



Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta



Multi-objective optimization of spatial sampling



R.M. Lark

British Geological Survey, Keyworth, Nottinghamshire, NG12 5GG, UK

ARTICLE INFO

Article history:

Received 12 April 2016

Accepted 6 September 2016

Available online 22 September 2016

Keywords:

Spatial sampling

Multi-objective

Pareto set

ABSTRACT

The optimization of spatial sampling by simulated annealing has been demonstrated and applied for a range of objective functions. In practice more than one objective function may be important for sampling, and there may be complex trade-offs between them. In this paper it is shown how a multi-objective optimization algorithm can be applied to a spatial sampling problem. This generates a set of solutions which is non-dominated (no one solution does better than any other on all objective functions). These solutions represent different feasible trade-offs between the objective functions, and a subset might be practically acceptable. The algorithm is applied to a hypothetical example of sampling for a regional mean with the variance of the mean and the total distance travelled between sample points as the two objective functions. The solutions represent a transition of sample arrays from a loose grid to a tight loop. The potential to develop this approach and apply it to other spatial sampling problems is discussed.

© 2016 British Geological Survey a component part of NERC.

Published by Elsevier B.V.

This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. The problem and its motivation

The selection of locations for spatial sampling can be optimized numerically for cases, such as geostatistical surveys, where probability sampling is neither necessary nor desirable (De Gruijter et al., 2006). van Groenigen and Stein (1998) showed how the simulated annealing algorithm

E-mail address: mlark@nerc.ac.uk.

(Aarts and Korst, 1989) can be adapted to find spatial distributions of sample points which optimize some objective function. They called the method spatial simulated annealing (SSA). The approach has since been used widely. For example, Lark (2002) used SSA to investigate optimal sample designs for the variogram and Boer et al. (2002) used it to thin an existing sampling network optimally. Marchant and Lark (2007) and Brus and Heuvelink (2007) used SSA to optimize geostatistical surveys, and Wang et al. (2014) applied the method to optimize sampling for regional estimation of gross primary production. Marchant et al. (2013) used SSA in phased sampling designs for contaminated urban soils.

The SSA method finds a sampling design which minimizes a criterion, the objective function, such as the average kriging variance across the area of interest. However, for some sampling problems the performance of a particular distribution of sample sites may be judged by more than one criteria, all relevant to the data user, particularly when the data user is concerned not only with the quality of the resulting information but also the cost of acquiring it. This paper is concerned with how more than one objective function can be addressed in the optimization of spatial sampling. This requires some method for multiobjective optimization. First I briefly outline three exemplar problems which provide the motivation for this study.

1.1.1. Field soil sampling

The question of how best to sample an agricultural field to form an aggregate (bulk) sample from which to estimate available nutrient concentrations is of interest to the arable and horticultural sectors (Marchant et al., 2012). This sampling, certainly in the UK, is often undertaken by commercial contractors. The efficiency with which fields are sampled (staff time) is therefore important, as well as the statistical quality of the estimate, which may be measured by its standard error. Total distance walked is a major factor influencing the time taken to complete a survey (Beckett, 1981), and in a general investigation of soil sampling logistics Lark and Knights (2015) used figures from recent survey work by the Geological Survey of Ireland and the British Geological Survey which showed that daily staff costs are a unit of magnitude larger than unit analytical costs. In the case considered here, where sampling one field produces one bulk sample, there may therefore be considerable interest in reducing the time taken to sample each field. However, if the sampling time (equivalently the distance walked) is regarded as an objective function for a field sampling design, it will be in conflict with another objective function, the standard error of the sample estimate. The latter objective function is minimized by a sampling design which achieves good spatial coverage (De Gruijter et al., 2006), but the distance walked around such a sampling design will be large. A real practical question, for the land manager and the commercial organizations which provide a service, is how a sampling design can be in some sense optimized according to both criteria. The previously-cited study by Marchant et al. (2012), addressed only statistical criteria when optimizing sampling designs, so there would be considerable scope to extend this, given a suitable multiobjective optimization framework, to show how logistical efficiency can be considered as well. This particular example is used as an illustration of the approach to multiobjective optimization of sampling discussed in this paper.

1.1.2. Wireless sensor networks

There is increasing interest in the value of in situ sensor arrays to provide real-time information on soil properties (Shaw et al., 2016). More generally it is recognized that wireless sensor networks (WSN) have considerable potential for environmental monitoring (e.g. Othman and Shazali, 2012). A WSN is an array of autonomous sensors capable of passing data cooperatively to a central location. The particular value of WSNs is their capacity to deliver environmental information in real time from a distributed set of locations without a substantial infrastructure for data communication. As with any system for sampling an environmental variable in space, the value of the information provided by a WSN will depend on the spatial distribution of the sensors. This will be true whether the target variable is a spatial mean, the temporal trend of the spatial mean or the value of the target variable at a particular location in time and space (De Gruijter et al., 2006). A specific objective function can be determined for a particular quality measure of interest, which could be the standard error of a spatial mean or temporal trend parameter, or the mean prediction error variance of a spatio-temporal prediction. However, other considerations are important when considering the spatial

layout of a WSN. These include the routing distance and time around the network, the per-node energy consumption and the robustness of the network. Robustness may be increased if there are more alternative routes for data if one node fails, or scope to share processing tasks around the network if one subset of nodes becomes overloaded or, in the case of solar-powered sensors, if a subset is shaded. The potential to optimize network design with respect to such technical objective functions has been investigated (e.g. [Yang et al., 2013](#); [Yang and McCann, 2014](#)). It is likely that there may be conflicts between these technical objective functions and at least some possible statistical quality measures. For example, the scope for redundancy in data paths around a network would be close to minimal in a layout which maximizes spatial coverage. As WSN are increasingly used in research and application, the question of how to account for both technical performance and data quality in the design of the spatial layout can be expected to become a pressing one, requiring a conceptual framework and computational methodology. As the work reported by [Shaw et al. \(2016\)](#) is to be developed further in collaboration with the agricultural industry, this is a pressing question.

1.1.3. Unmanned aerial vehicles

In recent years the value of unmanned aerial vehicles (AEV) for the collection of spatial data has attracted increasing interest ([Anderson and Gaston, 2013](#)). Sensors can be mounted on an AEV and used to collect data on specific properties (e.g. [Faye et al., 2016](#)). As with the previous examples, the route that an AEV follows around a study site will affect the quality of the information that it returns (e.g. the spacing of adjacent passes of the sensor). A good flight path for an AEV bearing an environmental sensor must take account of the spatial variability of the measured variable and its implications for the standard error of estimated quantities or the error variance of spatial predictions. However, in practice it is also necessary to take account of logistical considerations (e.g. [Babel, 2013](#); [Cares and Dickmann, 2016](#)), which will include power demand given the path length and flying conditions. As in previous examples, the optimization of an overall sampling scheme is a multiobjective problem in practice because factors in addition to information quality measures cannot be ignored. This question has recently become practically important after research which has shown the potential of AEVs for monitoring soil erosion features ([McShane et al., 2016](#)). This research was part of a programme to pilot national-scale soil erosion monitoring in the UK. If the technology is to be deployed in that context then logistical and statistical considerations must be jointly considered when optimizing flight plans for monitoring sites to ensure that the national programme is practically feasible and produces adequate data.

1.2. Previous work on multiobjective sampling optimization, and Pareto-optimality

Some attention has been paid to the problem of optimizing a spatial sampling design with respect to more than one outcome. In some cases one can identify a single objective function which truly characterizes two aspects of a sample design which may be regarded as distinct or even conflicting. For example, many approaches to designing sampling for geostatistical survey aim to achieve appropriate spatial coverage. However, the sample must also support estimation of covariance parameters of the model, which is not well served by a sample design with no close pairs ([Stein, 1999](#)). In practice the two problems have been kept separate, with a spatial coverage design subsequently supplemented by some close pairs, constituting about 10% of the total sample according to the rule of thumb of [Stein \(1999\)](#). However, [Marchant and Lark \(2007\)](#) showed that the prediction error variances of the final map can be shown to depend on both the spatial coverage of the survey and the reliability of the covariance model. A single objective function was defined to quantify this. When SSA was used to optimize the sampling design according to this objective function the resulting sampling designs combined spatial coverage with some locally clustered observations.

In other cases it may be possible to combine two or more objective functions into a single one for optimization when both can be converted to an economic cost to the end-user. For example, [Ballari et al. \(2012\)](#) considered the value of information collected by mobile sensor networks. Networks were designed to minimize the expected costs from wrong predictions made from the network, but the costs of implementing and operating the network were also considered.

