Geophysical Journal International

Geophys. J. Int. (2020) **223,** 944–958
Advance Access publication 2020 July 29
GJI General Geophysical Methods

doi: 10.1093/gji/ggaa358

Experimental design for fully nonlinear source location problems: which method should I choose?

H. Bloem [1], A. Curtis[1] and H. Maurer[2]

[1] *School of Geosciences, University of Edinburgh, Edinburgh EH8 9XP, UK. E-mail: hugo.bloem@ed.ac.uk*
[2] *Department of Earth Sciences, ETH Zürich,* 8092 *Zürich, Switzerland*

Accepted 2020 July 23. Received 2020 July 23; in original form 2020 April 7

SUMMARY

Statistical experimental design (SED) is the field of statistics concerned with designing experiments to obtain as much information as possible about a target of interest. SED algorithms can be divided into two categories: those that assume a linear or linearized relationship between measured data and parameters, and those that account for a fully nonlinear relationship. We compare the most commonly used linear method, Bayesian D-optimization, to two nonlinear methods, maximum entropy design and D_N-optimization, in a synthetic seismological source location problem where we define a region of the subsurface in which earthquake sources are likely to occur. Example random sources in this region are sampled with a uniform distribution and their arrival time data across the ground surface are forward modelled; the goal of SED is to define a surface monitoring network that optimally constrains this set of source locations given the data that would be observed. Receiver networks so designed are evaluated on performance—the percentage of earthquake pairs whose arrival time differences are above a threshold of measurement uncertainty at each receiver, the number of prior samples (earthquakes) required to evaluate the statistical performance of each design and the SED compute time for different subsurface velocity models. We find that D_N-optimization provides the best results both in terms of performance and compute time. Linear design is more computationally expensive and designs poorer performing networks. Maximum entropy design is shown to be effectively impractical due to the large number of samples and long compute times required.

Key words: Statistical methods; Computational seismology; Earthquake source observations.

1 INTRODUCTION

Over the past 50 yr, a variety of techniques have been introduced to find optimal designs for geophysical experiments (Curtis 2004a,b; Maurer *et al.* 2010). These techniques stem from a field of statistics called statistical experimental design (SED) and were initially developed to optimize industrial processes (Cox 1958; Kackar 1985). The first application in geophysics was by Kijko (1977) who used SED to find optimal receiver locations in a source localization problem. Since then, SED has been used to design a variety of source location experiments (Steinberg *et al.* 1995; Curtis *et al.* 2004; Winterfors & Curtis 2008; Toledo *et al.* 2018), surveys for seismic tomography (Curtis 1999a,b; Sirgue & Pratt 2004; Maurer *et al.* 2009), electromagnetic and electrical resistivity tomography (Stummer *et al.* 2004; Coles & Morgan 2009; Ren & Kalscheuer 2019) and for CO_2 monitoring (Romdhane & Eliasson 2018). SED methods are also used to minimize the data requirements for other algorithms. For example, Maurer *et al.* (2017) use experimental design to limit the data requirements for full-waveform inversion (FWI) by using SED to select the most informative subset from a data set that contains as much information about the subsurface target as possible. Guest & Curtis (2010a,b, 2011) optimized the subset of source-to-receiver offset that should be used from an active-source seismic survey so as to preserve amplitude-versus-offset (AVO) information about subsurface reflector properties. Thus, in each case fewer data can be processed to obtain similar information.

SED algorithms can be divided into two categories: those that assume a linear or linearized relationship between measured data and parameters, and those that account for fully nonlinear relationships. Linear methods are typically presumed to be easier to compute, while nonlinear methods are presumed to be more accurate. A comparison between these categories of SED methods has not been made for any geophysical problem, which is the goal of this work. We compare such methods for seismic source localization problems, and the research herein is the first to test and recommend which class of SED methods to use.

Source locations are used in various fields of seismology (Tong *et al.* 2016), for example, when studying tectonic processes and earthquake dynamics (Waldhauser & Ellsworth 2000) or when studying how earthquakes are related to structural features (Huang

& Zhao 2013; Lin 2013). Source localization techniques are also used to resolve political issues, for example, when locating and discriminating nuclear tests (Wen & Long 2010; Zhang & Wen 2014). In exploration and production for subsurface earth resources, locations of induced seismicity are of importance to check for possible leaks in the cap-rock (Wuestefeld *et al.* 2018) or to attribute seismicity to anthropogenic activity such as shale gas production (Wilson *et al.* 2015). These applications all require a receiver network to record the seismic events, and hence methods to design experiments that acquire data with the most information content.

We compare one linearized design method, Bayesian D-optimization (Box & Lucas 1959; Kijko 1977; Steinberg *et al.* 1995), and two nonlinear methods, maximum entropy design (Shewry & Wynn 1987; van den Berg *et al.* 2003, 2005; Guest & Curtis 2009, 2010b, 2011) and D_N-optimization (Coles & Curtis 2011). Bayesian D-optimization is the most commonly used linearized method that makes this research relevant to a broad range of SED users. Maximum entropy design is the only criterion that directly maximizes the Shannon information about the parameters of interest while using fully nonlinear physics. D_N-optimization is an approximation to maximum entropy design that is computationally efficient and does not assume linear relationships.

Optimal designs are calculated using each method in a synthetic source localization example. We compare the performance of the generated networks in terms of the remaining source location ambiguity, evaluating how many example sources in the subsurface region of interest are needed by each method to obtain a stable design, and the overall computation time required for each method. Thus, we draw conclusions on which method to use under various scenarios.

The next section introduces source location problems, then we describe inversion and optimal design theory in general, sequential design strategies, and the above three specific SED methods. Thereafter we explain the experimental methodology used to obtain the test results, followed by the results themselves. Finally, we discuss the implications of this work and conclude.

2 SOURCE LOCATION PROBLEM

In source location problems, we wish to establish the location of a source (usually an earthquake) in the subsurface. This is done by recording the arrival times of seismic waves on a receiver network; differences in arrival times at different receivers allow us to estimate the source location given the subsurface velocity structure.

There are usually four model parameters in a source location problem: the x, y and z location, and the time t_0 when the earthquake occurred. The location parameters are described by vector $\mathbf{m}$ and the data are differences in arrival times $\mathbf{d}$. The relation between data $\mathbf{d}$ and parameters $\mathbf{m}$ is a mathematical function

$$\mathbf{d} = \mathbf{F}_S(\mathbf{m}), \tag{1}$$

where $\mathbf{F}_S$ is the nonlinear forward function that calculates the arrival time differences at the receivers of seismic energy from the source. Note that we only consider $\mathbf{m}$ as independent parameters of $\mathbf{F}_S$: subsurface seismic velocities are considered implicit and fixed within the forward function. The subscript S denotes the design of the receiver network, and in this work the design consists of locations of a network of seismometers used to record arrival times.

To avoid including t_0 in the set of model parameters, there are at least two possibilities: we can subtract the mean arrival time from the arrival times across the group of stations. Or, we can consider only arrival time differences between the first arriving S and P waves (t_s and t_p). We use the latter and assume a constant v_p/v_s ratio (i.e. common ray paths L for P and S waves between the same source and receiver), the traveltime difference can be written as

$$t_s - t_p = \int_{r \in \mathbf{L}(S,\mathbf{m})} \frac{1}{v_s(r)} - \frac{1}{v_p(r)} \mathrm{d}r, \tag{2}$$

where r is a position vector and the integration is carried out along L, and where $\mathbf{m}$ contains only the location of the source.

3 THEORY

3.1 Introduction to inverse theory

Consider some recorded arrival time data $\mathbf{d}^{\mathrm{obs}}$ from which we want to infer the values of parameters $\mathbf{m}$. Assume that parameters $\mathbf{m}$ are linked to synthetic data $\mathbf{d}^{\mathrm{syn}}$ by the forward model $\mathbf{d}^{\mathrm{syn}} = \mathbf{F}_S(\mathbf{m})$. We relate $\mathbf{d}$ to $\mathbf{m}$ by finding the likelihood of observing data $\mathbf{d}$ given the parameter values $\mathbf{m}$: $\rho(\mathbf{d}|\mathbf{m})$, which is also known as the likelihood function (Tarantola 2005). $\rho(\cdot)$ denotes the probability density function (pdf) of a continuous variable or the probability distribution function of a discrete event (e.g. earthquake) being in a particular set, and $\rho(\mathbf{d}|\mathbf{m})$ is the probability of observing data $\mathbf{d}$ in the current experiment given that model $\mathbf{m}$ is true. The form of the likelihood therefore depends on the uncertainty in recorded data $\mathbf{d}$, since it must define how likely it is that data $\mathbf{d}^{\mathrm{syn}}$ *might have* been recorded in the experiment.

Before conducting an experiment we already have some information about the parameters. This is described by a so-called prior pdf $\rho(\mathbf{m})$. Possible parameter prior information might include that there is a higher probability that earthquakes occur close to a known fault or that earthquakes always occur below ground level.

To combine the information on parameter space M and data space D we use the fact that in $M \times D$, $\rho(\mathbf{m}, \mathbf{d}) = \rho(\mathbf{m}|\mathbf{d})\rho(\mathbf{d}) = \rho(\mathbf{d}|\mathbf{m})\rho(\mathbf{m})$ from which we derive:

$$\rho(\mathbf{m}|\mathbf{d}) = \frac{\rho(\mathbf{d}|\mathbf{m})\rho(\mathbf{m})}{\rho(\mathbf{d})}. \tag{3}$$

This equation is known as Bayes' theorem for probability densities (Bayes 1763) and it gives the so-called posterior pdf $\rho(\mathbf{m}|\mathbf{d})$ in terms of the likelihood $\rho(\mathbf{d}|\mathbf{m})$, the prior $\rho(\mathbf{m})$ and a normalization factor $\rho(\mathbf{d})$ that is constant for any particular fixed data set. The normalization factor, called the evidence, is given by

$$\rho(\mathbf{d}) = \int \rho(\mathbf{d}|\mathbf{m})\rho(\mathbf{m}) d\mathbf{m} \tag{4}$$

3.2 Introduction to optimal design theory

To better understand the optimization of $\mathbf{F}_S$ consider the schematic in Fig. 1. The horizontal-axis depicts a parameter, the vertical-axis a datum, and lines $\mathbf{F}_a$ (Fig. 1a) and $\mathbf{F}_b$ (Fig. 1b) represent $\mathbf{F}_S$ for two different receiver network designs, denoted a and b. The blue area represents the recorded datum d^{obs} and its measurement uncertainty $\pm\sigma$. The red areas represent the corresponding uncertainty on the model parameter if we infer its range of possible values from the data using the two different designs. We see $\mathbf{F}_a$ produces greater parameter uncertainty than $\mathbf{F}_b$, which means that we can better constrain the parameter with design b, despite making the same nominal experimental effort (the same number of data with the same σ) for both designs.

