Efficient spatial designs using Hausdorff distances and Bayesian optimisation

Jacopo Paglia[1], Jo Eidsvik[1], and Juha Karvanen[2]

[1] Norwegian University of Science and Technology
[2]University of Jyvaskyla

Abstract

An iterative Bayesian optimisation technique is presented to find spatial design configurations of data that carry much information. Within this setting, Gaussian process approximations enable fast calculations of expected improvement for a large number of designs, while the full-scale evaluations are only done for the most promising designs at each iteration. The Hausdorff distance is here used to model the similarity between design configurations in the Gaussian process covariance representation, and this allows the suggested algorithm to learn across different designs. The applications are related to natural resources, and we use the decision theoretic notion of value of information as a design criterion. We study properties of the Bayesian optimisation design algorithm in a synthetic example and real-world examples from forest conservation decisions and petroleum drilling operations. In the synthetic example we consider a model where the exact solution is available and we run the algorithm under different versions of this example and also compare it with existing approaches such as sequential selection and exchange algorithm. In the forestry and petroleum applications, we discuss the results obtained by the algorithm and compare it with others. Overall the suggested methodology allows an efficient selection of design with large value of information.

1 Introduction

This paper is inspired by challenging decision situations in the earth and environmental sciences. In these situations, data are gathered to support decisions about resource management. Data acquisition and processing is often costly, and it is then important to choose the sampling design wisely. There exist several common design or information criteria, see e.g. Ryan et al. (2016) for a recent review. For decision makers, value of information (VOI) analysis is useful in this context (Abbas and Howard, 2015; Eidsvik et al., 2015), as it is directly connected with the information gain associated with the decision situation and it provides a bound on the expected monetary amount one should be willing to pay for data to aid in resolving this decision situation.

We focus on designing experiments in spatial domains. Using VOI analysis, we aim to provide the decision maker with efficient survey designs including the optimal number of measurement locations and their spatial configuration. We assume that the spatial domain is discretised to a grid so that there is a finite set of possible observation locations. Moreover, we limit scope to static designs (Diggle and Lophaven, 2006; Dobbie et al., 2008; Huan and Marzouk, 2013), where the experimental configuration is selected once, at the onset of data gathering. The alternative is sequential data gathering, where the design can be adapted based on the observations made in the first (batches of) measurements (Drovandi et al., 2013; Eidsvik et al., 2018; Binois et al., 2019), but this is not always possible in practical experimental planning, which must comply with project management and budgetary limitations.

As pointed out by several others, this design problem is not trivial as the number of possible designs grows combinatorially fast. Royle (2002) proposed a random exchange algorithm to search for the optimal design. García-Ródenas et al. (2020) presented an interesting overview of some of the main algorithms for finding efficient designs. Weaver et al. (2016) and Overstall and Woods (2017) applied Bayesian optimisation to focus

the search for good designs. We use a Gaussian process (GP) model enabling fast computation of the expected improvement (EI) in Bayesian optimisation. This is combined with techniques from search algorithms, to find efficient spatial designs. A novel idea of the current paper, which is an important building block in our approach, is to use the Hausdorff distance between various designs to correlate outcomes of similar point configurations, within a realistic statistical model. Even though our focus is on spatial decision situations and design, we believe that this approach can also be applicable to other big-data challenges (Drovandi et al., 2017) and active learning approaches (Settles, 2012; Bouneffouf, 2016), where the challenge is more related to which data to process for learning and improved classifications.

The paper is organized as follows. In Section 2 we describe the spatial design problem in mathematical detail and define the VOI criterion which we use as a practically relevant information measure. In Section 3 we outline the Bayesian optimisation approach using Hausdorff distances to borrow information among similar designs. In Section 4 we study the properties of the methodology via simulations. In Section 5 we show results on two examples to demonstrate possible applications of the methods. The first one regards forest management and conservation (Kangas et al., 2008; Eyvindson et al., 2017). The decision maker must choose if stands in a forest area should be conserved or harvested (Eyvindson et al., 2019). Biological data are valuable to learn the uncertain ecological profits related to species diversity. The second application regards decision making during drilling operations in the petroleum industry. Here the decision maker must choose the best alternative to complete a well in a trade off between cost and risks (Mondal and Chatterjee, 2019). Geophysical data from neighboring wells can be valuable to infer the uncertainties in the subsurface pore pressure (Paglia et al., 2019), which is a key parameter in the quantification of risks. Section 6 has closing remarks on the methodological contributions presented here, including viable opportunities future work.