Another approach is to define an objective function which is a weighted combination of two or more distinct objective functions. For example, Melles et al. (2011) considered sampling designs for monitoring radioactive releases. One criterion on which a design can be assessed is the mean prediction error variance for measurements across the region. A second is the probability of detecting a particular event. Melles et al. (2011) formed a single objective function which was a weighted combination of these two. They suggested that an organization responsible for monitoring could select appropriate weights, and they chose weights for demonstration purposes. In practice it may be difficult to form weights that really reflect the requirements of a particular end-user of data, particularly when the weights are used to combine fundamentally different quantities such as a prediction error variance and a detection probability.

It may not always be possible, however, to reduce the properties of a proposed sample scheme to a single combined criterion which can then be optimized. In these circumstances the sponsors and end-users of the sampling must agree on trade-offs between criteria by a process of evaluation which is not necessarily reducible to weighting. In these circumstances numerical optimization can be useful by producing a set of alternative designs which are Pareto optimal. This is discussed in detail in Section 2.1, but in short the solutions in a Pareto-optimal set cannot be improved with respect to any of the objective functions under consideration without making at least one of the others worse. If one finds a Pareto-optimal set of solutions then one may identify a subset of acceptable solutions, those for which specified threshold values of the objective functions are passed, and then select a solution to implement on the basis of explicit discussion of the trade offs required.

In this paper I propose an approach to multi-objective optimization of spatial sampling that uses the AMOSA algorithm of Bandyopadhyay et al. (2008) as presented by Bandyopadhyay and Sahaha (2013). This algorithm is based on simulated annealing, and also incorporates an archive of solutions which is modified and updated as the algorithm runs. In the methods Section I discuss the problems of multi-objective optimization, the AMOSA approach, and how it is applied to the spatial sampling problem. I then consider an exemplar case study based on the problem considered in Section 1.1.1.

2. Materials and methods

2.1. Pareto optimal solutions

Fig. 1 shows a plot with feasible solutions (discs) to an optimization problem with two objective functions o_1 and o_2 . Both functions are to be minimized. Consider the black disc, it achieves a smaller value of both objective functions than the solution represented by the white disc, and so is said to *dominate* this solution. If the cross on the plot represented a feasible solution, then this would dominate the black disc. Neither grey disc dominates the black one because, while each grey disc has a smaller value of one of the objective functions than does the black, it also has a larger value of another objective function. We say that a solution is *Pareto-optimal* if there is no feasible solution which dominates it. The *Pareto frontier* for a problem is the set of its Pareto-optimal solutions. Note that, by definition, the solutions on the Pareto frontier must be mutually non-dominating.

Consider a case where we minimized some compound objective function

$$o_3 = \lambda_1 o_1 + \lambda_2 o_2,$$

where λ_1 and λ_2 are real non-negative weights normalized to sum to 1, so o_3 is a convex combination of o_1 and o_2 . If the solution found does indeed minimize o_3 then it must represent a solution on the Pareto frontier for o_1 and o_2 , because any solution which dominated it with respect to o_1 and o_2 would necessarily have a smaller value of their convex combination. This suggests that one might investigate the Pareto frontier by optimizing different convex combinations of the constituent objective functions. However, this approach has two drawbacks (Das and Dennis, 1997). First, while any solution to such an optimization is a point on the Pareto frontier, not every solution on the Pareto frontier corresponds to the optimization of a convex combination of the separate objective functions. This is because the Pareto frontier is not necessarily convex. Second, it is inefficient because evenly-spread solutions on the Pareto frontier, even when this is convex, do not in general correspond to evenly-spaced weights.

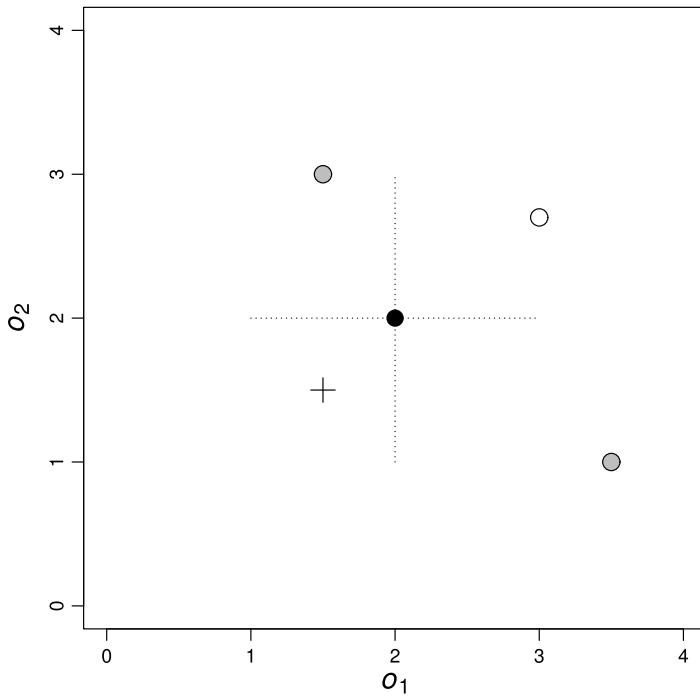


Fig. 1. Some feasible solutions (discs) to a problem minimizing two objective functions, o_1 and o_2 . The black disc and the two grey discs comprise a non-dominating set of solutions. The black disc dominates the white disc. If the cross represented a feasible solution then it would dominate the black disc.

For this reason specific methods have been proposed to find points on the Pareto frontier without using convex combination. One such is the AMOSA algorithm, which I now consider.

2.2. The AMOSA algorithm

The AMOSA algorithm (Bandyopadhyay et al., 2008; Bandyopadhyay and Sahaha, 2013) is widely used as a method to find a set of solutions from the Pareto front of multi-objective optimization problems. For example, Hermoso et al. (2015) applied AMOSA to the problem of identifying Pareto-optimal sets of management strategies for catchment rehabilitation. The solutions in the set could then be evaluated to find a preferred set of trade-offs between objective functions comprising ecological quality, sediment load and economic opportunity costs. In a quite different setting Martins et al. (2015) used the algorithm to solve problems in the design of integrated circuits, where criteria include size and proportions of the overall circuit and proximity between different groups of components.

A key concept in AMOSA is the notion of the *amount of domination* between solutions. In a case with M objective functions $f_i(\mathbf{a})$, $i = 1, 2, \dots, M$, the total amount of domination between two solutions, \mathbf{a} and \mathbf{b} is

$$\Delta d(\mathbf{a}, \mathbf{b}) = \prod_{i=1, \dots, M; f_i(\mathbf{a}) \neq f_i(\mathbf{b})} \frac{|f_i(\mathbf{a}) - f_i(\mathbf{b})|}{R_i}, \quad (1)$$

where R_i is the range of values of $f_i(\cdot)$, which may be adapted as new solutions are found by the algorithm.

Another distinctive feature of the AMOSA algorithm is its use of an archive of solutions. An archive is used to store non-dominated solutions which have so far been discovered. Over the course of the optimization the archive will tend to grow in size. The total number of stored solutions has a hard

limit, H , the largest number of solutions that the algorithm will return, and a soft limit, S , the largest number of solutions that can be held in the archive at any time. The soft limit on the archive is enforced by undertaking a single-linkage cluster analysis in the space of the objective function values, when the number of archived solutions exceeds S , so as to find H clusters. The solution closest to the centroid of each of the H clusters is retained and the rest are discarded. At points in the optimization solutions may be added to the archive, or removed and replaced by new solutions which dominate them.

The AMOSA algorithm uses simulated annealing (Aarts and Korst, 1989) for optimization. The particular advantage of simulated annealing is that it does not automatically reject solutions which are inferior to existing ones, but accepts or rejects them probabilistically, the probability of acceptance depending on how much worse the new solution is, and the magnitude of a ‘temperature’ parameter. The temperature is reduced while the algorithm runs, which reduces the probability of acceptance of a poorer solution. However, the possibility of accepting poorer solutions means that the algorithm is less likely to stick at locally optimal solutions.

A fully-detailed account of AMOSA is presented by Bandyopadhyay and Sahaha (2013), here I outline the key steps.