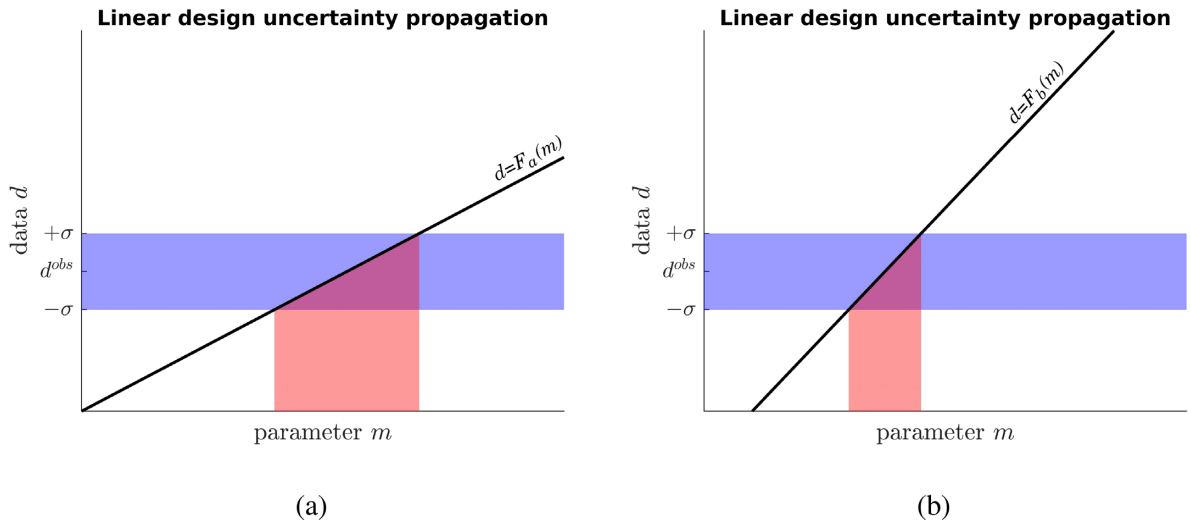

Figure 1. Two representations of the parameter–data relationship $\mathbf{F}_S$ for two receiver network designs a and b. The blue area represents a datum d^{obs} with its corresponding measurement uncertainty $\pm\sigma$. The red areas show the inferred projected parameter uncertainty for designs a and b.

Because the model–data relations shown in Fig. 1 are linear, it is relatively easy to infer a heuristic rule for optimal design: a steeper gradient of $\mathbf{F}_S$ is more favourable as this reduces the resulting uncertainty on the parameters. This logic does not hold for nonlinear functions per se as nonlinear functions can have a different gradient for each value of $\mathbf{m}$; it is therefore difficult to devise a simple rule of thumb, and to find optimal designs for nonlinear relations we need to use formal SED methods (van den Berg *et al.* 2003, 2005).

3.3 Sequential design

If multiple receivers are used in a network, we would ideally calculate the quality of information for every possible combination of receiver locations within a specified set (here a grid), then select the best as the survey design. In our experiment below, 3.67×10^{18} possible designs exist, which rules out an exhaustive search. Additionally, the optimization of receiver coordinates is usually difficult due to the strongly nonlinear relationship between receiver placement and the expected information. Therefore, nonlinear algorithms must be used to optimize receiver placement [previously genetic algorithms (Curtis 1999a,b) or simulated annealing (Barth & Wunsch 1990), for example].

Curtis *et al.* (2004) reduce the dimensionality of the search by starting with a design containing all possible receiver locations, then reducing the design by sequentially deleting the least informative receiver until a practical number of receivers is reached. Guest & Curtis (2009) reduce computational demand still further by adding receivers one at a time to a design, starting from the first receiver. The latter algorithm is referred to as sequential design in the statistical literature (Atkinson *et al.* 2007). In this algorithm, the first receiver is placed in the location that is expected to provide most information and is thereafter fixed at that location. In each subsequent iteration, the algorithm searches for where to add another single receiver to the network for optimal effect, given the receiver locations already fixed in the network. This means that for each receiver a 1-D domain is searched (the vector of possible receiver locations), thus reducing the search so as to scale proportionally to the number of receivers. This method results in only $3721 \times 6 = 22\,326$ design combinations to be searched. Mathematically, to find an optimal design consisting of N receiver locations the following schema is used herein:

$$\begin{aligned} S_1 &= \arg_{S_1} \max[\Phi(S_1)] \\ &\vdots \\ S_i &= \arg_{S_i} \max[\Phi(S_i|S_1, \cdots, S_{i-1})] \\ &\vdots \\ S_N &= \arg_{S_N} \max[\Phi(S_N|S_1, \cdots, S_{N-1}] \end{aligned} \tag{5}$$

where $S = \{S_1, \cdots, S_N\}$ is the vector containing the N optimal receiver locations, Φ is a quality measure for each receiver location and $\arg_{S_i} \max[\cdot]$ is a mathematical function that denotes the receiver location S_i where the quality is maximized for receiver number i. While not all possible designs are evaluated, the method does take into account previously fixed receiver locations when assigning a new location. This method greatly reduces the number of designs to be tested for optimality. However, since at each stage only one receiver position is optimised (we do not re-adjust previously located receivers in the light of information from the new receiver), it is possible that a local quality maximum is achieved rather than a global maximum. It has been postulated that the sequential deletion approach of Curtis *et al.* (2004) is less susceptible to this problem since at iteration 1 it considers all possible receiver locations that are never considered in the sequential addition approach of Guest & Curtis (2009). However, in early iterations, the former method requires quality to be evaluated in a far higher dimensionality of data space, which we show below is impractical for at least one quality measure. We therefore apply the sequential method of Guest & Curtis (2009), defined in eq. (5), for all tests herein.

3.4 Linear design: D-optimization

In Fig. 1, we saw that a general rule for optimal design could be to maximize the gradient in $\mathbf{F}_S$ as this leads to a smaller parameter uncertainty for a given data uncertainty. This is the idea behind the D-optimal linearized design method introduced to statistics by Box & Lucas (1959), to geophysics by Kijko (1977) and the extension for multiple sources by Steinberg *et al.* (1995).

In D-optimization, we assume the nonlinear model parameter–data relationship $\mathbf{F}_S$ to be locally linear around $\mathbf{m}$. Eq. (1) can then be written as a linear system of equations

$$\mathbf{d} = \mathbf{X}_S\mathbf{m} + \epsilon, \tag{6}$$

where $\mathbf{X}_S$ is the so-called planning or design matrix and consists of partial derivatives of data with respect to model parameters, the subscript S indicates that the elements are functions of receiver locations and ϵ is the data error which is assumed to follow a Gaussian distribution. In experimental design, the model parameters are chosen from the prior and the data are then forward modelled; by perturbing $\mathbf{m}$ we can construct finite-difference approximations of the gradient to calculate $\mathbf{X}_S$. Alternatively, ray theory could be applied to obtain analytical partial derivatives (Cerveny 2005).

When parameters $\mathbf{m}$ are estimated by inverting eq. (6) in a least-squares sense, their covariance matrix is

$$\mathbf{D} = \sigma^2(\mathbf{X}_S^T\mathbf{X}_S)^{-1}, \tag{7}$$

where σ^2 is variance of the measurement uncertainty of each datum recorded by a receiver which, for the sake of this comparative synthetic study, may be taken constant. A point source is therefore efficiently monitored by a network if $\mathbf{X}_S^T\mathbf{X}_S$ is 'large' in some sense so that $(\mathbf{X}_S^T\mathbf{X}_S)^{-1}$ is 'small' (Steinberg *et al.* 1995). An appropriate measure to use is the matrix determinant

$$\Phi^S = \arg_S \max[|\mathbf{X}_S^T\mathbf{X}_S|], \tag{8}$$

where $|\cdot|$ denotes the determinant, which is then maximized. This is appropriate because the volume of parameter space defined by covariance matrix $\mathbf{D}$ and corresponding to any given confidence level is proportional to $|\mathbf{X}_S^T\mathbf{X}_S|^{-1}$ (Box & Lucas 1959); by minimizing the determinant with respect to the receiver locations we therefore minimize this volume.

Sequential design can be carried out efficiently by noting that if a single receiver location i would be added to the network of design S, matrix $\mathbf{X}_S$ is augmented by a single row, and so the determinant of the augmented matrix becomes

$$|\mathbf{X}_S^T\mathbf{X}_S|(1 + v_i), \tag{9}$$

where the increase of the determinant is

$$v_i = \boldsymbol{f}_i^T(\mathbf{X}_S^T\mathbf{X}_S)^{-1}\boldsymbol{f}_i \tag{10}$$

with $\boldsymbol{f}$ the partial derivatives of the arrival time with respect to the parameters for the new receiver location i. Instead of maximizing the full determinant for each location on the grid, we can therefore maximize v_i (Steinberg *et al.* 1995). This is more efficient because once $(\mathbf{X}_S^T\mathbf{X}_S)^{-1}$ has been calculated then the quality criterion for every potential receiver location i is found by matrix-vector multiplications.

However, this method only works for a single source point. We would like to design a survey that would perform well for many different possible event locations. Steinberg *et al.* (1995) extended the work of Kijko (1977) to make eq. (8) suitable for the case with multiple sources:

$$S = \sum_{j=1}^{k} a_j \ln(|\mathbf{X}_{S,j}^T\mathbf{X}_{S,j}|), \tag{11}$$

referred to as the D-optimum for multiple source (DMS) criterion. The summation is over all source locations that are considered within the design with a relative importance factor a_j for the jth source. The importance factor can be used to prioritize optimization for some sources over others, where a larger factor means that the resulting design will better constrain source locations around that site. If the a_j are assigned to the prior probability of source j, then the sum in eq. (11) approximates the expected value of the logarithm term. Using that criterion to design a survey is called Bayesian (linearized) design (Chaloner & Verdinelli 1995) and is the method used in this paper. For brevity, we refer to this simply as D-optimization from hereon.

As for single sources, we can calculate the quality criterion more efficiently for sequential design by only looking at the change in DMS when a single receiver location i is added. This results in the following equations:

$$\mathrm{DMS}_i = \textstyle\sum_{j=1}^{k} a_j \ln[|\mathbf{X}_{S,j}^T\mathbf{X}_{S,j}|(1 + v_{i,j})] \tag{12}$$

$$= \textstyle\sum_{j=1}^{k} a_j \ln[|\mathbf{X}_{S,j}^T\mathbf{X}_{S,j}|] + \sum_{j=1}^{k} a_j \ln[(1 + v_{i,j})], \tag{13}$$

where

$$v_{i,j} = \boldsymbol{f}_{i,j}^T(\mathbf{X}_{S,j}^T\mathbf{X}_{S,j})^{-1}\boldsymbol{f}_{i,j}\,. \tag{14}$$

Subscripts i and j denote the receiver location i being evaluated and the source location j. The second term of eq. (13) is the quality measure and will be evaluated for every receiver location on the grid; the location with the largest value is the location at which a receiver is added to the network.