2 Spatial design of experiments

2.1 Spatial survey designs

We consider a situation as illustrated in Figure 1, with a spatial phenomenon allocated to a two-dimensional domain divided in grid cells or sites. The approach presented in the paper can be extended to other dimensions with minor changes. The spatial variables of interest are represented at n sites, denoted $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$

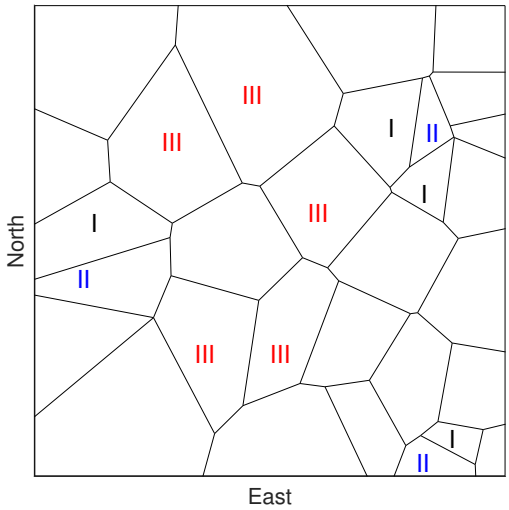


Figure 1: Illustration of a spatial domain split in 40 regional units of varying size. Three different designs are indicated (Design I, II and III) of different cardinality and spatial allocation.

with $\boldsymbol{s}_i = (\text{north}_i, \text{east}_i)$, $i = 1, \ldots, n$. In our applications, these sites have a particular interest to the decision maker. For instance, in the forestry example, the governmental institute must choose at each of the n sites

whether this forest unit should be harvested or left for conservation. Because there is much at stake and uncertain outcomes, the decision maker is likely to benefit from doing surveys at (a subset of) the sites.

Without lack of generality, the possible survey design locations are assumed to be identical to the n sites in our description, and a design is defined as any subset of these n sites (other cases can be constructed similarly, see e.g. Section 5.2). The possible spatial designs then include no points, single points, couples, triplets, and so on, up to all n points in the design. We denote these by $\mathscr{D} = \bigcup_{i=0}^{n} \mathscr{D}_i$, defined by;

$$
\begin{aligned}
&\mathscr{D}_0 = \emptyset, && \text{no sites in design,}\\
&\mathscr{D}_1 = \{(\boldsymbol{s}_1), (\boldsymbol{s}_2), \ldots, (\boldsymbol{s}_n)\}, && \text{one site in design,}\\
&\mathscr{D}_2 = \{(\boldsymbol{s}_1, \boldsymbol{s}_2), (\boldsymbol{s}_1, \boldsymbol{s}_3), \ldots, (\boldsymbol{s}_{n-1}, \boldsymbol{s}_n)\}, && \text{two sites in design,}\\
&\quad\vdots && \quad\vdots\\
&\mathscr{D}_n = \{(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)\}, && \text{all sites in design.}
\end{aligned}
$$

There are n possible designs of cardinality one, $\binom{n}{2}$ possible designs of cardinality two, etc. This means that there are 2^n possible designs in $\mathscr{D}$. We will further denote a general design by $D \in \mathscr{D}$ and its cardinality by $|D|$. The sites in this design are then $\boldsymbol{s}_{D,1}, \ldots, \boldsymbol{s}_{D,|D|}$. The number of sites shared by designs C and D is $|C \cap D|$, while the number of sites in at least one of the designs is $|C \cup D|$.

In our setting we compare the information gain obtained by different designs, and it makes sense that similar spatial designs contain almost the same information. In Figure 1 three different designs are indicated (I, II, III). Design I and II appear very similar in the spatial allocation of survey locations even though they have different cardinalities (three and four). Most likely, Design I will not have much to offer over Design II, unless there is much noise in the data or large gain in capturing additional covariate information which could be important for predictive purposes. Say, in the forestry example, a biologist would spend time doing one more experiments in Design I, at an extra cost. But unless she learns substantially more about the model, there is not much additional spatial information in Design I compared with doing just the three measurements in Design II. The last survey plan, Design III, is spatially very different from the two others because it allocates the measurements in the central parts of the domain. The value of this design could be very different from that of Design I and II.

To find the optimal design one must evaluate the information gain and cost for all possible design sets, but in practice one can only evaluate it for a fraction of all possible designs. We suggest a statistical approach for this optimisation problem, where we utilize the similarity of spatial designs to estimate the information gain.

2.2 Value of information

The goal of spatial design is to choose a survey plan for information gathering. This choice must balance expected information gain with the cost of data acquisition and processing. To evaluate the expected information gain associated with designs, one must formulate a value or utility function. Valuable designs tend to increase the expected utility substantially, while poorly selected designs provide hardly any additional utility over what is available with the current information. In the applications that we consider here, it is relatively straightforward to relate the question about information gain to an underlying decision situation, meaning that data are only valuable when their outcome can materialize in different decisions. For instance, in the forestry example the underlying decision is to conserve forest units or not, and data can help the decision maker to decide one or the other, depending what the information reveals. Managers are further often willing to phrase these decision situation in terms of monetary units, and then the VOI which gives the expected gain in information is directly comparable to the cost of data gathering. If the VOI exceeds this cost, the experiment is worthwhile and the decision maker should commit to gather the information, if the budget permits the cost. We next define the VOI formally through a model for the uncertainties involved, the decision alternatives and the information gathered by a chosen design.