1. The archive is initialized with a set of solutions (which might not be Pareto optimal). The temperature of the system is set to some value, τ .
2. One of the solutions in the archive is selected at random as the ‘current solution’, \mathbf{c} . The values of the objective function are computed for this solution.
3. The current solution is perturbed randomly to generate a new solution, \mathbf{n} , and each of the M objective functions are evaluated for this solution. The two solutions, \mathbf{n} and \mathbf{c} are compared with respect to values of the objective functions to see if one dominates the other. There are three possible outcomes, addressed in the three following paragraphs.
 - (a) If \mathbf{c} dominates \mathbf{n} then \mathbf{n} is accepted as the current solution, or \mathbf{c} is retained the decision being made at random with probability of acceptance of \mathbf{n} set as

$$P_{\text{acc}} = \left(1 + \exp \left\{ \frac{\overline{\Delta d}_{k+1}}{\tau} \right\} \right)^{-1}, \quad (2)$$

where τ is the temperature parameter and $\overline{\Delta d}_{k+1}$ denotes the mean dominance of \mathbf{n} by \mathbf{c} and the archive. If k solutions in the archive dominate \mathbf{n} , and these are denoted by $\tilde{\mathbf{a}}_i$, $i = 1, \dots, k$ then

$$\overline{\Delta d}_{k+1} = \frac{\Delta d(\mathbf{c}, \mathbf{n}) + \sum_{i=1}^k \Delta d(\tilde{\mathbf{a}}_i, \mathbf{n})}{k+1}. \quad (3)$$

- (b) If \mathbf{c} and \mathbf{n} are non-dominating (i.e. neither dominates the other), then \mathbf{n} is compared with the solutions in the archive. If \mathbf{n} is non-dominating with respect to all solutions in the archive, then \mathbf{n} is added to the archive, and becomes \mathbf{c} . If \mathbf{n} dominates some solutions in the archive then \mathbf{n} is added to the archive and becomes \mathbf{c} , and the dominated solutions are removed from the archive. If k solutions, $\mathbf{a}_1, \dots, \mathbf{a}_k$, in the archive dominate \mathbf{n} then the mean dominance of \mathbf{n} by the archive $\overline{\Delta d}_k$ is computed:

$$\overline{\Delta d}_k = \frac{\sum_{i=1}^k \Delta d(\mathbf{a}_i, \mathbf{n})}{k}. \quad (4)$$

The solution \mathbf{n} is accepted as the new \mathbf{c} with probability

$$P_{\text{acc}} = \left(1 + \exp \left\{ \frac{\overline{\Delta d}_k}{\tau} \right\} \right)^{-1}, \quad (5)$$

otherwise the current \mathbf{c} is retained.

- (c) If \mathbf{n} dominates \mathbf{c} then various outcomes are possible. If \mathbf{n} dominates some of the solutions in the archive (one of which might be \mathbf{c}) then it is added to the archive, the dominated solutions are removed and \mathbf{n} becomes \mathbf{c} . If \mathbf{n} does not dominate any solutions in the archive (which implies that \mathbf{c} is not in the archive), and is not dominated by any solutions in the archive, then it is

added to the archive and becomes \mathbf{c} . It is possible that this new solution is already present in the archive, and if so the duplicate entry is removed. If k solutions, $\mathbf{a}_1, \dots, \mathbf{a}_k$, in the archive dominate \mathbf{n} (which indicates that \mathbf{c} is not in the archive) then the solution is found which dominates \mathbf{n} least, with total domination Δd_{\min} . This solution becomes \mathbf{c} with probability

$$P = \left(1 + \exp \left\{ \frac{-\Delta d_{\min}}{\tau} \right\} \right)^{-1},$$

otherwise \mathbf{n} becomes \mathbf{c} .

4. If the number of solutions in the archive exceeds S then the clustering procedure is followed to thin the archive to H solutions.
5. The process from step 3 to step 4 above is repeated some specified number (N) of times.
6. The temperature τ is reduced by multiplying it by a factor $\theta < 1$. If τ exceeds some minimum value the process above from step 3 to step 5 is repeated.

2.3. Application of AMOSA to spatial sampling

In the study reported here I applied the standard AMOSA algorithm to the spatial problem as follows.

One solution to the multi-objective optimization problem in the spatial case consists of a set of spatial coordinates for n sample points. The output of the AMOSA algorithm for a spatial problem therefore consists of H (the hard limit on the archive size) different sets of spatial coordinates, each for n sample points. Each such set is a proposed sample design representing some trade-off between the objective functions, and not dominating, not dominated by any other solution in the set. As described in Section 2.2, there are N repeats of steps 3 to step 4 at some fixed temperature τ_k , each of which is initiated by perturbing the current solution. In the spatial adaptation of AMOSA I specified $N = K \times n$ repeats (recall that n is the number of sample sites to be positioned), which comprise K cycles in each of which the current solution was perturbed by changing the location of the i th sample location, where i takes values $1, 2, \dots, n$ in each cycle. I selected a value of K by trial and error so that the overall convergence time of the algorithm was acceptable.

The sample location which is changed is moved by a random distance in a random direction. Note that the coordinates of sample points, as implemented here, are continuous rather than discrete. The maximum distance, d_m is fixed and the actual distance is generated at random from a uniform distribution bounded by $[0, d_m]$. If the perturbation removes the point out of the domain of interest then it is replaced and a different random perturbation is tested. To avoid degenerate solutions a minimum distance between any two sample points within a solution is also enforced at this stage.

It is necessary to select an initial temperature for the algorithm, a minimum temperature and a factor θ by which the temperature is periodically reduced. In standard simulated annealing it is common practice to use a 'cooling schedule', rules of thumb for the value of the initial temperature and the rate of cooling (Kirkpatrick et al., 1983). In similar fashion, Bandyopadhyay and Sahaha (2013) recommend that the probability of acceptance of new solutions that are dominated should be close to 0.5 in initial iterations of the algorithm. Note from Eq. (2) that this is the limit of the probability of acceptance as $\tau \rightarrow \infty$. They propose that $0.5 \leq \theta \leq 0.99$. These guidelines were followed here. There is no established rule to define the target minimum temperature, and so the total number of iterations at step (6). I followed Bandyopadhyay et al. (2008) by running the algorithm until the proportion of accepted dominating changes was close to zero for several hundred iterations.

2.4. Case study

To illustrate the application of the AMOSA algorithm to spatial sampling I consider a hypothetical case study. The basic problem is the one set out in Section 1.1.1, sampling to form an aggregate (bulk) sample of soil material from a region of interest. The domain to be sampled is irregular—a 1-ha square plot excluding a quarter-circle area of radius 40 m centred on one corner of the plot (location $\{100, 100\}$ in the local coordinate scheme). It is assumed that the variable to be sampled shows spatial correlation with a spherical variogram range 100 m, nugget variance 500 and correlated variance 4500. The individual who collects the samples will enter the domain at the corner $\{0, 0\}$ in the local

coordinate scheme, and will leave the domain at the same location. The sample is of a medium such as soil from which an aggregate specimen can be formed of equal aliquots from each sample point, and for a variable for which a measurement on the aggregated specimen is equivalent to an arithmetic average of measurements on the separate aliquots. This measurement will be treated as an estimate of the mean of the variable across the domain. A total of 20 sample points are to be distributed across this domain. This is a reasonable sample size to form an aggregate sample for, for example, soil nutrient analysis for an agricultural field (Marchant et al., 2012).

The first objective function to be considered in optimization of the sample design is the variance of the sample mean. Because the sample sites are not selected independently and at random I computed a model-based estimate of the variance of the sample mean, given the variogram parameters, following Webster and Oliver (1990):

$$\sigma_m^2 = \frac{2}{n} \sum_{i=1}^n \bar{\gamma}(\mathbf{x}_i, \mathcal{B}) - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma(\mathbf{x}_i - \mathbf{x}_j) - \bar{\gamma}(\mathcal{B}, \mathcal{B}), \quad (6)$$

where \mathbf{x}_i is a vector that denotes the location of the i th out of n sample sites, $\gamma(\mathbf{h})$ is the variogram of the variable, \mathcal{B} denotes the domain to be sampled,

$$\bar{\gamma}(\mathbf{x}_i, \mathcal{B}) = \int_{\mathbf{x}_k \in \mathcal{B}} \gamma(\mathbf{x}_i - \mathbf{x}_k) d\mathbf{x}_k,$$

and

$$\bar{\gamma}(\mathcal{B}, \mathcal{B}) = \int_{\mathbf{x}_k \in \mathcal{B}} \int_{\mathbf{x}_l \in \mathcal{B}} \gamma(\mathbf{x}_k - \mathbf{x}_l) d\mathbf{x}_l d\mathbf{x}_k,$$

where the integrals are over the two-dimensional space of \mathcal{B} . The integrals are evaluated numerically. The third term on the right-hand side of Eq. (6) does not depend on the distribution of sample sites so is evaluated only once. Note that this variance is not a kriging variance, because all cores are equally weighted in the aggregate mean and no weights are computed.

The second objective function to be considered is the total distance walked to visit all points, starting and leaving from the specified corner of the site. As discussed in Section 1.1.1, it is expected that there would be a trade-off between distance walked around the sample points and the variance of the mean. The latter objective function will be minimized by a sample design which gives good spatial coverage, but the distance walked to visit all these sites will be larger than for a design with greater clustering of sites. In this study the total distance to visit a proposed set of sample points was evaluated by finding the shortest route around the points, including the entrance to the domain. This route was found with the solve_TSP procedure in the TSP package for the R platform (Hahsler and Hornik, 2007; R Core Team, 2014). I used the arbitrary insertion algorithm with the two-edge exchange improvement procedure (Hahsler and Hornik, 2007).