At the beginning of the sequential design process there will be fewer receivers in the network than parameters; hence, the problem is underdetermined and the determinant is zero. Therefore, in our design algorithm the first three receivers are optimized together using the DETMAX algorithm (Mitchell 1974). DETMAX generally starts with a random network of n receiver locations. An optimal location $n + 1$ is added according to eq. (13). The $n + 1$ network is reduced to n again by removing the location that contributes the least information to the network. This procedure is iterated until a sufficiently high quality network is found. In our linear case, we use DETMAX with $n = 3$ to find the first three receiver locations and use sequential design as described above thereafter starting with S_4 in eq. (5). We note that this may give the linear design method an advantage over the other two methods since DETMAX may find a globally optimized solution for receivers 1–3 while the sequential design of the other methods might not.

3.5 Nonlinear design: maximum entropy design

In SED, we want to design an experiment such that the information I about the parameters $\mathbf{m}$ is maximized. All information about $\mathbf{m}$ is contained in the posterior pdf: we therefore need to quantify the information in that distribution. The entropy of a pdf of any random variable $\boldsymbol{y} \in \mathbf{Y}$ is related to Shannon's measure of information I (Shannon 1948) as

$$\mathrm{Ent}(\mathbf{Y}) = -\int_{\mathbf{Y}} g(\mathbf{y}) \log(g(\mathbf{y}))\mathrm{d}\mathbf{y} = -I(g(\mathbf{y})) + c_1, \tag{15}$$

where Ent is the entropy function, $g(\mathbf{y})$ is the pdf of $\mathbf{y}$ and c_1 is a constant (Guest & Curtis 2009). A fully nonlinear quality measure (Lindley *et al.* 1956) is therefore

$$\Phi(S) = -\int_D \mathrm{Ent}(\rho(\mathbf{m}|\mathbf{d}, S))\rho(\mathbf{d}|S)\mathrm{d}\mathbf{d}, \tag{16}$$

where $\rho(\cdot\,|S)$ represents the dependence on the receiver network design of both the posterior and the evidence. This quality measure is the information content in the posterior distribution over all data that might be recorded by a sensor network S. Calculating

this quality measure directly would require that we know the posterior distribution $\rho(\mathbf{m}|\mathbf{d}, S)$ for all possible sets of data $\boldsymbol{d}$, which is computationally intractable. However, for a fixed number of data,

$$-\Phi(S) + Ent(\rho(\mathbf{d}|S)) = c_2 \tag{17}$$

[Coles & Morgan (2009), after Shewry & Wynn (1987)] for some constant c_2, which means that instead of maximizing $\Phi(S)$ for each new receiver we can instead maximize $\mathrm{Ent}(\rho(\mathbf{d}|S))$. The evidence $\rho(\mathbf{d}|S)$—which describes the probability of data $\mathbf{d}$ being recorded—requires data $\mathbf{d}$ to be calculated from all possible sets of parameters $\mathbf{m}$ (or a representative subset of those) using $\mathbf{F}_S$, which is computationally more demanding than calculating the posterior pdf. Since the evidence is not known analytically, we sample it stochastically to calculate a numerical approximation. To do this, samples of $\mathbf{m}$ are drawn from the prior pdf and forward modelled through $\mathbf{F}_S$; the resulting set of discrete sampled data is used for entropy estimation as shown next. Note that if too few samples are used in entropy estimation the results become inaccurate, so it is necessary to assess the number of samples needed to obtain accurate results; in what follows we increase the number of samples until the result becomes stable.

3.5.1 *Entropy estimation*

The method to estimate entropy used here was introduced by Stowell & Plumbley (2009). For a random variable $y \in Y$. Let A be a partition of Y with $A = \{A_j | j = 1...n\}$, $A_j \cap A_k = \emptyset$ and $\bigcup_j A_j = Y$. Then, we approximate the continuous Shannon entropy $H = \mathrm{Ent}(Y)$ in eq. (15) with a discrete version

$$\hat{H} = \sum_{j=1}^{m} \frac{n_j}{N} \log\left(\frac{N}{n_j}\mu(A_j)\right), \tag{18}$$

where n_j is the number of sampled data points in A_j, N is the total number of data points and $\mu(A_j)$ is the D-dimensional volume of A_j. Normally, a fixed partition width or number of partitions is used to approximate integrals, but Stowell & Plumbley (2009) partition the data based on two criteria: (1) the distribution inside each partition element must be uniform and (2) the data must be split into a minimum number of partition elements. A test for uniformity of a distribution is (Chu *et al.* 1955)

$$Z_j = \sqrt{n_j}\frac{2 \cdot \mathrm{med}_d(A_j) - \min_d(A_j) - \max_d(A_j)}{\max_d(A_j) - \min_d(A_j)}, \tag{19}$$

where $\mathrm{med}_d(A_j)$, $\min_d(A_j)$ and $\max_d(A_j)$ are the median, minimum and maximum of partition element A_j along dimension d, respectively. If $|Z_j| > 1.96$, then element A_j is considered to show a nonuniform behaviour at a significance level of 95 per cent and criterion (1) is not met. This uniformity test is weak and prone to errors. Therefore, the data have to be divided into a minimum number of partitions such that every partition has $\sqrt{N}$ data points (criterion 2). This corresponds to a branching level L_N (the number of partitions) of

$$L_N = \lceil \frac{1}{2}\log_2 N \rceil, \tag{20}$$

where $\lceil \cdot \rceil$ denotes the ceiling function. If either one of the criteria is not met then A_j is divided into two elements with the division along $\mathrm{med}_d(A_j)$.

Before this partitioning algorithm is executed, all data samples are placed in the same partition element $A = \{A_1\}$. The data are then partitioned following criteria (1) and (2) according to the following scheme:

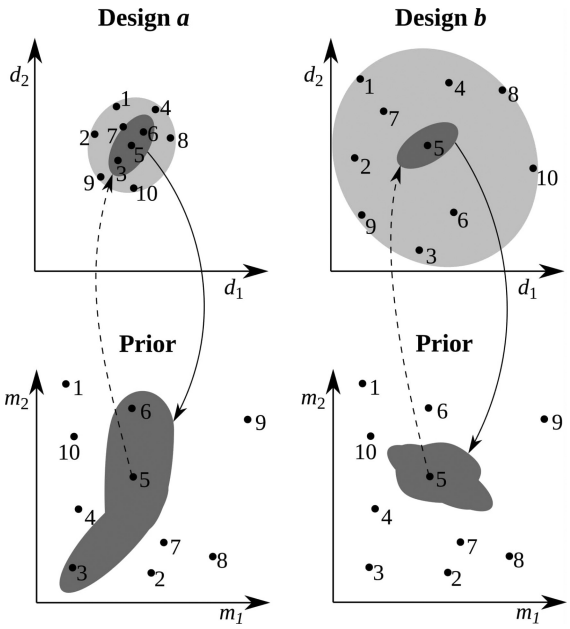

Figure 2. 10 prior model samples (bottom panels) and the data that would be observed for each model (top panels) using two different experimental designs *a* and *b* [modified after Coles & Curtis (2011)]. Top panels: the dark grey region represents expected data measurement uncertainty, in this case shown for data set 5 as an example under each design. The light grey region around the data points in each design represents the covariance of the entire set of samples, and is the measure that is maximized with D_N-optimization. Arrows: The dashed arrow represents the forward modelling operator which maps each prior model to the corresponding measured data and their expected uncertainties. The solid arrows represent inversion of the data (shown for data set 5): this maps the data uncertainties in the top panels onto the posterior model uncertainties represented by grey regions in the lower panels.

(1)Consider partition element A_j
(2)Check criteria:

(i)Check for uniformity: $|Z_j| \leq 1.96$
(ii)Check branching level: $n \geq L_N$

(3)If both criteria passed, increment j
Else, split the cell at the median along dth dimension and return to step 2 for the newly created cells.

If both criteria are met for all j then the entropy can be computed using eq. (18).

This method is called k-d partitioning; it works in high dimensions, can be computed efficiently and has low memory requirements (Stowell & Plumbley 2009), which is all advantageous as we evaluate entropy many times. An added benefit is that the partition element size is optimized by the algorithm, useful because choosing the partition element width incorrectly can result in incorrect estimates of the entropy (van den Berg *et al.* 2003, 2005).

3.6 Nonlinear design: D_N-optimization

D_N-optimization was proposed by Coles & Curtis (2011) as an approximation to maximum entropy design. Both methods are designed to maximize the entropy of the evidence, which in turn can be seen as maximizing the scatter (difference) between the sampled data points (Shewry & Wynn 1987). Coles & Curtis (2011) method is illustrated conceptually in Fig. 2, where 10 samples from the

prior are taken $\{m_i: i = 1, \ldots, 10\}$ (representing 10 earthquake locations). These samples are recorded using design a and design b, each of which records two observations d_1 and d_2 (e.g. arrival times of two receivers). The data generated by the two observations are scattered across the 2-D data space, and say the expected measurement uncertainty is shown as the grey shaded area. The latter is only shown for parameter set five, but is assumed to be predictable for all parameter sets. The design used for design a causes the data for the 10 parameter sets to be close to each other, whereas design b produces data which are more spread out (scattered). Design b is advantageous because it is easier to discriminate between the different models (different possible earthquake locations) using the measured data. Parameter set three, for example, lies within the measurement uncertainty region of parameter set five under design a meaning that earthquake locations 3 and 5 can be discriminated using design b, but this is not the case for design a. Coles & Curtis (2011) therefore argue that data for parameter samples in design a are more similar to each other than in the case for data generated by design b, so design b is preferred.

Consider two parameter sets $\mathbf{m}_i$ and $\mathbf{m}_j$ with corresponding data sets $\mathbf{d}_i$ and $\mathbf{d}_j$ distributed according to pdfs $p_i(\mathbf{d}_i)$ and $p_j(\mathbf{d}_j)$, and assume both p_i and p_j are multivariate Gaussians with the same covariance $\Sigma_d(S)$. Then, the relative entropy between the two is analytic, given by (Goldberger *et al.* 2003)

$$D(p_i||p_j) = \frac{1}{2}[\mathbf{F}_S(\mathbf{m}_j) - \mathbf{F}_S(\mathbf{m}_i)]^T \times [\Sigma_d(S)]^{-1} \times[\mathbf{F}_S(\mathbf{m}_j) - \mathbf{F}_S(\mathbf{m}_i)], \quad (21)$$

where Σ_d is the data-noise covariance. We can approximate $D(p_i||p_j)$ with (Wunsch 1996)

$$D(p_i||p_j) \approx \frac{1}{2}\delta(S)^T \Sigma(S)\delta(S), \quad (22)$$

where

$$\delta(S) = [\Sigma_{\mathbf{F}}(S)]^{-\frac{1}{2}}[\mathbf{F}_S(\mathbf{m}_j) - \mathbf{F}_S(\mathbf{m}_i)] \quad (23)$$

$$\Sigma(S) = [\Sigma_{\mathbf{F}}(S)]^{\frac{1}{2}}[\Sigma_d(S)]^{-1}[\Sigma_{\mathbf{F}}(S)]^{\frac{1}{2}}, \quad (24)$$

where $\Sigma(S)$ is the nonlinear data covariance matrix. $\Sigma(S)$ can be decomposed into two covariance matrices, $\Sigma_{\mathbf{F}}(S)$ and $\Sigma_d(S)$, where the former is the deterministic covariance of the data governed by the forward function $\mathbf{F}_S$, and the latter is the data-noise covariance and describes the stochastic component of the data (Coles & Curtis 2011).