The uncertain variables of interest are denoted by $\boldsymbol{x} = (x_1, \ldots, x_n)$, where $x_i = x(\boldsymbol{s}_i)$, $i = 1, \ldots, n$. Assuming a continuous sample space for this variable, we denote its density function by $p(\boldsymbol{x})$, with marginal density $p(x_i)$ for each sites $\boldsymbol{s}_i$. The decision alternatives are generally denoted by $\boldsymbol{a} \in \mathcal{A}$, where $\mathcal{A}$ is the

set of all possible alternatives. In some situations, the alternatives decouple (Eidsvik et al., 2015), involving for instance local decisions about harvesting units in our forest conservation example. In general, the prior value (PV), without any additional information, is defined as the value from doing the optimal decisions. Assuming a risk-neutral decision maker (Abbas and Howard, 2015), the PV is calculated from expected values as follows;

$$\text{PV} = \max_{\boldsymbol{a}\in\mathcal{A}} \{\mathbb{E}(\nu(\boldsymbol{x},\boldsymbol{a}))\}, \quad \mathbb{E}(\nu(\boldsymbol{x},\boldsymbol{a})) = \int \nu(\boldsymbol{x},\boldsymbol{a})p(\boldsymbol{x})d\boldsymbol{x}. \tag{1}$$

Here, $\nu(\boldsymbol{x},\boldsymbol{a})$ represents the value function, which could be quite general, but in our application it is the monetary profits associated with choice $\boldsymbol{a} \in \mathcal{A}$ when the uncertain outcome is $\boldsymbol{x}$. In the forestry example, the decision maker will choose to conserve the units that have high preservation value, while the others are harvested.

It is difficult to make decisions under uncertainty, and one can choose to purchase information that facilitate decision making. We here let $\boldsymbol{y}_D$ denote the data gathered by design $D \in \mathscr{D}$. This data is relevant to the decision situation in the sense that it is indicative of the underlying uncertainty $\boldsymbol{x}$. In the applications below, the model for data is given as a conditional probability density or mass function $p(\boldsymbol{y}_D|\boldsymbol{x})$, and the marginal model for data is then $p(\boldsymbol{y}_D) = \int p(\boldsymbol{y}_D|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$.

When the data are available, the conditional value (CV) is

$$\text{CV}(\boldsymbol{y}_D) = \max_{\boldsymbol{a}\in\mathcal{A}} \{\mathbb{E}(\nu(\boldsymbol{x},\boldsymbol{a})|\boldsymbol{y}_D)\}, \tag{2}$$

and the expected posterior value (PoV) before the data gathering is obtained by taking the expectation of expression (2) over the possible data outcomes:

$$\text{PoV}(D) = \mathbb{E}_{\boldsymbol{y}_D}\left[CV(\boldsymbol{y}_D)\right] = \mathbb{E}_{\boldsymbol{y}_D}\left[\max_{\boldsymbol{a}\in\mathcal{A}} \{\mathbb{E}(\nu(\boldsymbol{x},\boldsymbol{a})|\boldsymbol{y}_D)\}\right]. \tag{3}$$

The VOI is defined as the difference between the expected PoV in (3) and the PV in (1):

$$\text{VOI}(D) = \text{PoV}(D) - \text{PV}. \tag{4}$$

The goal is to choose a valuable design D. Keeping in mind that data comes with a cost we should compare the VOI with the cost $C(D)$ of design D. This means that the objective is to optimise

$$D^* = \text{argmax}_{D\in\mathscr{D}} I(D), \quad I(D) = \text{VOI}(D) - C(D). \tag{5}$$

and this is what we use in our applications, but other approaches are possible. For instance, a decision maker might have a fixed budget for the design, and the goal would then be to maximize the VOI among all designs that have a cost less than the budget.

With large opportunities for data gathering, it is extremely difficult to find the optimal design. First, the complexity grows extremely fast with the number of sites. Second, in common settings, the calculation of the information design criterion in (5) for a fixed design typically requires quite a bit of computational effort as is emphasized by the complexity of the integral expectation expressions required in (3). In practice one must often turn to heuristic approaches to such design problems (García-Ródenas et al., 2020). Instead of computing $I(D)$ exactly, we suggest to use a statistical approximation strategy that evaluates (5) only for a few promising designs which are extracted by in a much faster Bayesian optimisation approach building on Gaussian processes and expected improvement.

3 Bayesian optimisation for designs

We develop a Bayesian optimisation approach to guide the search for the maximum of $I(D)$ in (5). We combine computational search algorithms with the EI acquisition criterion to select which designs to evaluate in an iterative optimisation workflow. In doing so, we suggest to model the information measure $I(D)$ using a GP. This is in line with common approaches for Bayesian optimisation (Brochu et al., 2010; Frazier, 2018). The benefits of using a GP emulator for the information measure is that it enables:

- efficient model updating based on evaluations (Section 3.1),
- learning across different but similar designs (Section 3.2),
- computing EI in closed form, to focus on evaluating promising designs (Section 3.3),
- framing a useful algorithmic description of the overall procedure (Section 3.4).

3.1 Gaussian process emulator

A GP representation for the information gain I relies on mean and variance-covariance specifications for all possible designs. In the current setting with Bayesian optimisation, the GP model parameters are updated sequentially when more evaluations become available.