The AMOSA algorithm, adapted for the spatial sampling problem as described in Section 2.3, was encoded in the R environment (R Core Team, 2014). It was run to find a Pareto optimal set of solutions. As noted, 20 sample points were to be deployed. Three sets of sample locations were found to initiate the algorithm by a weighted spatial simulated annealing. The sets of weights applied to the variance and the total distance respectively were {0.17, 0.83}, {0.83, 0.17} and {0.98, 0.02}. These were found by trial and error to give contrasting solutions. The spatial distribution of the sample points for these three initial solutions is shown in Fig. 2. Maximum values of the two objective functions found in the course of this exploration were used to define the range of each function required in Eq. (1).

The soft length limit on the archive was 60 solutions and the hard limit was 40. At any given temperature of the system a total of 20 iterations of the algorithm were run, one cycle over the 20 indices for sample locations. The iterations at any temperature were initialized by selecting an archived solution at random as a current set \mathbf{c} . The maximum distance by which any sample point could be moved in a single perturbation was set at 5 m. A minimum distance of 1 m between any sample site and its nearest neighbour was imposed. The initial temperature of the annealing was found by trial and error to give initial acceptance rates of proposed \mathbf{n} dominated by \mathbf{c} which varied either side of 0.5. The temperature was modified after each of $N = 20$ iterations from τ_k to $\tau_{k+1} = \theta \tau_k$

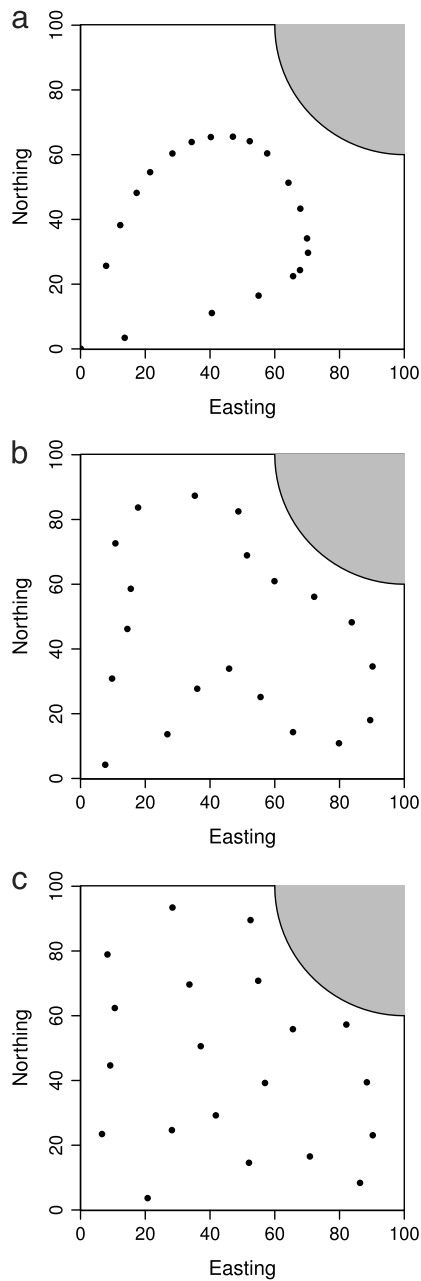


Fig. 2. Initial sample designs obtained by simulated annealing on a weighted combination of variance and distance, then used to initiate the AMOSA algorithm. Weights on variance and distance are, respectively (a) {0.17, 0.83}, (b) {0.83, 0.17} and (c) {0.98, 0.02}. The square region, excluding the quarter-circle shown in grey, is the domain for sampling.

with $\theta = 0.99$. In this case study I allowed 2000 successive cooling steps of the annealing. I then examined a plot of the proportion of dominated \mathbf{n} accepted at each successive temperature, and ensured that the algorithm had run sufficiently long for this proportion to have fallen to zero for most iterations.

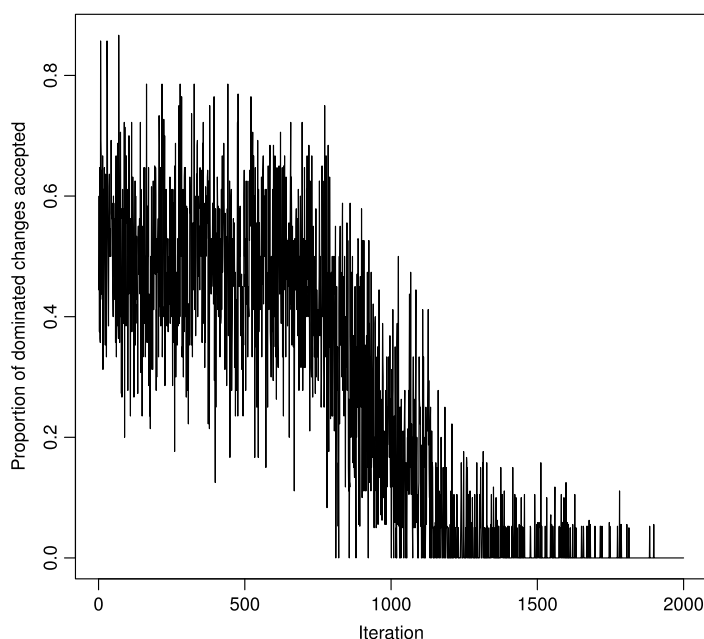


Fig. 3. Proportion of dominated changes accepted in successive iterations of the AMOSA algorithm in the first run of the case study.

The output of any numerical optimization algorithm depends on the initial conditions. To test the stability of the output of the algorithm I ran it a further four times, each time initializing the archive with a different set of three solutions.

The dispersion variance of the variable within the domain (the third term in Eq. (6)) was 3360, so the variance of the mean of a simple random sample of 20 points would be 168. For comparison with points on the Pareto front I also generated simple random samples within the domain and computed the total distance to visit them, starting and ending at the specified corner of the domain. I did this 10 000 times. The mean (and median) distance was 400 m, and the first and third quartiles were 380 m and 420 m respectively.

3. Results

Fig. 3 shows the proportion of dominated \mathbf{n} accepted in each of 2000 successive iterations in the first run of the algorithm. At the start the proportion varied either side of 0.5, and then declined. During most of the last 200 or so iterations no dominated \mathbf{n} was accepted.

These 2000 iterations took approximately 10 h to run (3.40 GHz processor with 8 GB RAM) with a hard limit of 40 solutions in the archive. One run of the spatial simulated annealing algorithm to optimize a weighted combination of the two objective functions took approximately 15 min (there was no evidence that the speed of the algorithm varied with the weights assigned to the different objective functions). It follows that generating the same number of sample designs by successive spatial simulated annealing to optimize a weighted combination of the two objective functions would take about the same length of time as the AMOSA algorithm. However, it should be pointed out that there may be scope to come up with more efficient stopping rules for the application of AMOSA to optimizing spatial sampling.

Fig. 4 shows the set of forty solutions obtained by the first run of the algorithm as an estimate of the Pareto front with respect to the two objective functions (black discs). The three crosses are the solutions used to initialize the algorithm, which had been obtained by minimizing weighted combinations of the two objective functions. One of these (the solutions with the shortest total

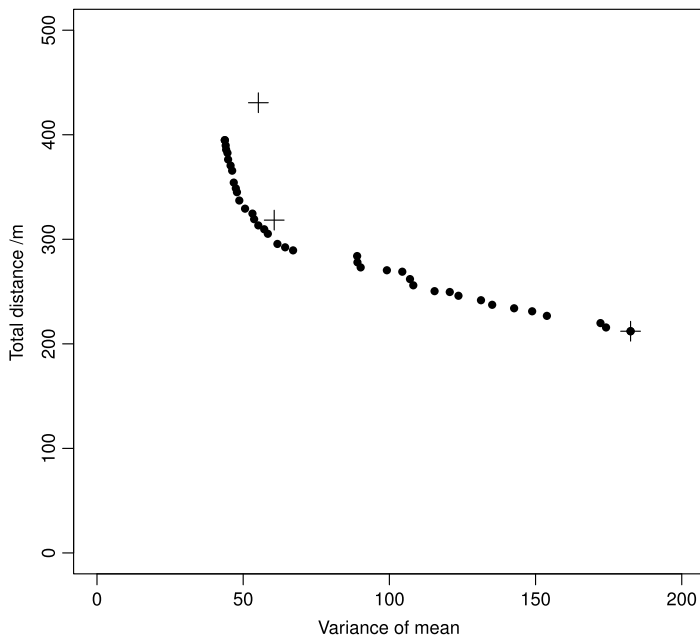


Fig. 4. Values of the two objective functions for forty solutions from the first run of the algorithm (black discs). Crosses show solutions used to initiate the algorithm.

distances) was non-dominating with respect to the rest of the set (and remained a member of it). However, the other two initial solutions (longest distance) were dominated by solutions from the AMOSA algorithm which achieve smaller variances with a smaller total distance.

