and 36 receivers were removed, respectively. In this particular case there was no weighting of the model parameters and the data uncertainties were all constant (no anelastic attenuation occurred because $B = 0$). Hence, similar to the tomography example, all information relevant to the experimental design is contained within matrix $\mathbf{A}$ so analysis of the eigenvalues of $\mathbf{A}^T\mathbf{A}$ makes sense. Similarly to Figs 2 and 3, spectra from 20 random designs are also shown in Fig. 8.

Again we see that in all cases the design algorithm maximizes small eigenvalues. However, in the current example this trades off strongly with minimizing large eigenvalues. Indeed, there is an almost exact symmetry between the small and large eigenvalues, with a reflection point at 7.5 on the eigenvalue number axis and the average of the maximum and minimum eigenvalues on the eigenvalue axis. All eigenvalue spectra pass through this point.

To analyse this curious behaviour further, in Fig. 9 we show the eigenvalue and eigenvector spectrum for the particular design after 31 receivers have been removed (lower-left plot in Fig. 8). Symmetry in the eigenvalue spectrum is reflected in symmetry in the eigenvector spectrum. Each eigenvector is shown in the lower plot as a column vector with each row representing a model parameter. The model parameters are ordered (x_1, z_1), (x_2, z_2), ... where (x_i, z_i) are the horizontal and vertical coordinates of event i, and event numbers increase away from the well. Hence, each eigenvector consists of a linear combination of the x and z coordinates of a single event.

This experimental design will provide least information (smallest eigenvalue) about the x-coordinate of the closest event, and provides most information (largest eigenvalue) about the z-coordinate of the same event. The point at 7.5 between eigenvalues 7 and 8 is the point at which we switch from eigenvectors that are sensitive mainly to event x-coordinates to those sensitive mainly to event z-coordinates. The trade-off through the point at (7.5, 2) observed on the lower-left plot of Fig. 8, therefore, approximately represents a trade-off between constraining x and z coordinates of events. Furthermore, it appears that this trade-off is a necessary feature of any design that is a subset of the full experimental geometry shown in Fig. 4 because it appears on all of the random designs shown in Fig. 8. Therefore, in this particular experiment it is possible to change the design to better constrain x-coordinates of any events at the expense of their z-coordinates, or vice versa, but it is not possible to better constrain both x- and z-coordinates of any events simultaneously.

5 DISCUSSION

The examples presented above show that in cases that are sufficiently simple for us to have intuition about expected results, the algorithm indeed produces results consistent with this intuition. However, our

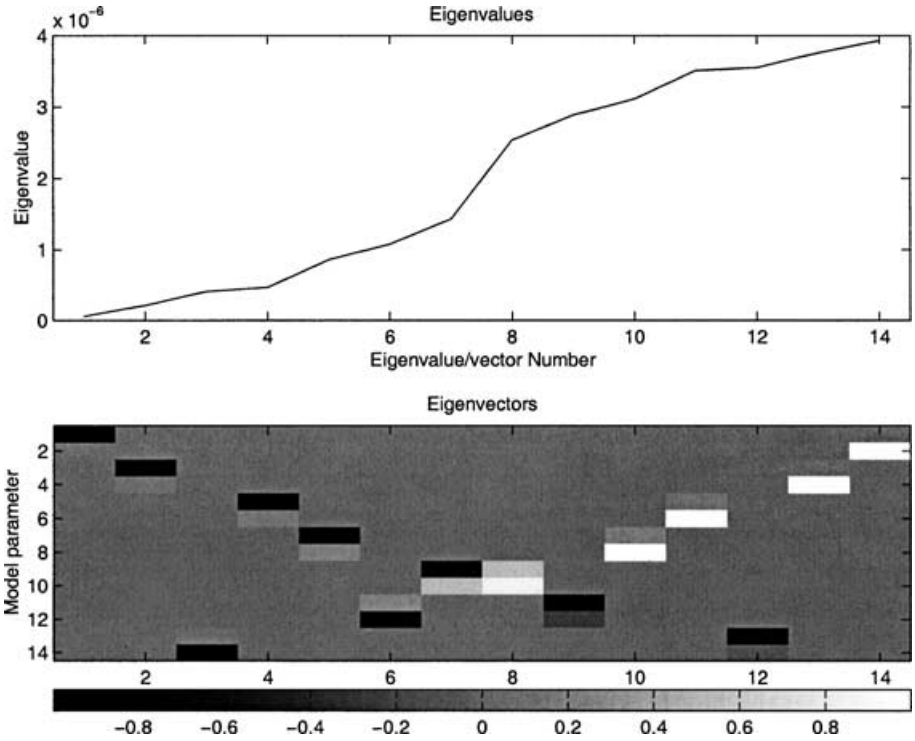

Figure 9. Eigenvalues (top plot) and eigenvectors (bottom plot) from the design found by the new algorithm after 31 receivers were removed (corresponding to the lower-left plot in Fig. 8). On the lower plots eigenvectors are plotted as columns with colour-coded values.

intuition is generally only qualitative whereas the algorithm presented here provides quantitative experimental designs (would we be able to guess intuitively exactly how asymmetric in depth the microseismic monitoring design should be for a given linear velocity gradient?). In more complicated velocity structures with several high- and low-velocity layers and, hence, potentially highly non-linear ray paths, we have little hope of applying intuition and are forced to use quantitative design algorithms to design both tomographic and microseismic monitoring surveys.

The algorithm presented here can be applied to any linear or linearized experimental design problem, is simple to use and converges to a unique, deterministic result that is influenced by the range of all possible experimental designs. It results in designs that significantly increase small eigenvalues without ever having to calculate those eigenvalues. Such designs are desirable if we expect our data quality to be sufficiently high that these eigenvalues and the associated eigenvectors will not be removed from the inverse problem. Our method, therefore, complements the suite of previous algorithms described in the introduction that have different key features.