Eq. (22) is an approximation of the entropy between two data sets. To obtain an approximation for all parameter values in M, we integrate over δ. The expectation of the entropy can then be found by (Coles & Curtis 2011)

$$E_\delta[\delta^T \Sigma\delta] = \int \delta^T \Sigma p(\delta)\mathrm{d}\delta = \mathrm{Tr}\,\Sigma. \quad (25)$$

Thus, the expected entropy of a design is the trace of the nonlinear data covariance matrix. However, if $\delta(S) = 0$, the integral is maximized for any design S, which occurs when $\mathbf{m}_i = \mathbf{m}_j$ or $f(\mathbf{F}_S)^{-1}$ is non-unique. The former is trivial as the same parameters produce identical data sets. The latter case was mitigated by Coles & Curtis (2011) by modifying eq. (25) to the following nonlinear criterion which is maximized for D_N-optimization:

$$\Phi_{D_N}(S) = \mathrm{Tr}\ln\Sigma(S) = \ln\det\Sigma(S). \quad (26)$$

4 NUMERICAL EXPERIMENTS

Now, we describe the numerical experiments to find what quality measure is best to design source location problem surveys. We first present the synthetic modelling domain and methods used to obtain arrival times. Then, we evaluate the number of prior samples needed for stable designs, the compute time needed for each method and the various quality measures applied in a sequential design algorithm. Finally, we assess the performance of the networks designed using the different quality measures.

4.1 Synthetic models

Cross-sections through the synthetic velocity models within which we conduct tests are shown in Fig. 3. The velocity structure is shown by the colour map, the region in which earthquakes occur is defined by the red region and a cross-section through the grid of possible receiver locations is shown by red triangles. Velocities vary only with respect to the z-axis, thus we only consider horizontally layered subsurfaces. Four different models are used: one homogeneous, a two-layered and two three-layered subsurface structures, with velocities between 2 and 4 km s^{-1} that increase with depth. Properties for the four subsurface models are shown in Table 1. The number of samples needed to stabilize each design method and compute time comparisons for the three SED algorithms are evaluated in the homogeneous model, whereas the design performance comparison between the algorithms is made for all four subsurface velocity models.

The location parameter prior is defined using a region in the subsurface where earthquakes are likely to occur. For this test, we define the region as a rectangular cuboid with horizontal dimensions of 20 km in x and y directions, and 10 km in depth z, centred at $x = y = 0$ km, $z = 17.5$ km for the homogeneous, two- and three-layered models a–c. Model d has the sources not in the half-space but in the second layer; for that model the cuboid dimensions are the same but it is centred at $x = y = 0$ km, $z = 10$ km. We assume the probability of an event at a particular location to be uniform within the cuboid and zero outside. Sample events are drawn on a grid with equidistant spacing chosen such that there is always a sample at the centre and at the borders of the cuboid to make sure the full extent of the cuboid is considered in every comparison.

Receivers may be placed on the surface $z = 0$ km at any of the set of pre-defined locations on a regular square grid shown in Fig. 3(e). The spacing between possible locations is 2 km in both x and y directions and the maximum offset from the centre of the source cuboid is 60 km along the x and y axes. These maximum dimensions were fixed by testing when all three SED algorithms stop placing receivers at the grid boundaries (except for the receivers placed at infinite offset which occurs due to the lack of energy attenuation in our synthetic test).

4.2 Arrival time calculation

Arrival times at the surface are calculated using a finite-difference scheme on a subsurface 2-D vertical cross-section (Podvin & Lecomte 1991) by solving the Eikonal equation

$$(\nabla t)^2 = s^2, \quad (27)$$

where t are the traveltimes and s is the slowness of each location in space. We add Gaussian noise to t with a standard deviation of 0.1 s and a mean of 0 s. The 1-D nature of the subsurface velocity

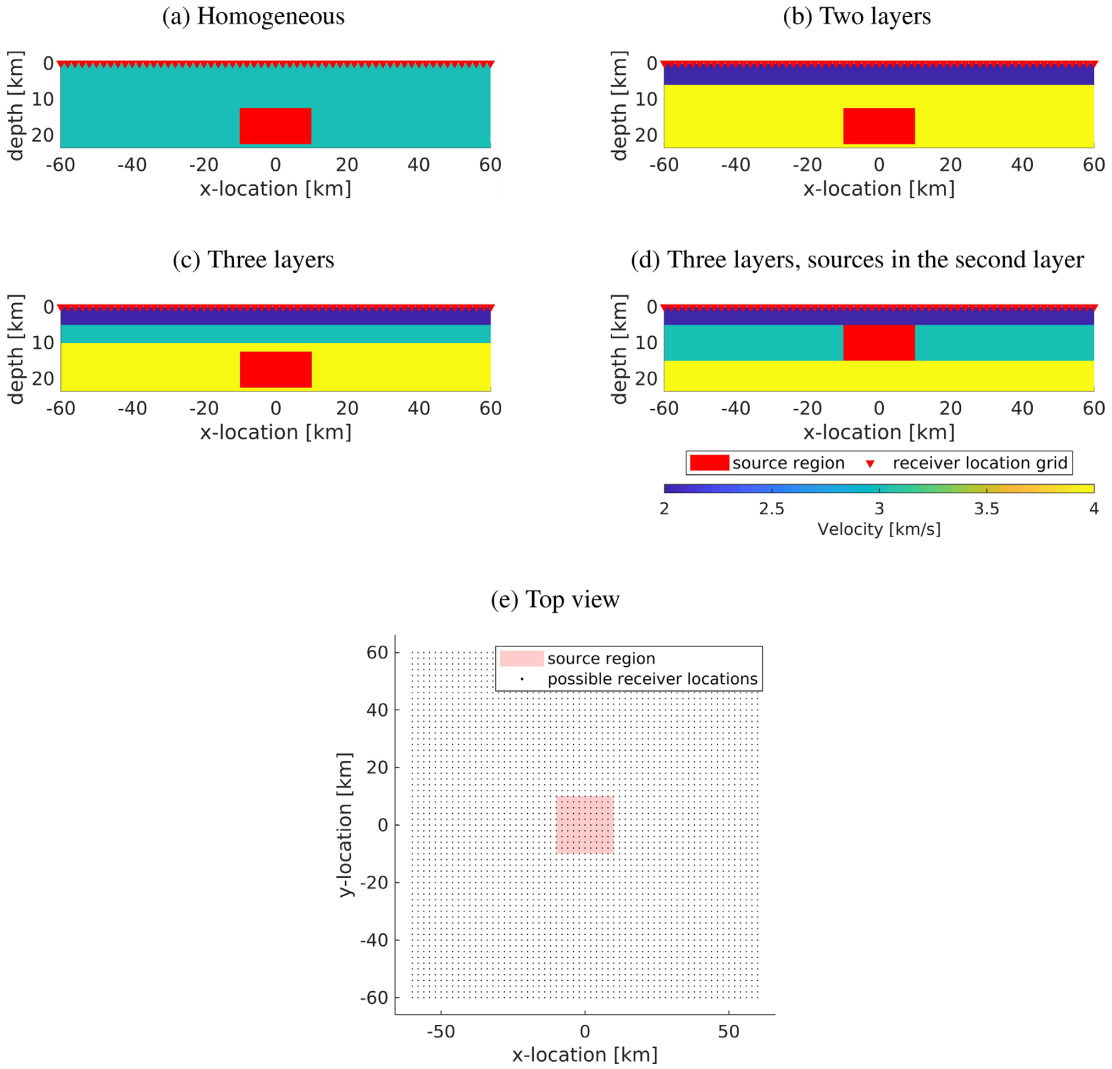

Figure 3. Cross-section at $y = 0$ of the four synthetic velocity models (a–d), where the red box shows the cuboid within which sources occur, red triangles show the possible receiver location grid and the colour map indicates the subsurface velocity model. Panel (e) shows a top view of the model where the grid of points shows all possible receiver locations and the red box is the top projection of the source cuboid.

Table 1. Four different subsurface velocity models, one homogeneous and three horizontally layered. Velocities and corresponding layer thicknesses are shown. Deepest layers are half-spaces indicated by infinite thickness.

Model	Layer velocities	Layer thicknesses
Homogeneous	3 km s^{-1}	∞ km
Two layers	2, 4 km s^{-1}	6, ∞ km
Three layers	2, 3, 4 km s^{-1}	5, 10, ∞ km
Three layers, sources in second layer	2, 3, 4 km s^{-1}	5, 15, ∞ km

model then allows us to calculate the arrival times to any receiver by rotating the 2-D cross-section between the event and receiver around the epicentre location. Furthermore, source samples on the same z-plane produce the same arrival times, but shifted in space by their (x, y) coordinates. Thus, for sources on the same depth level only one arrival time calculation is made across the 2-D grid to obtain arrival times at all receiver locations for all such sources.

The arrival times are thus effectively calculated on a radial 2-D cross-section around the source whereas the receiver grid is rectangular. To assign each receiver grid point an arrival time we take the arrival time closest to each receiver grid point, and to minimize the error the arrival times are computed with a much smaller spacing compared to the receiver grid. With an arrival time calculation spacing of 10 m and a receiver grid spacing of 2 km, we accrue a maximum error of 0.05 per cent in the modelled arrival time.

4.3 Number of samples

We sample the source location cuboid defined by $\rho(\mathbf{m})$ as described above, and forward model those samples to estimate $\rho(\mathbf{d})$ which is in turn used to assess the experimental designs. The number of

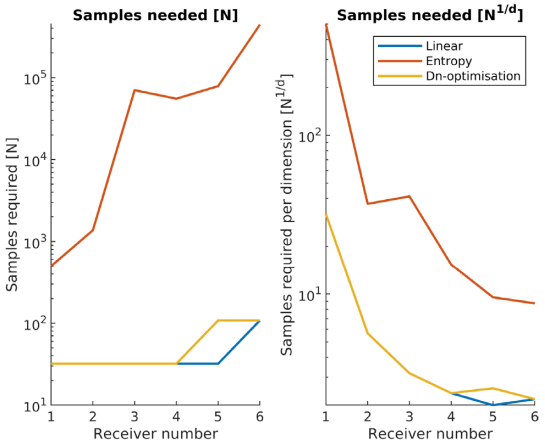

Figure 4. The number of samples N required to find a stable network. The left- and right-hand panels show the same data, but on the right the number of samples is shown per data space dimension where the dimensionality is d; that is, $\sqrt[d]{N}$ is plotted. The number of samples required is calculated for a uniform velocity model in all cases.

source samples defines how accurately we sample the parameter prior distribution and hence $\rho(\mathbf{d})$, and we wish to find the number of samples needed to create a stable design (a design that does not change when the prior is sampled more densely). We evaluate the number of prior samples needed for each receiver added sequentially to the network.