Fig. 5(a) shows the values of variance and distance for the original run of the algorithm (same values as in Fig. 4) and for the four additional runs. The outputs of the five different runs are reasonably consistent. However, the solutions from the original run with variances between about 65 and 105 are dominated by solutions from one or more of the four additional runs. There is rather more variation between runs for solutions with variances between about 90 and 120 than there is elsewhere in the full set of solutions. I extracted the non-dominated subset of all 200 solutions obtained from the five runs. The variances and distances for these solutions are plotted in Fig. 5(b). This is treated as an estimate of the Pareto front.

Fig. 6 shows five solutions from the subset of solutions on the estimated Pareto front that achieve a variance of no more than 50 with a total distance walked of no more than 375 m. Fig. 7 shows five solutions that achieve a variance of no more than 100, with the total distance walked no more than 300 m. Fig. 8 shows the five solutions with smallest total variance, and Fig. 9 shows the five solutions with smallest distance.

4. Discussion

The form of the estimated Pareto front shows how trade-offs between the two objective functions work for this particular problem. Note, for example, that a doubling of the variance of Pareto solutions from 65 to 130 is accompanied by a reduction in total distance of about 50 m, less than 20%. Similarly substantial increases in distance are required to achieve the smallest variances. These very sharply diminishing returns in the trade-off between objectives suggest that is inefficient to operate in that part of the Pareto front where variances are much less than 50 or distances are much less than 250 m. In this case the expected variance and distance for a simple random sample (168 and 400 respectively) are some way above the Pareto front, reflecting the strong spatial correlation in this example. Note that the Pareto front, as estimated from the multiple runs, appears to be convex, at least over the range

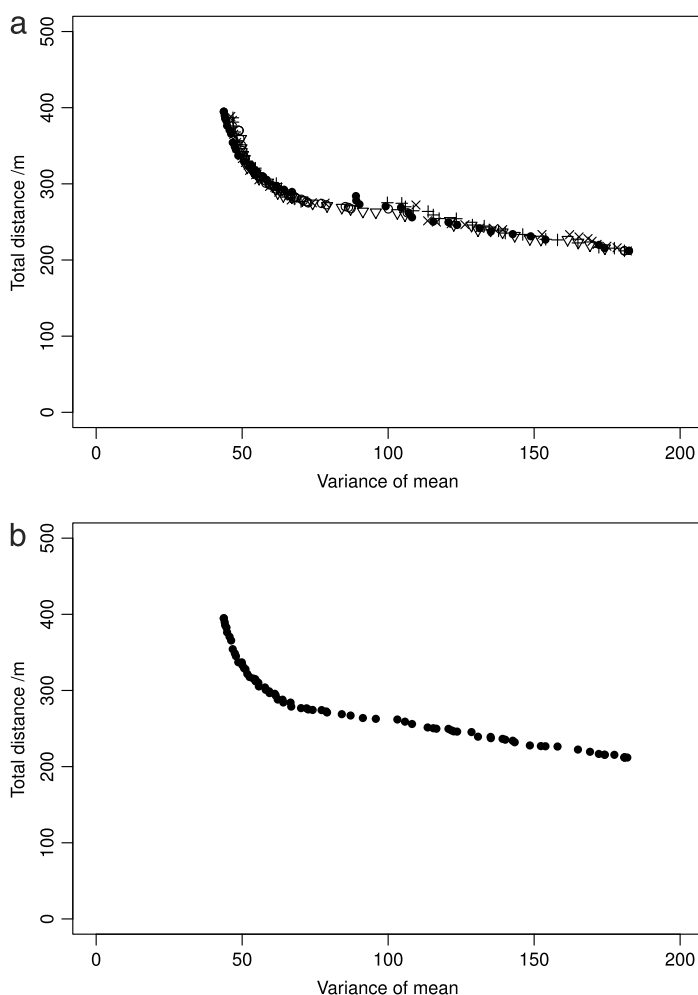


Fig. 5. (a) Forty solutions from each of five separate runs of the algorithm starting from different initial solutions. The initial run (in Fig. 4) is shown as black discs. (b) Non-dominating subset of solutions from those shown in (a), an estimate of the Pareto front.

of variances and distances considered. This implies that all points on the front can, at least in principle, be discovered by optimizing a weighted combination of the objective functions.

Having identified the form of the front, one could then identify a range of acceptable values for the two objective functions, given constraints. In the case illustrated by Fig. 6 the maximum variance is specified as 50, and the route around the sites should be no more than 375 m. Fig. 6 shows 5 solutions from the estimated Pareto front that meet these criteria. Note that none of the initial solutions (Fig. 2) achieve these criteria—all have too large a variance. The design in Fig. 2(c), which has the smallest variance of the initial designs, appears to be a loose grid. The designs in Fig. 6 which achieve the smallest variances, Figs. 6(e) and 5(d), have similar form. The other three designs in this set, with slightly larger variances, have sample points on a markedly non-convex closed ‘loop’ which explores the domain.

In Fig. 7 we consider a case where larger variances are accepted (less than 100) and the total distance must be less than 300 m. All the solutions involve a much tighter loops than in Fig. 6. There is something of a discontinuity in this subset. The two designs with smallest variances, Fig. 7(d)–(e)

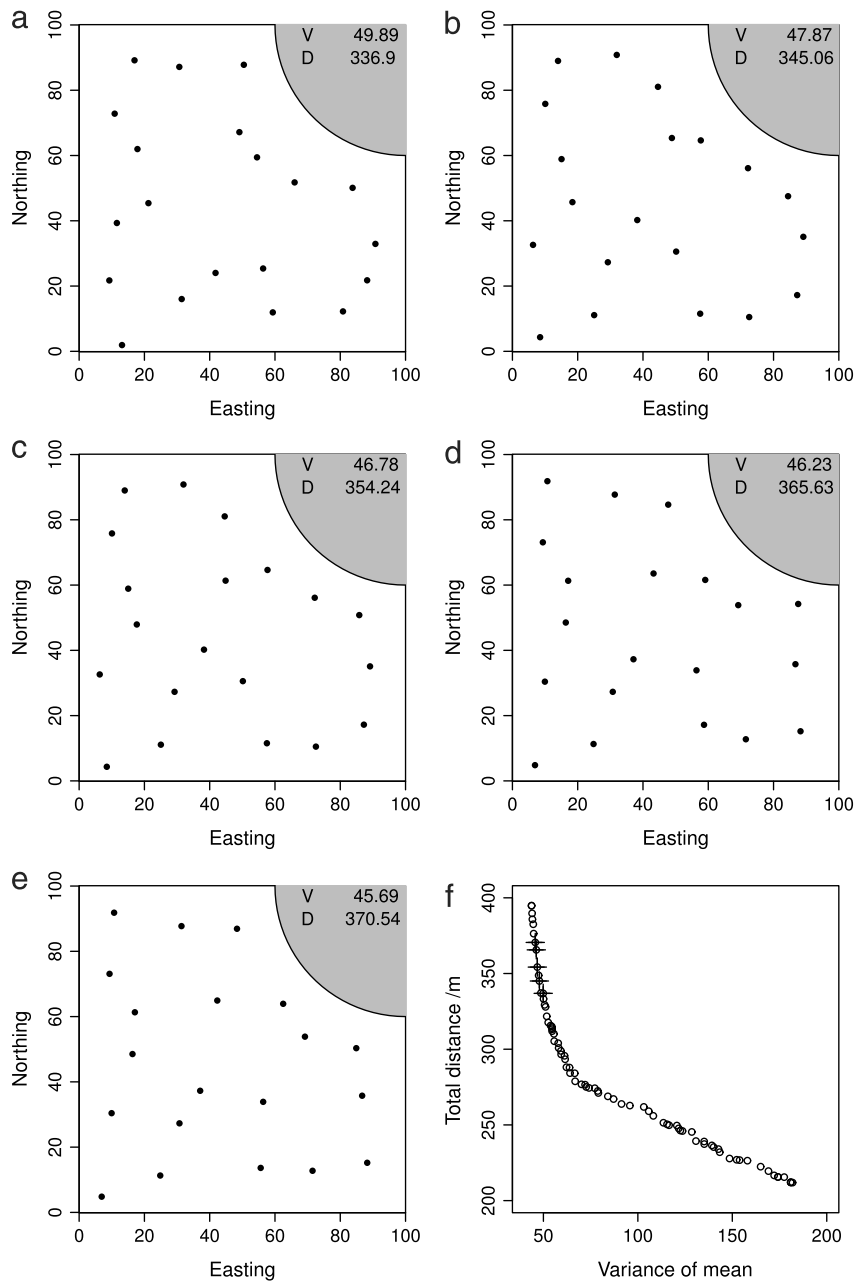


Fig. 6. (a–e) Five sample designs from the estimated Pareto front which meet the criterion that the variance is less than 50 and the distance walked is no more than 375 m, and (f) plot of the objective functions on the estimated Pareto front (excluding the origin for clarity) with the illustrated designs shown by crosses.

resemble the strongly non-convex loops illustrated in Fig. 6(a)–(c), but are markedly less involuted. The three designs with shortest total distances, Fig. 7(a)–(c), have a loosely pentagonal form with one or two additional points near some of the vertices which are closer than the vertices to the edges of the domain.