The representation then requires an initial specification of the mean μ and the variance σ^2. Before any data are observed, they are usually assumed constant for all designs. The GP model further needs a correlation function $K(C, D)$ between two different designs C and D to be specified (see Section 3.2). An starting batch of evaluations is used to find the initial maximum likelihood estimates for the model parameters involved in the mean and variance-covariance.

When m designs $D_{(1)}, \ldots, D_{(m)}$ have been evaluated, the knowledge is denoted $\mathcal{F} = \left\{(I_{(j)}, D_{(j)}); j = 1, \ldots, m\right\}$. By standard multivariate Gaussian theory, the conditional distribution for the information measure at design D is then Gaussian with mean and variance

$$\begin{aligned} \mu(D; \mathcal{F}) &= \mu + \boldsymbol{k}_{D,\mathcal{F}}^t \boldsymbol{K}_{\mathcal{F}}^{-1} (\boldsymbol{I}(\mathcal{F}) - \mu \mathbf{1}), \\ \sigma^2(D; \mathcal{F}) &= \sigma^2 \left(1 - \boldsymbol{k}_{D,\mathcal{F}}^t \boldsymbol{K}_{\mathcal{F}}^{-1} \boldsymbol{k}_{D,\mathcal{F}}\right). \end{aligned} \tag{6}$$

Here, $\boldsymbol{I}(\mathcal{F}) = (I_{(1)}, \ldots, I_{(m)})$ is the length m vector of information gain evaluations, $\boldsymbol{K}_{\mathcal{F}}$ the $m \times m$ correlation matrix between evaluations of designs, $\boldsymbol{k}_{D,\mathcal{F}}$ the length m vector of correlations between the evaluations and the information gain for design D, and $\mathbf{1}$ is a length m vector of 1 entries.

3.2 Distance between designs

The correlation function that gauges the similarity between designs, as defined via $\boldsymbol{k}_{D,\mathcal{F}}$ and $\boldsymbol{K}_{\mathcal{F}}$ in (6). This specification of a correlation function is a common task in spatial statistics and Bayesian optimisation over a regular input space. In our setting with spatial designs, it is not obvious how to assign this correlation function, and a main contribution of this paper is to formulate a distance measure between designs which is useful in the context of Bayesian optimisation. Our proposed distance measure for this task is the Hausdorff distance which is presented next, but we also outline other distance measures below to discuss this topic in a more general context. Throughout this description, we consider two general designs $D = (\boldsymbol{s}_{D,1}, \ldots, \boldsymbol{s}_{D,|D|})$ and $C = (\boldsymbol{s}_{C,1}, \ldots, \boldsymbol{s}_{C,|C|})$. For two sites $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, we let $\|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ be the Euclidean distance between the two points.

The Hausdorff distance is commonly used to measure the distance between curves, images or point sets (Huttenlocher et al., 1992). In our context it represents the maximum of the minimal distances from points in one set to points in the other set, and it hence measures similarity of designs:

$$h = \mathrm{dist}_1(D, C) = \max\left\{h_H(D, C), h_H(C, D)\right\}, \tag{7}$$

$$h_H(D, C) = \max_{i=1:|D|} \left\{ \min_{j=1:|C|} \|\boldsymbol{s}_{D,i} - \boldsymbol{s}_{C,j}\| \right\}. \tag{8}$$

Figure 2 illustrates several designs of size 1, ... 4. For each subplot the maximum distance from points in one point set (D, marked as circle) to the other (C, marked as cross) is calculated and shown. The Hausdorff distance in (7) is printed in the displays, and $h_H(D, C)$ and $h_H(C, D)$ are indicated. We note that in some cases the maximum distances from one set to the other are identical (upper right display and bottom middle display), but for most of these design configurations this symmetry is not present. For instance, in the upper left display, the circle is relatively close to the lowermost point in the cross set, but the highest point in the cross points is quite far from the circle. Similarly, in the center display, both points in the circles set are

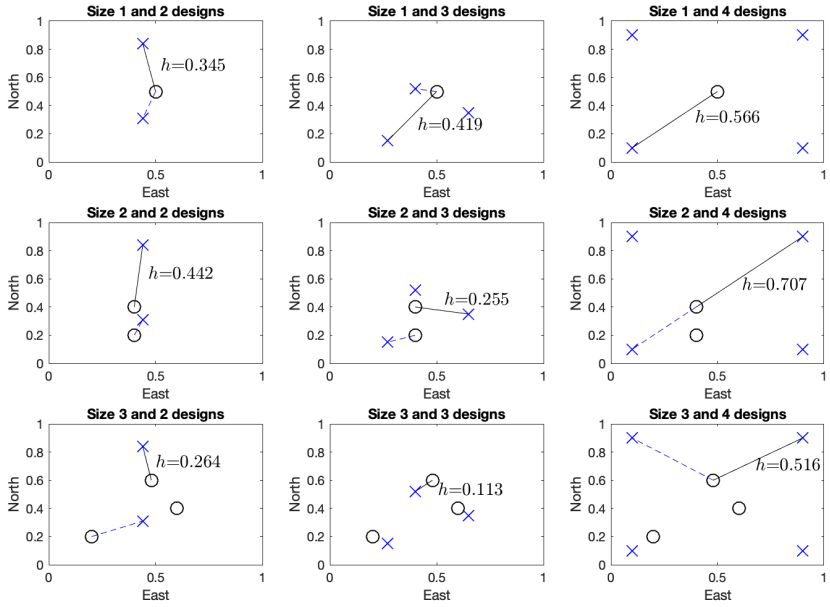


Figure 2: Hausdorff distance between various designs, h is the Hausdorff distance between the two sets, marked as circle for D and cross for C. The solid lines represent the maximum of the minimal distances between C and D, while the dashed line represents the maximum of the minimal distances between D and C.