The number of samples needed to obtain a stable design is found by visual inspection. We compute designs for an increasing number of prior samples and observe whether the design changes when more samples are taken. Designs are considered stable for a given number of prior samples if taking more prior samples does not change the design. During the visual inspection, we take into account the rotational symmetries caused by our symmetric velocity structure and prior pdf. That is, we count designs as identical if they are rotations of each other by $\pm 90^\circ$ or 180°, or if they are symmetric reflections across any of a number of symmetry planes.

In Fig. 4, the number of samples needed to reach stability for each sequential receiver location is shown for a homogeneous subsurface velocity. The left-hand panel shows the number of samples N required to place each receiver stably, while the right-hand panel shows the average number of samples per data space dimension $\sqrt[d]{N}$ (where the dimensionality d is equal to the number of receivers) for increasing network size. The figure gives insight into how the sample requirements change with the desired network size.

D_N-optimization needs about 32 samples for each iteration until the fourth receiver, whereafter 108 samples are needed to place receivers stably. Maximum entropy design needs exponentially more samples for every receiver added, with 500 samples for the first receiver and around 440 000 for the sixth receiver. However, for the latter method the number of samples per dimension decreases to a stable number of around 10 as more receivers are added.

DETMAX is used to locate the first three receivers in linear design and provided good results from 32 prior samples. DETMAX is initialized with a random design so we pick the best design from 4 runs of DETMAX for the first three receivers. Sequential linear design needs 32 samples for the fourth and fifth receiver and 108 for the sixth receiver. For computational simplicity, we now continue to use the number of samples required for a six receiver network for the full sequential process and for the other velocity models as shown in Table 2.

4.4 Optimal network designs

In Figs 5–7, the sequential receiver placement for each method is shown for the homogeneous subsurface. The colour map represents the quality measure for every receiver location on the grid: higher (yellow) values are more favourable locations to add a receiver to the network according to each SED algorithm's quality measure. The cross represents the highest quality measure value and hence the location where a receiver is placed and added to the network; the pink dots are the receiver locations placed in previous iterations.

Linear design quality measures are shown for sequential design from the fourth receiver onwards in Fig. 5. The receivers placed by DETMAX are the dots in the iteration 4 panel, and are placed close to three of the corners of the source cuboid. The fourth receiver is placed at a distance, presumably to better constrain x and y source parameters since trade-offs with source depth will be reduced at far offsets from the sources. Receiver five is placed close to the last corner of the source cuboid. Receiver six is placed at a position which we cannot explain intuitively.

The maximum entropy design quality measure (Fig. 6) for the first receiver location has high values on the diagonals which are highest at maximum offset. The second receiver is placed on the diagonal opposing the first, and the third receiver is placed above the centre of the source cuboid. This setup is intuitive as the middle receiver location is most sensitive to changes in the z parameter while the other two are more sensitive to x and y values of the source location. Iterations 4 and 5 might seem unintuitive as the receivers are placed almost at the same locations as iterations 2 and 3, but the near repeating of locations does give twice the constraint on the horizontal source location (for Gaussian data distributions this would correspond to increasing the signal to noise ratio by $\sqrt{2}$). The final receiver is not a repeat, but is still located at far offset from the sources.

D_N-optimization (Fig. 7) produces essentially the same first three receiver locations as maximum entropy design. Receivers 4–6 are thus placed in a smaller triangle pointing in the opposite direction compared to the triangle formed by receivers 1–3. The quality measure maps for D_N-optimization look like more dramatic versions of those from maximum entropy design for the first three iterations. The quality measure decreases after the third iteration, presumably due to the source localization problem becoming overdetermined and the extra receivers only adding marginal information about the source parameters. Interestingly, the designs are similar to the design found to be optimal by Lilwall & Francis (1978) for a single source location.

4.5 Design results

We now show the network designs then compare network performance for each quality measure and each velocity structure.

4.5.1 Network designs

The receiver networks found using the three design methods, plus a completely random network, and all four subsurface velocity models, are shown in Fig. 8. We rotated the designs by 90°, 180° or 270° to make them visually more similar. The random network design is the same for each velocity model. The designs for the homogeneous velocity model (Fig. 8a) are the same as shown in Section 4.4.

Linear design (D-optimization) shows designs for the two- and three-layered subsurfaces that have two or three receivers are placed

Table 2. The number of samples used for each method to calculate the optimal design of a six-receiver network for the four velocity models. We used slightly more samples than were required for us to be confident that a stable solution had been reached. The compute times on a small compute server (Dell PowerEdge R820, 64 cores, 256 GB RAM, using 12 MATLAB threads) for a six-receiver network for each SED method using the number of samples from the table. Note that the number of samples required for maximum entropy design is larger than the axis in Fig. 10.

Method	Number of samples	Compute time required
Linear	108	216 s
Maximum entropy design	442 368	3517 s
D_N-optimization	108	1.26 s

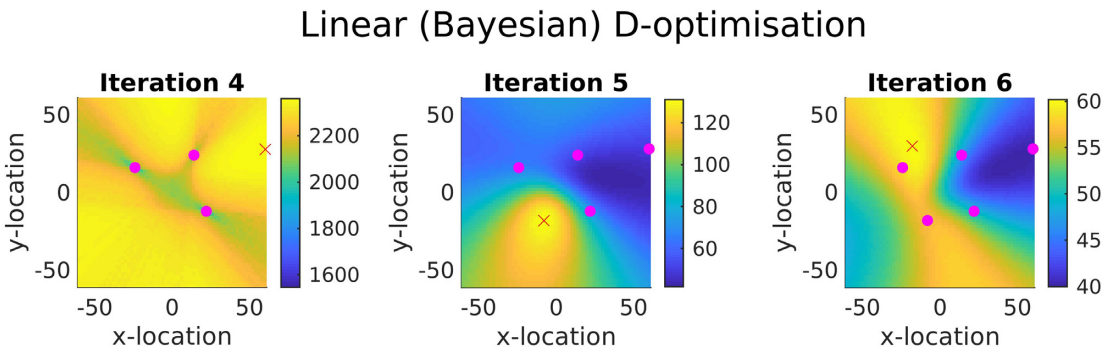

Figure 5. Receiver placement designed using linearized (Bayesian) *D*-optimality. The colour map shows the quality measure in eq. (13) calculated at each position on the grid of possible receiver locations; the maximum value defines where the next receiver will be placed (the cross). Dots show receiver locations already added to the network in previous iterations.

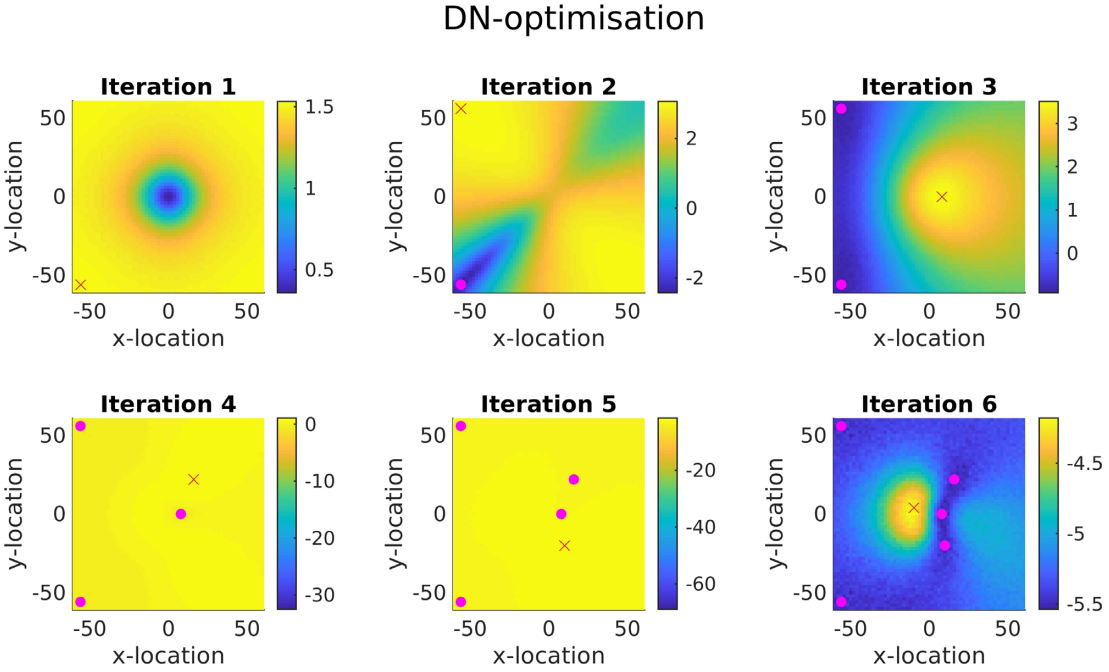

Figure 7. Receiver placement designed using D_N-optimization. The colour map shows the quality measure in eq. (26) calculated at each position on the grid of possible receiver locations; the maximum value defines where the next receiver will be placed (the cross). Dots show receiver locations already added to the network in previous iterations.

at far offset compared to just one for the homogeneous case. Interestingly, the design for the three-layered subsurface with the sources in the second layer (Fig. 8d) is very similar to the network for the homogeneous case.

Unlike the other two measures, maximum entropy design places more receivers at far offsets than over the source region. As the velocity structure becomes more complicated, receivers are moved towards the centre. This may be because refractions causing informative arrival times can be observed closer to the centre—particularly for model (d) in which the sources are closer to the surface than in the other models. The same behaviour is seen in linear design: model (a) has the receivers positioned in a larger circle compared to model (d). In all cases, maximum entropy design places receivers in almost repeated locations to improve those constraints.