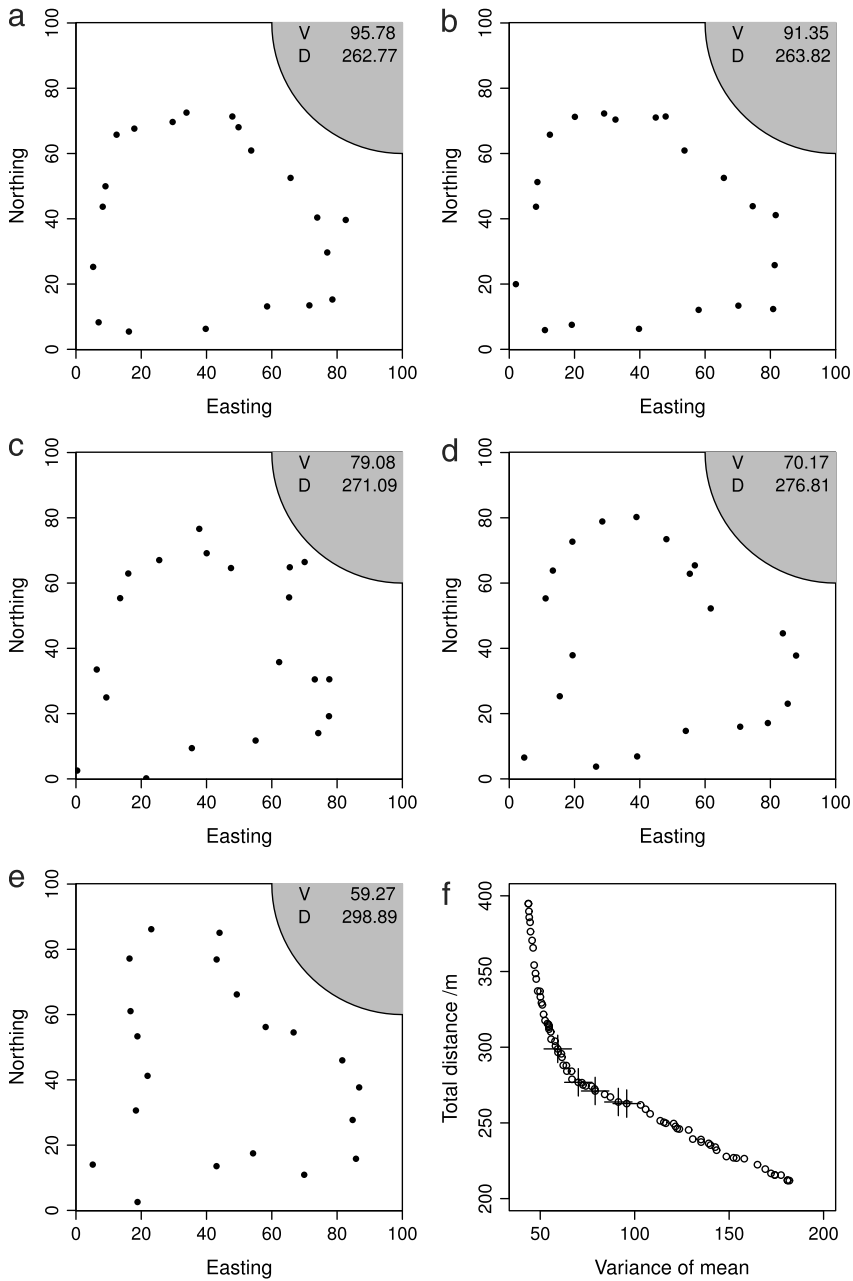


Fig. 7. (a–e) Five sample designs from the estimated Pareto front which meet the criterion that the distance walked is no more than 300 m and the variance is no more than 100, and (f) plot of the objective functions on the estimated Pareto front (excluding the origin for clarity) with the illustrated designs shown by crosses.

Figs. 6 and 7 illustrate how the solutions on the estimated Pareto front might be used in practice for sample design. Given the task, we may specify a maximum acceptable variance for the estimate; and, given the logistical and cost constraints, we may specify a maximum distance to walk. This allows us to identify the subset of solutions on the front that are acceptable by these criteria. We know that

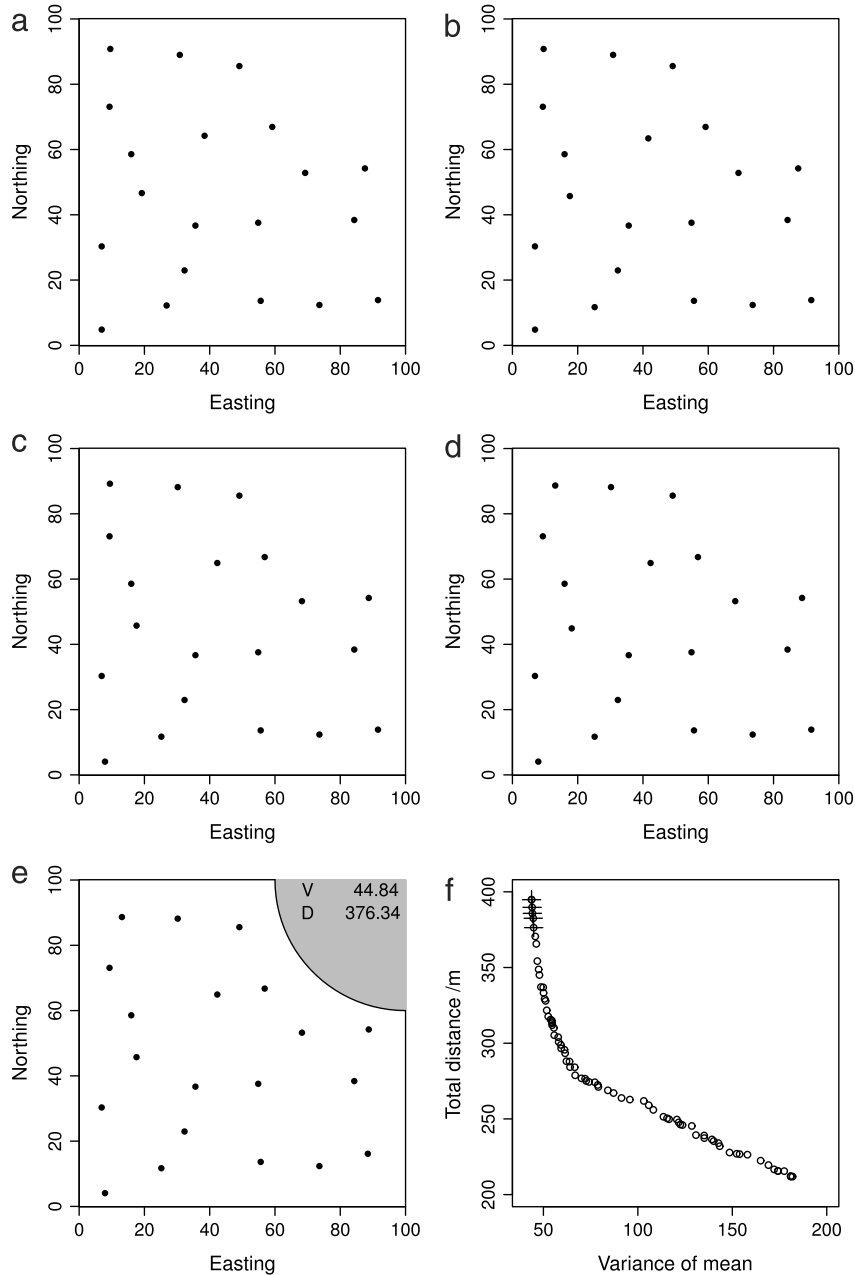


Fig. 8. (a–e) Five sample designs from the estimated Pareto front for which the variance is smallest, and (f) plot of the objective functions on the estimated Pareto front (excluding the origin for clarity) with the illustrated designs shown by crosses.

the set is Pareto efficient (switching between solutions will improve one of the objective functions, but make the other worse). The selection of a solution to use in practice may then depend on how data quality is prioritized relative to cost—one might select the solution with minimum variance if the quality of the resulting estimate is regarded as most important.