close to a point in the cross set, but one point in the cross set is far from the closest point in the circle set. Based on Hausdorff distances for point sets like that displayed in Figure 2, this seems to be a useful way to measure the difference between designs. In the bottom-middle display the point sets are rather similar, and the distance is small ($h = 0.113$). In all the right displays, the designs are very different, and the distances are large.

We next present some alternative distances. Min et al. (2007) and Fujita (2013) propose some distances that could potentially be used for our purpose. We explore these and compare their pros and cons.

The second distance we study is defined by the minimum of the distances between designs of the two sets:

$$\text{dist}_2(D, C) = \min \left\{ \|\boldsymbol{s}_{D,i} - \boldsymbol{s}_{C,j}\|, \boldsymbol{s}_{D,i} \in D, \boldsymbol{s}_{C,j} \in C \right\}. \tag{9}$$

It is not a proper metric since the triangular inequality does not hold. The main problem though is that designs with elements in common are not separated because the distance in this case will be zero. We will hence discard then this distance – it is not suitable for our purpose.

We similarly define another distance considering the maximum of the distances between designs instead

$$\text{dist}_3(D, C) = \max \left\{ \|\boldsymbol{s}_{D,i} - \boldsymbol{s}_{C,j}\|, \boldsymbol{s}_{D,i} \in D, \boldsymbol{s}_{C,j} \in C \right\}. \tag{10}$$

That is not a proper distance either. Here the problem is that it does not satisfy the axiom $\text{dist}(D, C) = 0 \iff D = C$. It is hence not convenient for us to use it as a measure of similarity because the distance between two equal sets is greater than 0.

An alternative distance could be the Jaccard distance (Levandowsky and Winter, 1971)

$$\text{dist}_4(D, C) = \frac{|D \cup C| - |D \cap C|}{|D \cup C|}. \tag{11}$$

It defines a proper metric, and a sensible way to consider dissimilarity between designs. The problem for our type of applications is that it does not take into account the spatial distance of the sites in the sets, but just count the number of elements in the sets. This distance will hence not be considered in the current study.

Fujita (2013) describes another metric based on the average distances between the two designs.

$$\text{dist}_5(D,C) = \frac{1}{|D \cup C||D|} \sum_{\boldsymbol{s}_{D,i} \in D} \sum_{\boldsymbol{s}_{C,j} \in C \setminus D} \|\boldsymbol{s}_{D,i} - \boldsymbol{s}_{C,j}\| + \frac{1}{|D \cup C||C|} \sum_{\boldsymbol{s}_{D,i} \in D \setminus C} \sum_{\boldsymbol{s}_{C,j} \in C} \|\boldsymbol{s}_{D,i} - \boldsymbol{s}_{C,j}\|. \quad (12)$$

This metric seems to work sensibly for our purpose. It is a bit more difficult to interpret, but projecting the distances in a lower dimensional space might help. We will study the possibility of using dist_5 in Section 4.

Finally we look at some modified Hausdorff distances. A variant is defined as the average of minimum distances

$$\text{dist}_6(D,C) = \max\{h_{mH}(D,C), h_{mH}(C,D)\}, \quad (13)$$

$$h_{mH}(D,C) = \frac{1}{N} \sum_{i=1}^{|D|} \left\{ \min_{j=1:|C|} ||\boldsymbol{s}_{D,i} - \boldsymbol{s}_{C,j}|| \right\}.$$

Dubuisson and Jain (1994) show that the modified Hausdorff distance (expression (13)) is a valid tool for object matching. The problem with this measure is that it not a metric since the triangular inequality does not hold. Moreover dist_6 smooths the effect of outlier sites, whereas we believe that even a single outlier site could add valuable information to the design, giving knowledge of a larger area, and that should then have an important impact on the distance. Another possible variant of the Hausdorff distance is obtained taking the average of minimum squares, which again does not define a metric. We discard this distance for the same reasons that apply to dist_6.