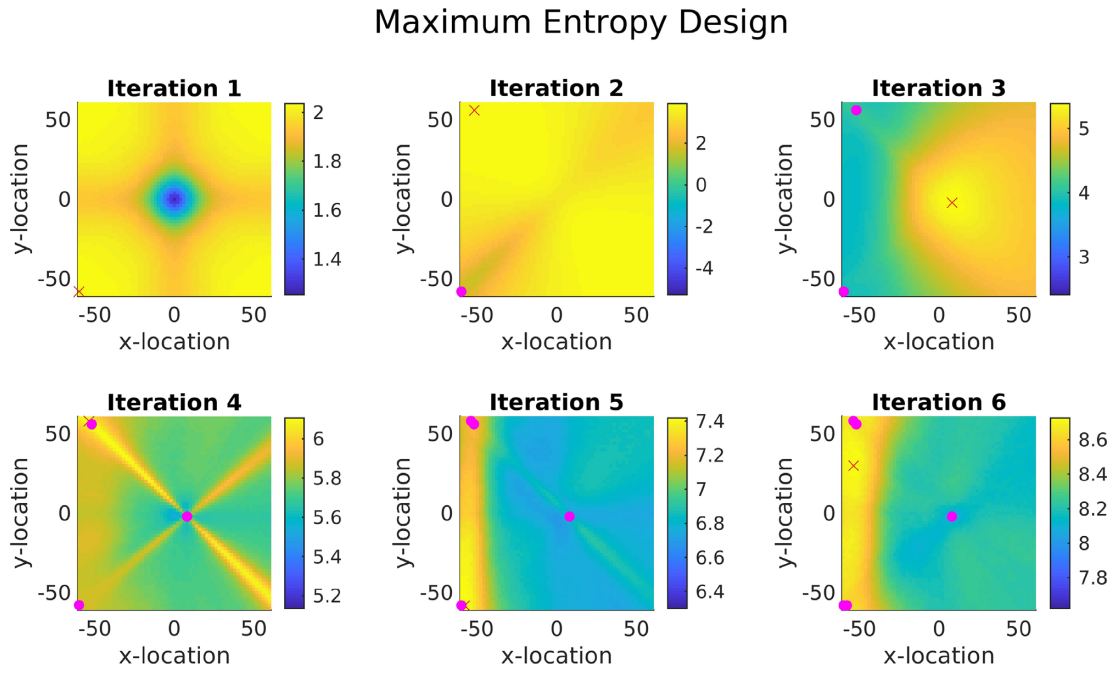

Figure 6. Receiver placement optimized using maximum entropy design. The colour map shows the quality measure Ent($\rho(\mathbf{d}|S)$) calculated at each position on the grid of possible receiver locations; the maximum value defines where the next receiver will be placed (the cross). Dots show receiver locations already added to the network in previous iterations.

D_N-optimization finds the same design for the homogeneous and two-layer subsurfaces. This design is the same as the optimal design for a single point source (Lilwall & Francis 1978) which has one receiver directly above the source region and the other receivers placed on concentric circles around it. The more complex the velocity structure the more receivers are placed towards the centre, but still with one receiver over the centre of the source region.

4.5.2 Network performance

To find which quality measure finds the best networks, we quantify the performance of the resulting network designs. We could evaluate the posterior model uncertainty for each model under every design (Fig. 2, bottom) to quantify receiver network performance. However, that would require that we then summarize the quality of those posterior distributions by one or more metrics or statistics for comparisons. In fact, by looking at the uncertainty in data space we can also evaluate metrics of network performance, without inverting data for the posterior pdf. We do this by evaluating the volume around each source sample defined by an arbitrary uncertainty threshold around the true data values for that source in data space. We choose the threshold to be 0.5 s. For this assessment, sources are sampled from the same prior distribution as used for designing the networks, but with far higher density leading to a much larger number of samples (864 000). These samples are discrete, so we can estimate the size of the volume by counting the number of source samples inside the volume. As an example consider Fig. 2: if we were to evaluate the volume of the posterior in Fig 2 under design *a* for source 5 and we count the number of other sources in its uncertainty region in data space defined by the dark grey region. So, for design *a* we count two other sources occupying the volume, whereas in design *b* no other sources occupy the dark grey region: thus design *b* has smaller posterior uncertainty for event 5 than design *a*. We repeat this by evaluating the posterior of all other source samples and count the average number of other sources within the posterior uncertainty volume. Thus we calculate the percentage of distinguishable source pairs, averaged over all source samples.

We note that some inversion algorithms would not be able to discriminate source pairs that give significantly different arrival times. For example, by definition linearized methods assume incorrect physical relationships between traveltimes and parameters in nonlinear problems, and hence implicitly assume incorrect arrival times across most of parameter space. Therefore, our performance number can be seen as an optimistic bound: it represents the performance that could be achieved by a network if locations were found using an ideal (non-linearized) inversion method.

We compute optimal designs for each SED method and each of the four subsurface velocity models, and evaluate their performance by looking at the percentage of distinguishable event pairs with results shown in Fig. 9. We present the number of distinguishable source pairs as a percentage of the total number of pairs evaluated; the best network has the highest such number. To understand the relative size of this number, we compare each set of results with the random network.

In all velocity structures, we see a similar pattern: all quality measures are able to find a network that performs better than a completely random network, network performance increases with the rate of increase that reduces with network size, and D_N-optimization finds designs that are best at distinguishing sources. Also, it seems that the rate of increase in performance converges to roughly the same rate for all quality measures (the increase from a five receiver network to a six receiver network is approximately equal for all measures). We therefore expect to see a similar comparison between SED method performance for larger receiver networks.

Linear design starts from a three receiver network as that is created by DETMAX instead of sequential design. After that, a large increase in performance is seen which flattens for a six receiver network. Maximum entropy design follows the same performance

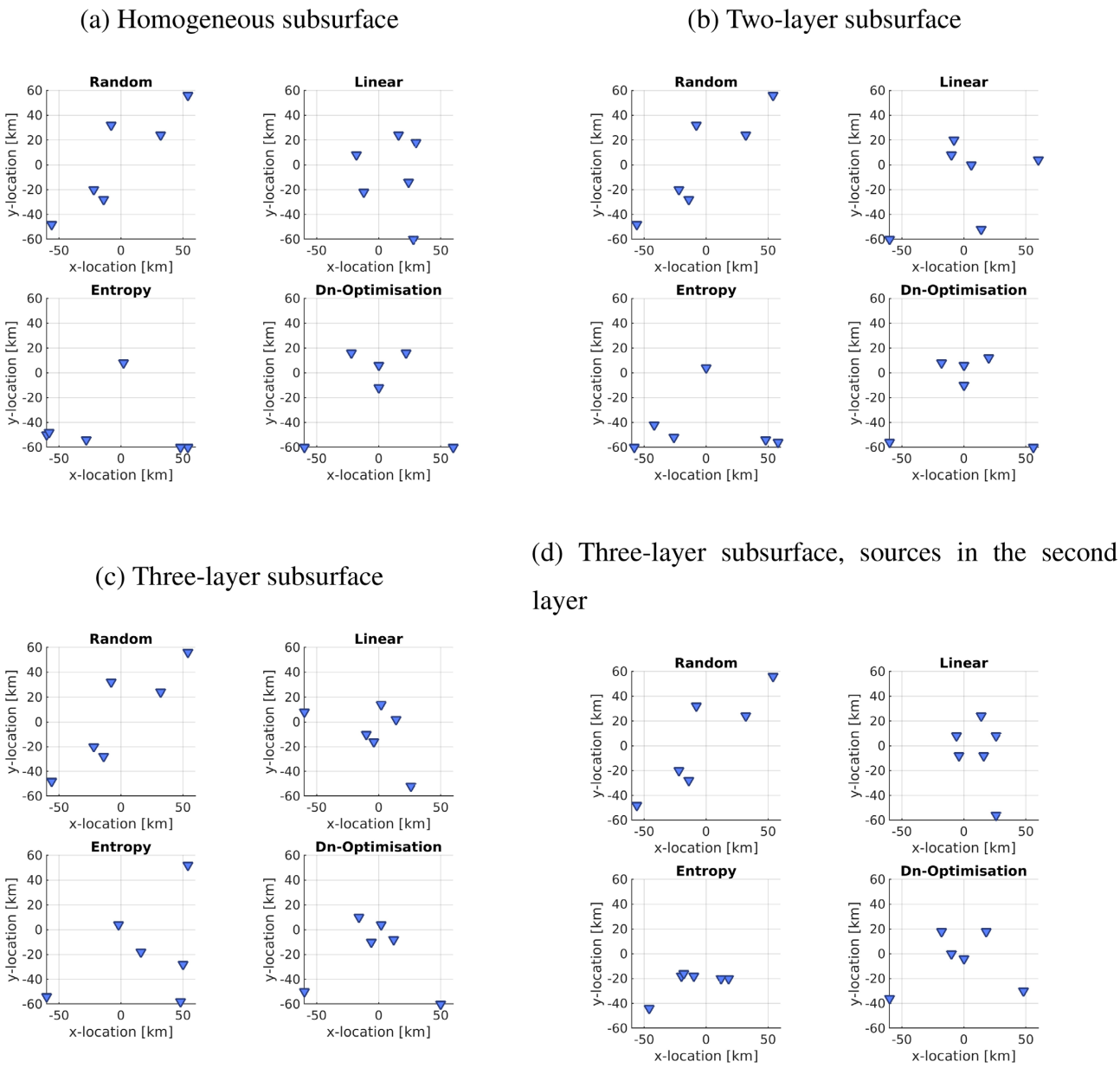

Figure 8. Network designs for each method and each subsurface model, rotated by 90°, 180° or 270° to make them visually more similar. The four subsurface models are: (a) homogeneous, (b) two-layered, (c) three-layered subsurface and (d) a three-layered subsurface with the sources in the second layer as shown in Fig. 3. Triangles represent receiver locations.

trend as D_N-optimization for the first three receivers, but performs worse for larger network sizes. Maximum entropy design finds the worst networks, except for the three-layer subsurface for which it performs better than linear design. We acknowledge that this relatively poor performance may be due to inaccurate calculation of the entropy related to a limited number of samples.

4.6 Compute time

For a six receiver network computed using different numbers of prior samples, we plot the compute times in Fig. 10. Receiver networks are computed for a range of prior sample numbers from 4 to 2916. For this range, the computation time ranges from 8 s to 2.8 h for linear design, from 4 to 14 s for maximum entropy design and from 1.2 to 3 s for D_N-optimization. Note that the compute time for the linear method depends on how many iterations DETMAX needs to reach a stable result, which is related to the random network from which it starts. Therefore, linear design computation time results are averaged over four runs. Both nonlinear methods follow similar trends but with maximum entropy design having a steeper gradient. As discussed earlier the number of samples needed is different for each measure, therefore compute times for the six receiver network are shown in Table 2. D_N-optimization is fastest to compute, followed by linear design, then maximum entropy design.