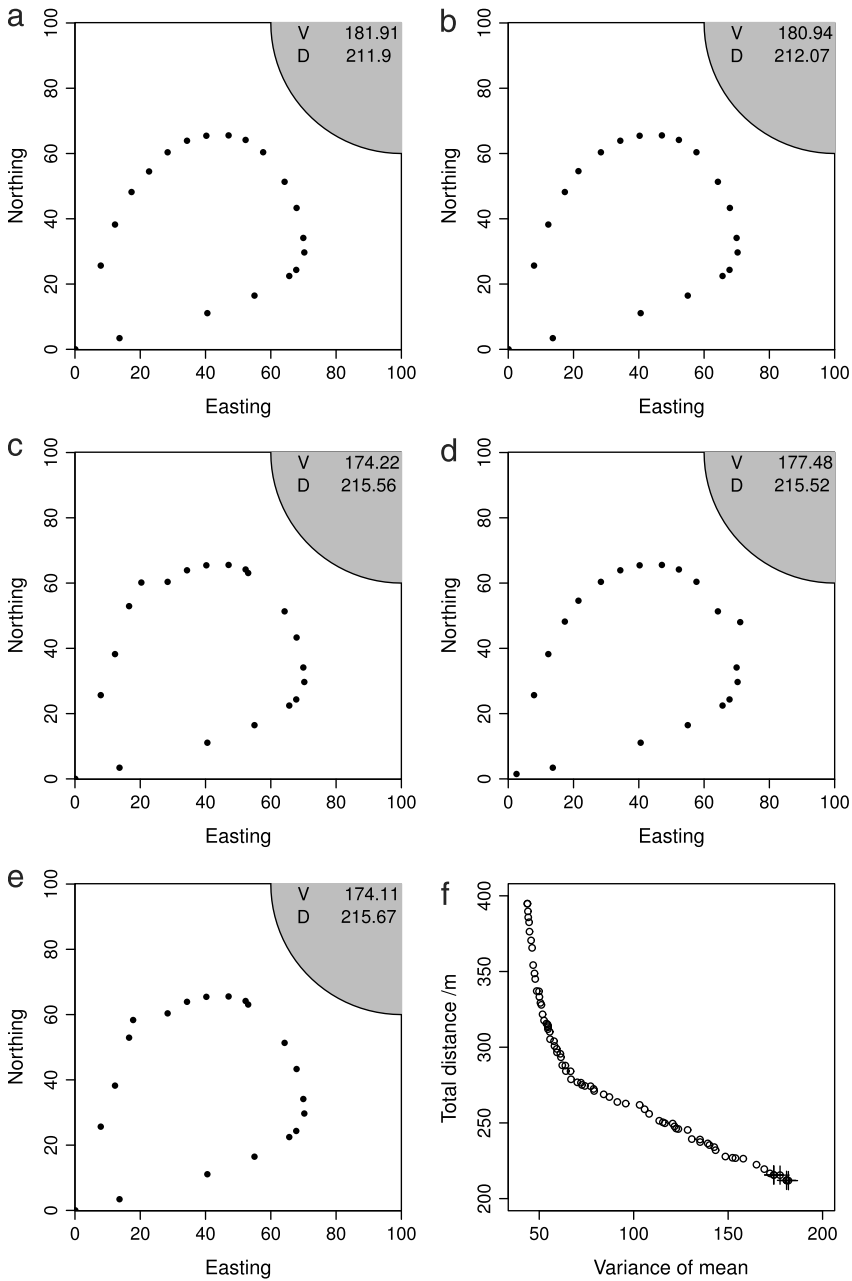


Fig. 9. (a–e) Five sample designs from the estimated Pareto front for which the distance is smallest, and (f) plot of the objective functions on the estimated Pareto front (excluding the origin for clarity) with the illustrated designs shown by crosses.

Fig. 8 shows the five solutions with smallest variance. These are all very similar with respect to variance, and differ only slightly with respect to total distance. They resemble loose grids, similar to the contiguous solutions on the Pareto front illustrated in Fig. 6(e) and (d).

Fig. 9 shows the five solutions with smallest distance. Fig. 9(a) is the solution inherited from the initiation of the algorithm, also in Fig. 2(a), which is a tight, smooth, loop. The solution in Fig. 9(b)

differs only slightly and the distribution of sample points is visually indistinguishable from 9(a). The three solutions with smallest variance, Fig. 9(d)–(e), have a rather less smooth path around the loop, and could be regarded as transitional to the irregular pentagonal shape seen in Fig. 7(a) and (b).

Fig. 5(a) shows that, while the results of five runs of the AMOSA algorithm are reasonably consistent, the full set of solutions from all runs is not completely non-dominated, i.e. some of the solutions dominate others. It is not unusual to have to run a numerical optimization procedure more than once to get robust solutions, and it is clear that we can have more confidence in a non-dominated subset of solutions from several runs of AMOSA than from a single run; Bandyopadhyay et al. (2008) do this.

The need to run the algorithm more than once raises a second practical problem. The runs of the AMOSA algorithm reported here took about 10 h to complete. It was interesting to note that this was commensurate with the time to complete a run of a simple spatial simulated annealing algorithm to minimize a weighted combination of the two objective functions to generate a single solution. It would take the same time to generate forty solutions by minimizing weighted combinations of the two objective functions as it does to generate forty solutions (the hard limit on the archive size here) by the AMOSA algorithm. Furthermore, as noted above, it can be difficult to specify weights *a priori* to give good coverage of the Pareto front, and optimizing a weighted combination of the objective functions cannot find non-convex regions of the front.

There might be scope to improve the efficiency of the AMOSA algorithm for spatial sampling problems by developing robust rules of thumb for suitable cooling schedules. In particular it is necessary to understand the sensitivity of the procedure to the cooling rate of the annealing, and to develop appropriate stopping rules. In this study I followed Bandyopadhyay et al. (2008) by running the algorithm until the proportion of dominated changes that are accepted is small. However, a more efficient stopping rule might be based on the magnitude of changes in the solutions in the archive. From observation of the five runs conducted in this study this might allow a smaller number of iterations. It might also be possible to improve the time efficiency of the overall analysis by running multiple runs of AMOSA in parallel, perhaps using a common archive which is updated at intervals.

It is interesting to see how this algorithm generates a set of solutions which show a transition from loose grids to tightly involuted closed loops, to a less involuted loop, to an irregular pentagon with some additional points near the vertices to a tight smooth loop around the domain. This continuum is conditioned, of course, by the size and shape of the domain, the sample size and the covariance parameters. Implicit in these solutions are rules for sample design, and it would be useful to develop an approach to extract those rules from the algorithm's output. These rules of thumb would be more computationally efficient to apply than the original AMOSA algorithm. This is the 'innovization' (innovation by optimization) approach which is gaining attention in engineering (e.g. Deb et al., 2014). It would be particularly useful for this problem where the optimization procedure is computationally demanding.

To illustrate, in cases where AMOSA provides evidence that the Pareto front for a particular problem is convex, at least in the region of solutions of practical interest (as seems to be true in this case study) then it is possible, in principle, to find solutions on the Pareto front by optimizing a weighted combination of the objective functions. The practical problem is how to increment the weights so as to get a practically useful, more-or-less even, distribution of points along the Pareto front (Das and Dennis, 1997). In an innovization study one might hypothesize that a particular rule for incrementing the weights (e.g. in a logarithmic sequence) will generate a reasonable spacing of Pareto optimal solutions, and then test this rule against the AMOSA output. Such a rule would then allow one to explore limited and interesting regions of the Pareto front for comparable problems by weighted optimization of the objective functions.

The objective of this paper was to introduce the AMOSA algorithm as a tool for multi-objective optimization of spatial sampling. One particular case study has been used, but clearly the approach could be extended to other problems where a logistical objective function could be added to a statistical quality measure. Two examples were already introduced in Section 1.1.2 (wireless sensor network design) and 1.1.3 (unmanned aerial vehicle deployment). Other possible cases include sampling for geostatistical interpolation (e.g. following Brus and Heuvelink, 2007), sampling to update existing information, where some parts of the region may require more intensive sampling than others (e.g. following Marchant et al., 2013), and possibly sampling schemes to investigate nested scales

of spatial variability in efficient reconnaissance surveys (e.g. following Lark, 2011). It would also be interesting to apply the AMOSA algorithm to other problems where the problem of multi-objective optimization has been addressed by weighting the different objective functions. One might expect that the probability of detection of plumes or similar contamination features would show a rather different trade-off with total distance than does the variance of the spatial mean, as in the previously-cited study by Melles et al. (2011).