In summary, we use the regular Hausdorff distance h in (7) to model design dissimilarities. This means that the correlation in the GP formulation (6) contains entries given by $K(D,C) = \exp\left(-\frac{h(D,C)^2}{2\theta_H^2}\right)$, which means a squared exponential correlation function with the Hausdorff distance $h(D,C)$ as input. For the expressions in (6) we must form this between all designs in the current evaluations $\mathcal{F}$ and for every design $D \in \mathscr{D}$ we want to predict. In a situation with spatial covariates $\boldsymbol{z}(\boldsymbol{s}_i) = \boldsymbol{z}_i$, $i = 1, \ldots, n$, we modify the expression to include the distance between covariates such that the correlation is

$$K(D,C) = \exp\left(-\frac{h(D,C)^2}{2\theta_H^2} - \frac{h(\boldsymbol{Z}_D, \boldsymbol{Z}_C)^2}{2\theta_Z^2}\right). \quad (14)$$

Here, $\boldsymbol{Z}_D = (\boldsymbol{z}_{D,1}, \ldots, \boldsymbol{z}_{D,|D|})$ contain the covariates in design D, and similarly for design C.

3.3 Expected improvement

The number of possible designs is huge, and in most situations the evaluation of $I(D)$ requires substantial computational resources involving an integral over the potential data outcomes which is typically solved by sophisticated analytical or numerical approximation methods or Monte Carlo sampling. In most practical applications it is hence not feasible to evaluate $I(D)$ for all designs D. We use EI as an acquisition function (Frazier, 2018) to guide the evaluation of designs.

The acquisition function uses the updated distribution for $I(D)$, given $\mathcal{F}$. Assuming m_0 starting evaluations, after t iterations with m evaluations of I each time, the EI is defined by

$$\text{EI}(D;\mathcal{F}) = \mathbb{E}\left(I(D) - I^+ | \mathcal{F}\right), \quad I^+ = \max\left\{I_{D_{(1)}}, \ldots, I_{D_{(m_0+mt)}}\right\}. \quad (15)$$

In the case of a GP model for $I(D)$, there is a closed form solution for EI, see e.g. Brochu et al. (2010). We have

$$\text{EI}(D;\mathcal{F}) = \left(\mu(D;\mathcal{F}) - I^+\right)\Phi(z) + \sigma(D;\mathcal{F})\phi(z), \quad z = \frac{\mu(D;\mathcal{F}) - I^+}{\sigma(D;\mathcal{F})}, \quad (16)$$

where Φ and ϕ are respectively the cdf and pdf of a standard Gaussian distribution; $\mu(D;\mathcal{F})$ and $\sigma^2(D;\mathcal{F})$ are the conditional mean and variance defined in (6).

By having the GP emulator, and accepting that EI is a useful acquisition function, the problem of maximizing I is now transformed to the problem of maximizing EI. The EI is relatively fast to compute for several designs, and the ones with large EI are selected for further evaluation.

3.4 Algorithm

The iterative algorithm is summarised in Algorithm 1 where we describe the methodology in pseudocode. Via the iterative procedure, the current maximum in information gain will not decrease and eventually reach the global maximum and return the optimal design. In practice, the algorithm terminates when the maximum value for information gain has not increased over a trailing buffer of iterations (ΔI^+) or if a maximum number of iterations (T_{max}) is reached.

The iterative scheme is initiated with a starting batch of m_0 evaluations, while each subsequent batch is of size m. For the initial batch, the designs are randomly selected from all possible designs $\mathscr{D}$. At each iteration, we augment the $\mathcal{F}$ set with a new evaluation, and then update the Hausdorff distances and the GP model, the current best design D^+ and the associated information gain I^+. To formalize this procedure we introduce $\mathcal{F}_t$ to denote all design evaluations done at iteration t, while $D_{t,(1)}, \ldots, D_{t,(m)}$ and $I_{t,(1)}, \ldots, I_{t,(m)}$ denote the design and information gain evaluations at iteration t.

At each iteration, the EI is computed for M designs while only the m ($m << M$) designs with largest EI, given the current evaluations, are selected for the batch $I(D)$ evaluation. The proposed new set of designs is obtained using a technique not dissimilar from a classical genetic algorithm (Goldberg, 1989), where the proposal set are mixed to create new combinations, while allowing random components to enter a design. In this step a large part of the proposed designs come from the set of all possible designs, while the remaining parts comes from a set obtained by mixing the best sites of previous steps. When we process a new design we check if the information gain has already been evaluated for that design, and if so we replace it with a new one. In this way we do not re-compute designs that have already been evaluated. In selecting new designs we adopt a weighted random selection of the design size to cover all cardinalities.

Algorithm 1: Search for designs by Bayesian optimisation.