Fig. 11 shows the time needed to place a single receiver in the fourth iteration. The right-hand plot shows the increase in compute time required for each receiver added to the network. The compute times for each receiver are normalized by the compute time needed for the fourth receiver so that we compare all quality measures using sequential design. For each receiver added to the network

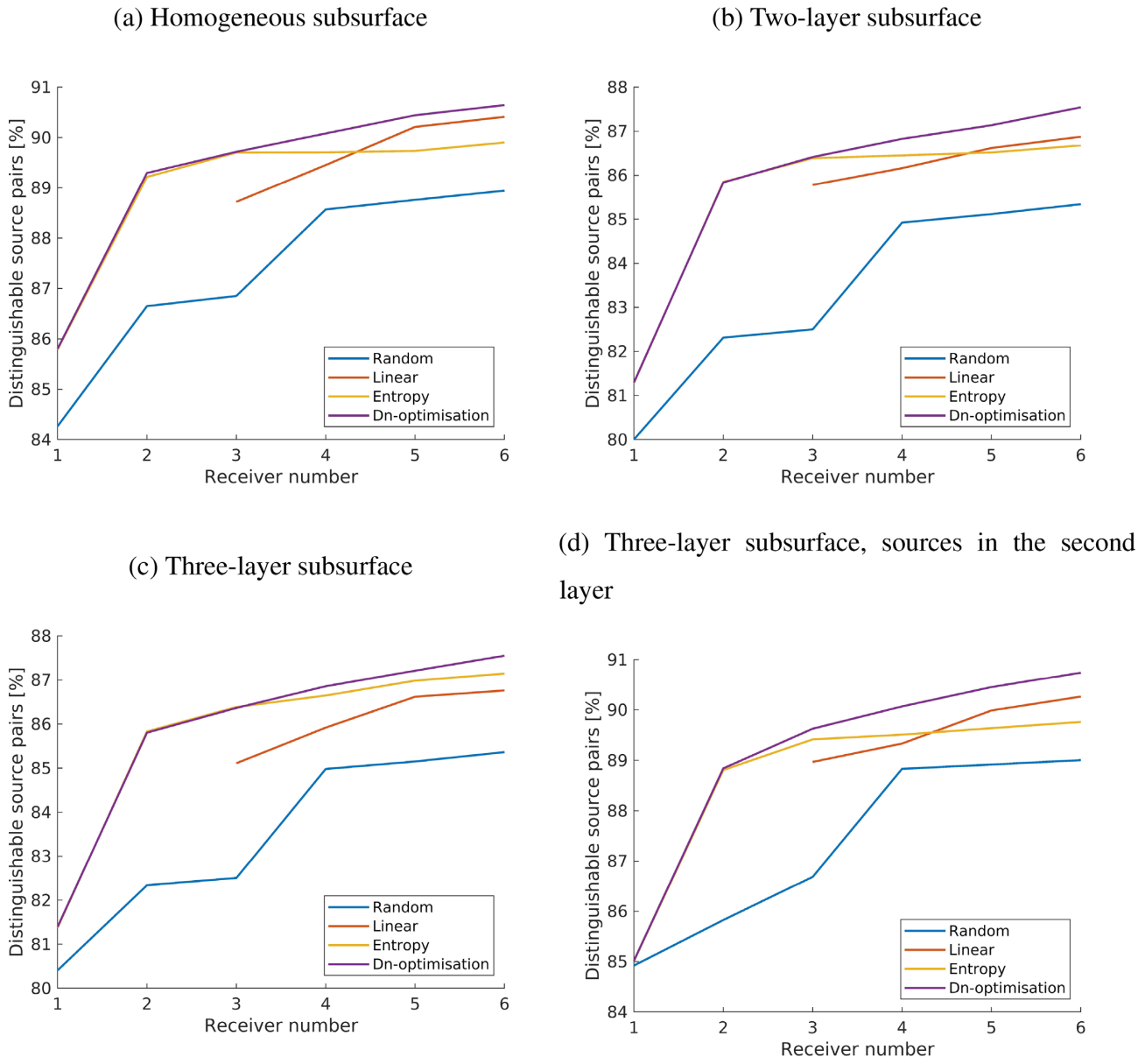

Figure 9. Performance of receiver networks generated by each of the three sequential SED methods, plus a random network, compared for (a) a homogeneous subsurface velocity model, (b) a two-layer model, (c) a three-layer model with the sources in the half-space and (d) a three-layer model with the sources in the second layer (see Fig. 3). Performance is expressed as the percentage of distinguishable source pairs from a total of 7.4×10^{11} pairs within the source cuboid.

with linear design the computation time is roughly constant, while maximum entropy design and D_N-optimization take more time for each receiver added.

5 DISCUSSION

The designs are symmetric in the sense that they can be rotated by 90°, 180° or 270° around the origin and can be mirrored through all symmetry axes of the square region that we considered, and under each reflection or rotation the quality measures will be identical. This is due to the symmetry of the square edges of the model, the square edges of the centrally located earthquake source region and due to our choice of one dimensional velocity models. We exploit this symmetry when finding the number of samples required to obtain a stable design: we rotated and mirrored each network to find whether each final network was similar to a network created with fewer samples.

However, the designs themselves are not symmetric, and to our knowledge neither should they be. For example, using the linearized design measure, the first three receivers were optimized simultaneously and the three optimal locations are shown by pink spots in the left-hand panel of Fig. 5. While there is clearly some apparent regularity in two of those locations relative to the centre of the grid, this design actually breaks all symmetries in the problem. Any of the rotations and reflections listed above results in a different design. A three-receiver design could have preserved at least two symmetries of the square grid (e.g. by rotating the array slightly to align it with the diagonals of the square), but the optimal design does not do so. This illustrates a key point about optimal designs: they may break symmetries in order to better constrain parameter combinations that

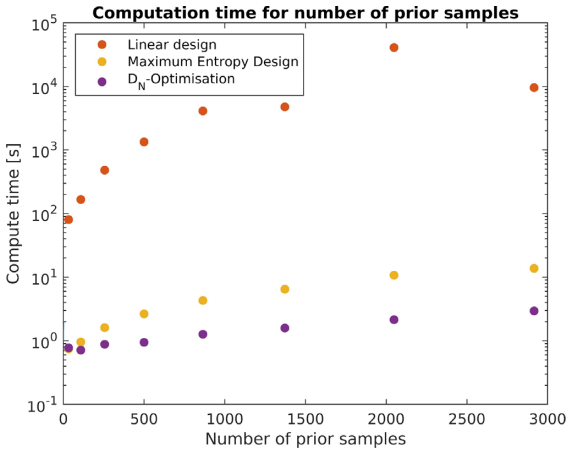

Figure 10. Compute time for a six-receiver network for a changing number of prior model samples. Linear design is shown by the red dots, maximum entropy design and D_N-optimization are shown with yellow and purple dots, respectively.

would be less constrained, or indeed be hidden by (lie in the null space of) symmetric designs.

The exact designs shown in Fig. 8 are on the whole not easily intuitively justified or explained. This is to be expected: the forward physics relating data to parameters is nonlinear, and the design metrics are all nonlinearly related to the designs as shown in Figs 5–7. This reveals the need for formalized design procedures such as those explored here; such methods can construct seemingly intuitive designs (e.g. D_N-optimization design in Figs 8a and b) but can operate just as effectively when intuition cannot help.

Fig. 9 shows that all three SED methods produce better than random networks. This proves that the methods are working correctly and produce credible results. Looking at Fig. 8, we see that the designs change when the velocity structure changes. This confirms that the methods are able to adapt to different scenarios.

It could be assumed that D_N-optimization performs similarly to maximum entropy design because the former is an approximation of the latter. However, this appears not to be the case as maximum entropy design performs worse than D_N-optimization. Three possibilities exist: (1) the sequential design method may not find the global maximum, (2) entropy is estimated inaccurately or (3) the performance metric used is not a good approximation to the Shannon information. The maximum entropy design algorithm could have got stuck in a local maximum. This is supported by the data (Fig. 9) because Maximum Entropy Design has approximately the same performance as D_N-optimization until the third receiver and only thereafter performs worse, which could be the point at which Maximum Entropy Design diverges towards a local maximum. Entropy is difficult to evaluate: it requires many samples and a well-chosen partition size (van den Berg *et al.* 2003, 2005). Entropy estimation by k-d partitioning solves the latter problem by choosing a partition size based on the samples themselves. Our performance metric is the same for all algorithms, and although it does not quantify the Shannon information directly it is a proxy for inversion performance which, in a geophysical setting, is more important. Future research could evaluate whether adding even more parameter prior samples would find the global maximum.

Fig. 10 shows the compute times needed for each method when we use different numbers of prior samples. Linear design is inefficient for large numbers of samples. This is because calculating gradients is computationally expensive, and in addition we must use DETMAX to design the first three receiver locations which requires more designs to be evaluated than are required for sequential design. The random initialization of DETMAX also makes computation times somewhat inconsistent as seen by the notable increase in time around 2000 prior samples (results given are averaged over four runs). Maximum entropy design and D_N-optimization are more efficient for larger numbers of samples, requiring lower compute times. All methods converge to a roughly log-linear increase in compute time with the number of samples. However, the large number of samples required by maximum entropy design (Table 2: 442 368 samples) causes that method to cost substantially more in compute time to design a six receiver network compared to the other two methods.

The marginal cost of adding a receiver to a given network using a fixed number of prior samples is shown in Fig. 11. Maximum entropy design and D_N-optimization require more compute time for each subsequent receiver, whereas linear design always takes roughly the same time for each added receiver. Given the similar prior sample requirements for linear design and D_N-optimization (Fig. 4), linear design could be faster to compute for large receiver networks.

Our performance metric quantifies the resolvability of source pairs, or in other words, the average volume of a contour in (x, y, z) around each event within which other events cannot be discriminated from the first given data uncertainty threshold of 0.5 s. However, the acquired data may also be used for other purposes, for example for seismic tomography or to estimate event magnitudes. De Landro *et al.* (2019) quantify the performance of seismic networks against multiple objectives which could be a useful extension of our work.

Nonlinear data–parameter relationships could cause multiple modes in the likelihood or posterior pdf. Linear design methods may perform less well in such cases because the locally determined gradient is assumed to be the same across the full parameter range. To improve network performance from linear design, Curtis & Spencer (1999) and Curtis (2004b) proposed to find the number of modes in the misfit function as a (negative) quality measure in addition to standard linear design measures. This might make the design scheme nonlinear while retaining low sample number requirements and constant compute times per receiver added, but more research is required to assess that potential.

6 CONCLUSION

We show that D_N-optimization designs best performing receiver networks with lowest compute times for a range of source location problems out of the three methods tested. Linear design using D-optimization have the advantage of constant compute times per added receiver, but produce worse designs and need substantially more total compute time than D_N-optimization. Maximum entropy design is inefficient and in our tests is effectively impractical as it requires many samples and very long compute times to obtain a stable design (which we did not achieve in this study) comparable to D_N-optimization. We conclude that D_N-optimization consistently designs the best source location experiments and does so with far less computational resources than linear or maximum entropy design. There therefore appears to be no reason not to adopt D_N-optimization in future survey and experimental design projects.