In the case study developed here the variable of interest is stationary in the mean and the variance. In some cases there may be non-stationarity in the mean which can be described by a partition of the study area into subregions (e.g. soil map units, different agricultural fields) or by considering a continuous covariate (e.g. a remote sensor image) with which the target variable is correlated. In such circumstances one may develop an alternative expression to Eq. (6) for the estimation variance of the sample spatial mean. This expression could be based on the universal block cokriging variance (Cressie, 1993). The AMOSA algorithm could then be used to find the Pareto front for this objective function and the total distance walked between sample points. A more interesting case may be when the subregions differ with respect to variance (and possible spatial autocorrelation as well). In this situation a total sample effort can be partitioned between the strata in some way. If the target quantity is a regional mean, then this is a relatively straightforward extension of the case study in this paper. If the target quantity of interest is a local (kriging) prediction, and the objective function to be minimized is the kriging variance, then it may be questionable whether simply averaging this over all strata is appropriate. This could give rise to a multiobjective optimization problem where the objective functions include mean kriging variances within distinct subregions. These extensions of the case study presented here would be an interesting topic for further research.

5. Conclusions

It has been shown how the AMOSA algorithm of Bandyopadhyay et al. (2008) can be adapted to the optimization of spatial sampling designs with more than one objective function. In the example considered here it was shown how a Pareto set of solutions, where the objective functions are total distance travelled for sampling and the variance of the sample mean, form a continuum from a tight loop (short distance) to a loose grid (small variance). The trade-offs between these functions show considerable sensitivities near the extremes (large increases in variance are incurred for small reductions in distance, and vice versa). It is possible to identify a subset of the Pareto set that meets some predetermined limits, and to select from these a sample design which could be used in practice. There is considerable scope to develop this area of work, in particular it would be interesting to see how the approach could be used to optimize the layout of wireless sensor networks or the deployment of unmanned aerial vehicles equipped with sensors.

Acknowledgement

This paper is published with the permission of the Executive Director of the British Geological Survey (NERC). I am grateful to Professor David Corne for pointing me to the literature on innovization.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.spasta.2016.09.001>.

References

- Aarts, E., Korst, J., 1989. *Simulated Annealing and Boltzmann Machines - A Stochastic Approach to Combinatorial Optimization and Neural Computing*. John Wiley & Sons, New York.
- Anderson, K., Gaston, K.J., 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Front. Ecol. Environ.* 11, 138–146.
- Babel, L., 2013. Three-dimensional route planning for unmanned aerial vehicles in a risk environment. *J. Intell. Robot. Syst.* 71, 255–269.

- Ballari, D., de Bruin, S., Bregt, A.K., 2012. Value of information and mobility constraints for sampling with mobile sensors. *Comput. Geosci.* 49, 102–111.
- Bandyopadhyay, S., Sahaha, S., 2013. *Unsupervised Classification Similarity Measures, Classical and Metaheuristic Approaches, and Applications*. Springer-Verlag, Berlin.
- Bandyopadhyay, S., Sahaha, S., Maulik, U., Deb, K., 2008. A simulated annealing-based multiobjective optimization algorithm: AMOSA. *IEEE Trans. Evol. Comput.* 12, 269–283.
- Beckett, P.H.T., 1981. Logistics of agricultural extension — Foreword and Part I: The component parts of a logistic model. *Agric. Administ.* 8, 177–208.
- Boer, E.P.J., Dekkers, A.L.M., Stein, A., 2002. Optimization of a monitoring network for sulfur dioxide. *J. Environ. Qual.* 31, 121–128.
- Brus, D.J., Heuvelink, G.B.M., 2007. Optimization of sample patterns for universal kriging of environmental variables. *Geoderma* 138, 86–95.
- Cares, J.R., Dickmann, J.Q., 2016. *Operations Research for Unmanned Systems*. John Wiley & Sons, Chichester.
- Cressie, N.A.C., 1993. *Statistics for Spatial Data*, revised ed. John Wiley & Sons, New York.
- Das, I., Dennis, J.E., 1997. A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Struct. Optim.* 14, 63–69.
- Deb, K., Bandaru, S., Greiner, D., Gaspar-Cunha, A., Tutum, C.C., 2014. An integrated approach to automated innovation for discovering useful design principles: case studies from engineering. *Appl. Soft Comput.* 15, 42–56.
- De Gruijter, J., Brus, D., Bierkens, M.F.P., Knotters, M., 2006. *Sampling for Natural Resource Monitoring*. Springer-Verlag, Berlin.
- Faye, E., Rebaudo, F., Yáñez-Cajo, D., Cauvy-Fraunié, S., Dangles, O., 2016. A toolbox for studying thermal heterogeneity across spatial scales: from unmanned aerial vehicle imagery to landscape metrics. *Methods Ecol. Evol.* 7, 437–446.
- Hahsler, M., Hornik, K., 2007. TSP - Infrastructure for the traveling salesperson problem. *J. Stat. Softw.* 23 (2), URL: <http://www.jstatsoft.org/v23/i02/>.
- Hermoso, V., Pantus, F., Olley, J., Linke, S., Mugodo, J., 2015. Prioritizing catchment rehabilitation for multi objective management: an application from SE-Queensland, Australia. *Ecol. Modell.* 316, 168–175.
- Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimisation by simulated annealing. *Science* 220, 671–680.
- Lark, R.M., 2002. Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. *Geoderma* 105, 49–80.
- Lark, R.M., 2011. Spatially nested sampling schemes for spatial variance components: scope for their optimization. *Comput. Geosci.* 37, 1633–1641.
- Lark, R.M., Knights, K.V., 2015. The implicit loss function for errors in soil information. *Geoderma* 251–252, 24–32.
- Marchant, B.P., Dailey, A.G., Lark, R.M., 2012. Cost-effective sampling strategies for soil management. Home-Grown Cereals Authority Research and Development Project Report No. 485. HGCA, London, Available at: <https://cereals.ahdb.org.uk/media/252469/pr485.pdf>.
- Marchant, B.P., Lark, R.M., 2007. Optimized sample schemes for geostatistical surveys. *Math. Geol.* 39, 113–134.
- Marchant, B.P., McBratney, A.B., Lark, R.M., Minasny, B., 2013. Optimized multi-phase sampling for soil remediation surveys. *Spat. Stat.* 4, 1–13.
- Martins, R., Lourenço, N., Horta, N., 2015. Multi-objective optimization of analog integrated circuit placement hierarchy in absolute coordinates. *Expert Syst. Appl.* 42, 9137–9151.
- McShane, G., James, M.R., Quinton, J., Anderson, K., DeBell, L., Evans, M., Farrow, L., Glendell, M., Jones, L., Kirkham, M., Lark, R.M., Rawlins, B.G., Rickson, J., Quine, T., Wetherelt, A., Brazier, R., 2016. Comparing and combining terrestrial laser scanning with ground-and UAV-based imaging for national-level assessment of soil erosion. *Geophys. Res. Abstr.* 18, 5576.
- Melles, S.J., Heuvelink, G.B.M., Twenhöfel, C.J.W., van Dijk, A., Hiemstra, P.H., Baume, O., Stöhlker, U., 2011. Optimizing the spatial pattern of networks for monitoring radioactive releases. *Comput. Geosci.* 37, 280–288.
- Othman, M.F., Shazali, K., 2012. Wireless sensor network applications: a study in environmental monitoring system. *Procedia Eng.* 41, 1204–1210.
- R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Shaw, R., Lark, R.M., Williams, A.P., Chadwick, D.R., Jones, D.L., 2016. Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture. *Agric. Ecosys. Environ.* 230, 294–306.
- Stein, M.L., 1999. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag, New York.
- van Groenigen, J.W., Stein, A., 1998. Constrained optimization of spatial sampling using continuous simulated annealing. *J. Environ. Qual.* 27, 1078–1086.
- Wang, J.H., Ge, Y., Heuvelink, G.B.M., Zhou, C.H., 2014. Spatial sampling design for estimating regional GPP with spatial heterogeneities. *IEEE Geosci. Remote Sens. Lett.* 11, 539–543.
- Webster, R., Oliver, M.A., 1990. *Statistical Methods in Soil and Land Resource Survey*. Oxford University Press, Oxford.
- Yang, S., Adeel, U., McCann, J.A., 2013. Selfish Mules: social profit maximization in sparse sensor networks using rationally-selfish human relays. *IEEE J. Sel. Areas Commun.* 31, 1124–1134.
- Yang, S., McCann, J.A., 2014. Distributed optimal lexicographic max–min rate allocation in solar-powered wireless sensor networks. *ACM Trans. Sensor Netw.* 11, 1550–4589.