Result: Design D^+ with the largest information gain I^+.
Iteration $t = 0$;
$\Delta I^+ = 1$;
Evaluate I for m_0 randomly selected designs $D_{t,(1)}, \ldots, D_{t,(m_0)}$ to get $I_{t,(1)}, \ldots, I_{t,(m_0)}$;
$I^+ = \max\{I_{t,(1)}, \ldots, I_{t,(m_0)}\}$;
$\mathcal{F}_t = \{(I_{t,(j)}, D_{t,(j)}); j = 1, \ldots, m_0\}$;
while $t \leq T_{max}$ *or* $\Delta I^+ = 0$ **do**
  $t = t + 1$;
  Mix existing design sites and random sites to suggest M designs ;
  Compute the Hausdorff distances for the suggested and available designs ; ▷ `expression` (7)
  Fit a GP model for I given all evaluations ; ▷ `expression` (6)
  Compute EI over I^+ for each of the M design ; ▷ `expression` (16)
  Find the m designs with largest EI to obtain $D_{t,(1)}, \ldots, D_{t,(m)}$;
  Evaluate $I(D_{t,(1)}), \ldots, I(D_{t,(m)})$; ▷ `expression` (5)
  $I^+ = \max\{I^+, I_{t,(1)}, \ldots, I_{t,(m)}\}$, $D^+ = \{D; I(D) = I^+\}$;
  $\mathcal{F}_t = \mathcal{F}_{t-1} \cup \{(I_{t,(1)}, D_{t,(1)})\} \cup \ldots \cup \{(I_{t,(m)}, D_{t,(m)})\}$;
  Compute an average increase over the last buffer of iterations ΔI^+;
end

4 Simulation study

We study the properties of the design algorithm in a situation with $n = 30$ spatial units of interest (Figure 1). The possible data gathering locations are $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$. At each of these locations there is an explanatory variable $z_i = z(\boldsymbol{s}_i)$ that classifies the units to one of four categories, largely inspired by the forestry application (Section 5.1) where ecologists know the age of the forest rather accurately.

The profits $x_i = x(\boldsymbol{s}_i)$, $x_i \in \mathbb{R}$, $i = 1, \ldots, n$, of the spatial units are the quantity of interest. The decision maker can choose, at each unit, not to take any action or to exploit that unit. We are hence in a situation where there is high decision flexibility and decoupled value, meaning that the decision maker is free to choose

the best alternative in a given location without accounting for the other parts. The set of alternatives is thus defined by $\mathcal{A} = \{a_i; i = 1, \ldots, n\}$, where $a_i = \{\text{not exploit}, \text{exploit}\} = \{0, 1\}$. The two-action value function is then

$$\nu(x_i, a_i) = \begin{cases} 0 & a_i = 0, \\ x_i & a_i = 1. \end{cases} \tag{17}$$

Data $\boldsymbol{y}_D$ can be gathered at any subset of $|D| \leq n$ units. These data will carry information at the design locations, and at other units through the spatial dependence and via learning the regression effect of the covariates. The optimal design size and configuration are chosen by the decision maker to maximize the VOI compared with the costs of data gathering.

For this simulation study we model the profits as Gaussian variables. The profits $\boldsymbol{x} = (x_1, \ldots, x_n)$ are represented by a hierarchical model with mean $\mathbb{E}(x_i|\boldsymbol{\beta}) = \beta_0 + \beta_1 z_i + \beta_2 z_i^2$, and the regression coefficients have a tri-variate Gaussian distribution $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$. Defining length n vectors $\boldsymbol{z} = (z_1, \ldots, z_n)$ and $\boldsymbol{z}^2 = (z_1^2, \ldots, z_n^2)$, the profits $\boldsymbol{x}$ are multivariate Gaussian with mean and covariance

$$\boldsymbol{\mu}_{\boldsymbol{x}} = \mu_{\beta_0} + \mu_{\beta_1}\mathbf{z} + \mu_{\beta_1}\mathbf{z}^2, \tag{18}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{\Sigma} + \begin{bmatrix} 1 & \mathbf{z} & \mathbf{z}^2 \end{bmatrix} \boldsymbol{\Sigma}_\beta \begin{bmatrix} 1 \\ \mathbf{z} \\ \mathbf{z}^2 \end{bmatrix}, \tag{19}$$

where matrix $\boldsymbol{\Sigma}$ holds the structural spatial variability with entries defined by a stationary variance term and a Matern correlation function: $\Sigma_{i,j} = \sigma_x^2 (1 + \eta \|\boldsymbol{s}_i - \boldsymbol{s}_j\|) \exp(-\eta \|\boldsymbol{s}_i - \boldsymbol{s}_j\|)$, where η is the spatial correlation decay parameter. Inspired by a forestry dataset, the parameters values are set to be $\boldsymbol{\mu}_\beta = (1.3 \cdot 10^4, -1.1 \cdot 10^4, 2.9 \cdot 10^3)'$, $\text{diag}(\boldsymbol{\Sigma}_\beta) = (6.4 \cdot 10^8, 5.4 \cdot 10^8, 0.2 \cdot 10^8)$, $\text{corr}(\beta_0, \beta_1) = -0.95$, $\text{corr}(\beta_1, \beta_2) = -0.98$, and $\text{corr}(\beta_0, \beta_2) = 0.89$, $\sigma_x^2 = 1.2 \cdot 10^8$ and $\eta = 0.3$. With the free selection of the various units, the PV in (1) hence involves a separate maximization for each unit, i.e. $PV = \sum_{i=1}^n \max\{0, \mu_{x_i}\}$.