Although in the examples contained in this paper we have designed surveys to provide either slowness structure from tomography or fracture locations from triangulation methods, it is simple to include both types of model parameters within the vector **m**. In this way, we might design a survey that provides information both about the fracture locations and the seismic velocity structure of the Earth in the vicinity of the survey and events. In such cases, matrix **A** would become partitioned (e.g. Menke 1989). Both this partitioning and the (typical) sparseness of matrix **A** in tomographic problems allow highly efficient, sparse matrix algorithms to be used. This would result in rapid computation of quality functions and their updates as receivers are removed from the design. This is important for practical survey design problems that may have many more data and model parameters than those considered above.

The current algorithm contains an underlying assumption that to within expected data uncertainties the relationship between model parameters and data can be approximated adequately by a relationship that is linearized around each model parameter value, similar to most of the other studies mentioned in the introduction. Generally speaking, such pseudo-linearity should only be assumed for the design problem if making the same assumption post-experiment would allow linearized inverse theory to be used robustly to constrain model parameters from the data in a single inversion step (i.e. without requiring iterated linearization and inversion).

If in reality this relationship depends on the model parameter values themselves in a more non-linear fashion, or on the values of other parameters that are intrinsic to the model–data relationship, this assumption could be relaxed slightly in the following Bayesian sense: the value of each receiver calculated in the above algorithm is replaced by the average value over the distribution of possible parameter values, where this distribution reflects our prior expectations of what the parameter values might be. Rankings then reflect the average performance of each receiver across this distribution of parameters. In the microseismic example above, for instance, we might average the range of quality functions obtained when a range of possible background seismic velocity structures were used. Such a design process is referred to in the statistics literature as Bayesian or non-linear experimental design (e.g. Atkinson & Donev 1992; Maurer & Boerner 1998).

However, this technique can only take account of mild non-linearity: specifically, only non-linearity that does not cause multiple, disconnected regions of model space to provide good fits to observed data (Curtis & Spencer 1999). Sometimes non-linearity encountered in real problems will not satisfy these conditions. In such situations, more computationally intensive methods must be employed (Curtis & Spencer 1999; van den Berg et al., 2004).

REFERENCES

Atkinson, A.C. & Donev, A.N., 1992. *Optimum experimental designs,*. Clarendon Press, Oxford.

Curtis, A., 1999a. Optimal experiment design: Cross-borehole tomographic examples, *Geophys. J. Int.,* **136,** 637–650.

Curtis, A., 1999b. Optimal design of focussed experiments and surveys, *Geophys. J. Int.,* **139,** 205–215.

Curtis, A., 2000. Optimizing the design of geophysical experiments: Is it worthwhile?, *EOS, Trans. Am. geophys. Un.,* Forum Article, **81**(20), 224–225.

Curtis, A. & Snieder, R., 1997. Reconditioning inverse problems using the genetic algorithm and revised parameterization, *Geophysics,* **62**(5), 1524–1532.

Curtis, A. & Spencer, C., 1999. Survey design strategies for linearized, nonlinear inversion, in: *Extended Abstracts,* 69th Ann. Internat. Mtg. Soc. of Expl. Geophys., pp. 1775–1778.

Curtis, A. & Wood, R., 2004. Optimal elicitation of prior information from experts, in *Geological Prior Information,* eds Curtis, A. & Wood, R., Geol. Soc. London. Special Publication, in press.

Kijko, A., 1977. An algorithm for the optimum distribution of a regional seismic network—I, *Pageoph,* **115,** 999–1009.

Maurer, H. & Boerner, D.E., 1998. Optimized and robust experimental design: a non-linear application to em sounding, *Geophys. J. Int.,* **132,** 458–468.

Maurer, H., Boerner, D.E. & Curtis, A., 2000. Design strategies for electromagnetic geophysical surveys, *Inverse Problems,* **16**(5), 1097–1117.

Menke, W., 1989. *Geophysical data analysis: Discrete inverse theory,* Revised edn), Vol. 45, International Geophysics Series, Academic Press Inc., Harcourt Brace Jovanovich Publishers, San Diego.

Mitchell, T.J., 1974. An algorithm for the construction of 'd-optimal' experimental designs, *Technometrics*,**16**(2), 203–210.

Rabinowitz, N. & Steinberg, D.M., 1990. Optimal configuration of a seismographic network: a statistical approach, *Bull. seism. Soc. Am.*, **80**(1), 187–196.

Sabatier, P.C., 1977. On geophysical inverse problems and constraints, *J. Geophys.,* **43,** 115–137.

Silvey, S.D., 1980. *Optimum design,* Chapman and Hall, London.

Smith, M.L., Scales, J.A. & Fischer, T.L., 1992. Global search and genetic algorithms, *The Leading Edge,* **11**(1), 22–26.

Steinberg, D.M., Rabinowitz, N., Shimshoni, Y. & Mizrachi, D., 1995. Configuring a seismographic network for optimal monitoring of fault lines and multiple sources, *Bull. seism. Soc. Am.,* **85**(6), 1847–1857.

Tarantola, A., 1987. *Inverse problem theory,* Elsevier Science Publishers B. V., Amsterdam.

Tarantola, A. & Valette, B., 1982. Inverse problems = quest for information, *J. Geophys.,* **50,** 159–170.

van den Berg, J., Curtis, A. & Trampert, J., 2003. Optimal, non-linear, bayesian experimental design with 1-D examples, *Geophys. J. Int.,* **155,** 411–421.