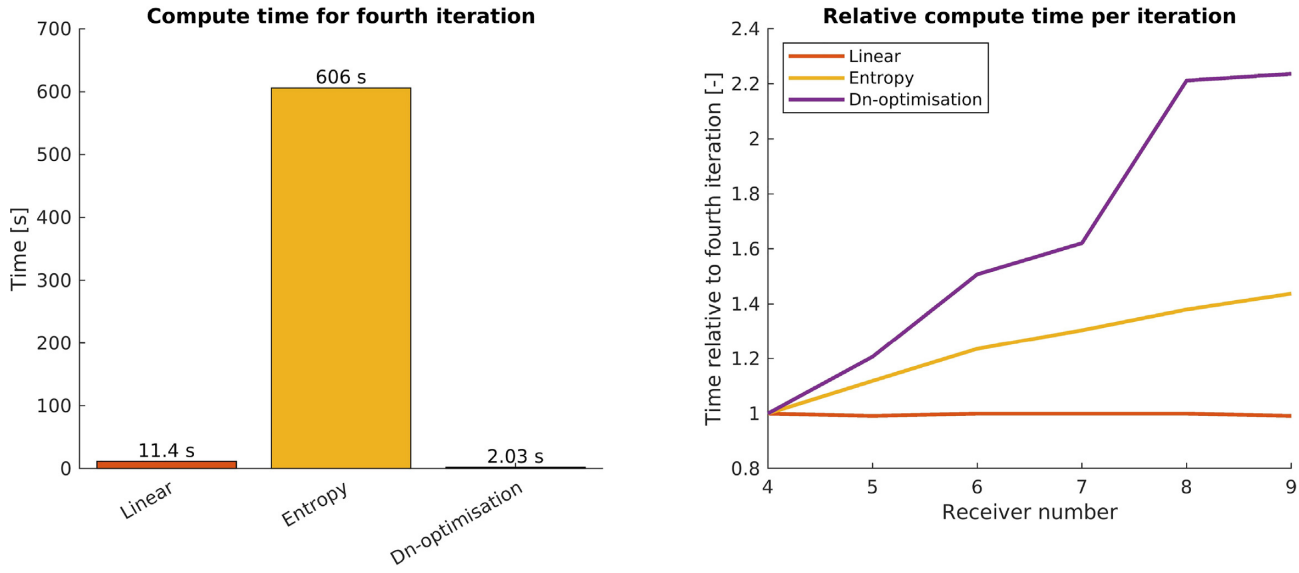

Figure 11. Left: the compute time needed for placing the fourth receiver. Right: the compute time needed per receiver added, relative to (divided by) the compute time required to add the fourth receiver to the network. A value of 2.5 would mean a compute time of 2.5 times that of the fourth iteration. Note that the fourth iteration compute times differ per method, and hence only the trends are compared here (absolute compute times for six receiver networks are shown in Fig. 10).

ACKNOWLEDGEMENTS

The authors would like to thank Malcolm Sambridge and an anonymous reviewer for constructive comments on the manuscript. The authors also thank the Edinburgh Interferometry Project sponsors (Schlumberger, BP and Total) for supporting this research.

REFERENCES

Atkinson, A., Donev, A. & Tobias, R., 2007. *Optimum Experimental Designs, with SAS,* Vol. 34, Oxford Univ. Press.

Barth, N. & Wunsch, C., 1990. Oceanographic experiment design by simulated annealing, *J. Phys. Oceanogr.,* **20**(9), 1249–1263, doi.org/10.1175/1520-0485(1990)020<1249:OEDBSA>2.0.CO;2.

Bayes, T., 1763. LII. An essay towards solving a problem in the doctrine of chances. by the late Rev. Mr. Bayes, F.R.S. communicated by Mr. Price, in a letter to John Canton, A.M.F.R.S., *Phil. Trans. R. Soc.,* **53,** 370–418.

Box, G.E. & Lucas, H., 1959. Design of experiments in non-linear situations, *Biometrika,* **46**(1/2), 77–90.

Cerveny, V., 2005. *Seismic Ray Theory,* Cambridge Univ. Press.

Chaloner, K. & Verdinelli, I., 1995. Bayesian experimental design: a review, *Stat. Sci.,* **10,** 273–304.

Chu, J.T., 1955. On the distribution of the sample median, *Ann. Math. Stat.,* **26**(1), 112–116.

Coles, D. & Curtis, A., 2011. Efficient nonlinear Bayesian survey design using D_N-optimization, *Geophysics,* **76**(2), Q1–Q8.

Coles, D.A. & Morgan, F.D., 2009. A method of fast, sequential experimental design for linearized geophysical inverse problems, *Geophys. J. Int.,* **178**(1), 145–158.

Cox, D.R., 1958. *Planning of Experiments,* Wiley.

Curtis, A., 1999a. Optimal design of focused experiments and surveys, *Geophys. J. Int.,* **139**(1), 205–215.

Curtis, A., 1999b. Optimal experiment design: cross-borehole tomographic examples, *Geophys. J. Int.,* **136**(3), 637–650.

Curtis, A., 2004a. Theory of model-based geophysical survey and experimental design: part 1—linear problems, *Leading Edge,* **23**(10), 997–1004.

Curtis, A., 2004b. Theory of model-based geophysical survey and experimental design: part 2—nonlinear problems, *Leading Edge,* **23**(10), 1112–1117.

Curtis, A. & Spencer, C., 1999. Survey design strategies for linearized nonlinear inversion, in *SEG Technical Program Expanded Abstracts 1999,* doi:10.1190/1.1820882.

Curtis, A., Michelini, A., Leslie, D. & Lomax, A., 2004. A deterministic algorithm for experimental design applied to tomographic and microseismic monitoring surveys, *Geophys. J. Int.,* **157**(2), 595–606.

De Landro, G., Picozzi, M., Russo, G., Adinolfi, G.M. & Zollo, A., 2019. Seismic networks layout optimization for a high-resolution monitoring of induced micro-seismicity, *J. Seismol.,* doi:10.1007/s10950-019-09880-9.

Goldberger, J., Gordon, S. & Greenspan, H., 2003. An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures, in *Proceedings Ninth IEEE International Conference on Computer Vision,* IEEE, Nice, France, p. 487.

Guest, T. & Curtis, A., 2009. Iteratively constructive sequential design of experiments and surveys with nonlinear parameter-data relationships, *J. geophys. Res.,* **114**(B4), doi:10.1029/2008JB005948.

Guest, T. & Curtis, A., 2010a. Amplitude variation with offset (AVO) optimal trace selection for AVA processing of shale-sand reservoirs, *Geophysics,* **75**(4), C37.

Guest, T. & Curtis, A., 2010b. Optimal trace selection for AVA processing of shale-sand reservoirs, *Geophysics,* **75**(4), C37–C47.

Guest, T. & Curtis, A., 2011. On standard and optimal designs of industrial-scale 2-D seismic surveys, *Geophys. J. Int.,* **186**(2), 825–836.

Huang, Z. & Zhao, D., 2013. Relocating the 2011 Tohoku-oki earthquakes (M 6.0–9.0), *Tectonophysics,* **586,** 35–45.

Kackar, R.N., 1985. Off-line quality control, parameter design, and the Taguchi method, *J. Qual. Technol.,* **17**(4), 176–188.

Kijko, A., 1977. An algorithm for the optimum distribution of a regional seismic network—I, *Pure appl. Geophys.,* **115**(4), 999–1009.

Lilwall, R.C. & Francis, T.J.G., 1978. Hypocentral resolution of small ocean bottom seismic networks, *Geophys. J. Int.,* **54**(3), 721–728.

Lin, G., 2013. Three-dimensional seismic velocity structure and precise earthquake relocations in the Salton trough, southern California, *Bull. seism. Soc. Am.,* **103**(5), 2694–2708.

Lindley, D.V. *et al.*, 1956. On a measure of the information provided by an experiment, *Ann. Math. Stat.,* **27**(4), 986–1005.

Maurer, H., Greenhalgh, S. & Latzel, S., 2009. Frequency and spatial sampling strategies for crosshole seismic waveform spectral inversion experiments, *Geophysics,* **74**(6), WCC79–WCC89.

Maurer, H., Curtis, A. & Boerner, D.E., 2010. Recent advances in optimized geophysical survey design, *Geophysics,* **75**(5), 75A177–75A194.

Maurer, H., Nuber, A., Martiartu, N.K., Reiser, F., Boehm, C., Manukyan, E., Schmelzbach, C. & Fichtner, A., 2017. Optimized experimental design in the context of seismic full waveform inversion and seismic waveform imaging, *Adv. Geophys.,* **58,** 1–45.

Mitchell, T.J., 1974. An algorithm for the construction of D-optimal experimental designs, *Technometrics,* **16**(2), 203–210.

Podvin, P. & Lecomte, I., 1991. Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools, *Geophys. J. Int.,* **105**(1), 271–284.

Ren, Z. & Kalscheuer, T., 2019. Uncertainty and resolution analysis of 2D and 3D inversion models computed from geophysical electromagnetic data, *Surv. Geophys.,* **41,** 47–112.

Romdhane, A. & Eliasson, P., 2018. Optimised geophysical survey design for CO_2 monitoring—a synthetic study, in *14th Greenhouse Gas Control Technologies Conference,* Melbourne.

Shannon, C.E., 1948. A mathematical theory of communication, *Bell Syst. Tech. J.,* **27**(3), 379–423.

Shewry, M.C. & Wynn, H.P., 1987. Maximum entropy sampling, *J. Appl. Stat.,* **14**(2), 165–170.

Sirgue, L. & Pratt, R.G., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies, *Geophysics,* **69**(1), 231–248.

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.

Stowell, D. & Plumbley, M.D., 2009. Fast multidimensional entropy estimation by *k*-d partitioning, *IEEE Signal Process. Lett.,* **16**(6), 537–540.

Stummer, P., Maurer, H. & Green, A.G., 2004. Experimental design: electrical resistivity data sets that provide optimum subsurface information, *Geophysics,* **69**(1), 120–139.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* Vol. 89, SIAM.

Toledo, T., Jousset, P., Maurer, H. & Krawczyk, C., 2018. Optimized experimental network design for earthquake location problems: applications to geothermal and volcanic field seismic networks, *J. Volc. Geotherm. Res.,* doi:10.1016/j.jvolgeores.2018.08.011.

Tong, P., Yang, D., Liu, Q., Yang, X. & Harris, J., 2016. Acoustic wave-equation-based earthquake location, *Geophys. J. Int.,* **205**(1), 464–478.

van den Berg, J., Curtis, A. & Trampert, J., 2003. Optimal nonlinear Bayesian experimental design: an application to amplitude versus offset experiments, *Geophys. J. Int.,* **155**(2), 411–421.

van den Berg, J., Curtis, A. & Trampert, J., 2005. Corrigendum, *Geophys. J. Int.,* **161**(2), 265.

Waldhauser, F. & Ellsworth, W.L., 2000. A double-difference earthquake location algorithm: method and application to the northern Hayward fault, California, *Bull. seism. Soc. Am.,* **90**(6), 1353–1368.

Wen, L. & Long, H., 2010. High-precision location of North Korea's 2009 nuclear test, *Seismol. Res. Lett.,* **81**(1), 26–29.

Wilson, M.P., Davies, R.J., Foulger, G.R., Julian, B.R., Styles, P., Gluyas, J.G. & Almond, S., 2015. Anthropogenic earthquakes in the UK: a national baseline prior to shale exploitation, *Mar. Pet. Geol.,* **68,** 1–17.

Winterfors, E. & Curtis, A., 2008. Numerical detection and reduction of non-uniqueness in nonlinear inverse problems, *Inverse Probl.,* **24**(2), 025016.

Wuestefeld, A., Greve, S.M., Näsholm, S.P. & Oye, V., 2018. Benchmarking earthquake location algorithms: a synthetic comparison, *Geophysics,* **83**(4), KS35–KS47.

Wunsch, C., 1996. *The Ocean Circulation Inverse Problem,* Cambridge Univ. Press.

Zhang, M. & Wen, L., 2014. Seismological evidence for a low-yield nuclear test on 12 May 2010 in North Korea, *Seismol. Res. Lett.,* **86**(1), 138–145.