We assume that data can be gathered and they are directly indicative of the profits, but measured with Gaussian additive noise. The conditional model for the data, given the profits is then defined by

$$\boldsymbol{y}_D = \boldsymbol{G}_D \boldsymbol{x} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \boldsymbol{T}_D), \tag{20}$$

where the size $|D| \times n$ matrix $\boldsymbol{G}_D$ picks the design locations by having one 1 entry in each row at the index of the sampled unit and otherwise 0 entries. Moreover, $N(0, \boldsymbol{T}_D)$ denotes a random Gaussian vector with zero-mean and covariance matrix $\boldsymbol{T}_D = \tau_D^2 \boldsymbol{I}_{|D|}$.

When the value function is in the form of (17) and the profits are Gaussian and measured with Gaussian additive noise, it is possible to compute the PoV in a closed form for each design (Bhattacharjya et al., 2013). The closed form calculation builds on the distribution of the conditional mean $\boldsymbol{\mu}_{x|y}$ with respect to the random data $\boldsymbol{y}$, which is Gaussian with $\mathbb{E}(\boldsymbol{\mu}_{x|y}) = \boldsymbol{\mu}_x$ and $\text{Var}(\boldsymbol{\mu}_{x|y_D}) = \boldsymbol{\Sigma}_x \boldsymbol{G}_D^t (\boldsymbol{G}_D \boldsymbol{\Sigma}_x \boldsymbol{G}_D^t + \boldsymbol{T}_D)^{-1} \boldsymbol{G}_D \boldsymbol{\Sigma}_x = \boldsymbol{R}$. The PoV in (3) is then

$$\text{PoV}(D) = \sum_{i=1}^n \mathbb{E}_{\boldsymbol{y}_D} \left[\max\{0, \mathbb{E}(\nu(x_i, a_i)|\boldsymbol{y}_D)\}\right] = \sum_{i=1}^n \left(\mu_{x_i} \Phi\left(\frac{\mu_{x_i}}{r_i}\right) + r_i \phi\left(\frac{\mu_{x_i}}{r_i}\right) \right), \tag{21}$$

where μ_{x_i} is element i in the mean vector and $r_i = \sqrt{R_{i,i}}$ is available from the ith diagonal entry of $\boldsymbol{R}$.

In the simulation study the costs increase with the size of the design, so that for designs with $|D| > 5$, the VOI never exceeds the cost. This means we can focus on all sites combinations up to size 5, and there are then around $1.7 \cdot 10^5$ possible designs. It is feasible to compute the exact VOI for all designs, and compare the optimal designs with the results obtained by Algorithm 1. For this purpose, we represent $I(D)$ by a GP where the covariance depends on the spatial distance between designs as well as their distance in covariate space, see (14). The Bayesian optimisation approach is run for 15 iterations and for a number of replicate re-starts. The results increase over iterations, and most replicates reach high information gain values. Each batch iteration consists of $m_0 = m = 50$ evaluations, so for each replicate there are $50 \cdot 16 = 800$ evaluations of $I(D)$.

We compare the results obtained by Algorithm 1 with that of other methods: (i) sequential selection algorithm, (ii) modified exchange algorithm (Mitchell, 1974; Royle, 2002), and (iii) using dist_5 (expression (12)) instead of the Hausdorff distance.

Method (i) sequentially chooses the best site in a forward selection: it first evaluates each of the single units and selects the one that maximizes information gain I. Next, it looks at all the couples that contain the selected unit, and finds which of these couples that has the largest value of information gain. It proceeds in this way until I does no longer increase. The computational cost of the sequential algorithm is relatively small as the total number of VOI evaluations is $\sum_{i=1}^{|D|+1}(|D|+1-i) = 140$. For our reference model, the sequential selection algorithm ends up with design $(1, 9, 23)$. This is quite far from the global maximum and also significantly below most of the replicate results achieved using the Bayesian optimisation approach.

Method (ii) iteratively exchanges, adds or removes random units to an existing design. When a random unit is added to the design, we have cardinality $|D| \to |D|+1$, while the cardinality $|D| \to |D|-1$ when a random unit is removed from the design. For a random exchange the cardinality remains the same. For each suggested design the information gain $I(D)$ is evaluated, and one keeps track of the best design so far. The solution paths of the exchange algorithm will change every time because of the random selection of moves. The exchange algorithm is often able to find the best design in less the 10^4 iterations, and could probably get there faster with some kind of weighted resampling. Still, it seems to require more evaluations than the Bayesian optimisation approach, and we believe it is difficult to tune this method for larger-size problems where the number of required evaluations will also increase dramatically.

Method (iii) using dist_5 has performance very similar to that of using the Hausdorff distance, but does not reach the optimum as often. The computational cost is about the same.

We compare the approaches using multidimensional scaling (Borg et al., 2018). In our context this helps us visualize the Hausdorff distances between designs in a 2 dimensional space that maintains the distances. Figure 3 shows the best 3000 designs (in grey) projected in this 2 dimensional space. The pink star represents the best design while the green diamond is the design obtained with the sequential selection method. In this display we indicate typical paths that Algorithm 1 (red) and the exchange algorithm (blue) take to get their final best results. It is interesting to observe the randomness of the path taken by the exchange algorithm to reach the maximum. We do not show all the sites explored to get to the maximum, only the locations of local maximum. Algorithm 1 reaches the maximum following a much more efficient path.

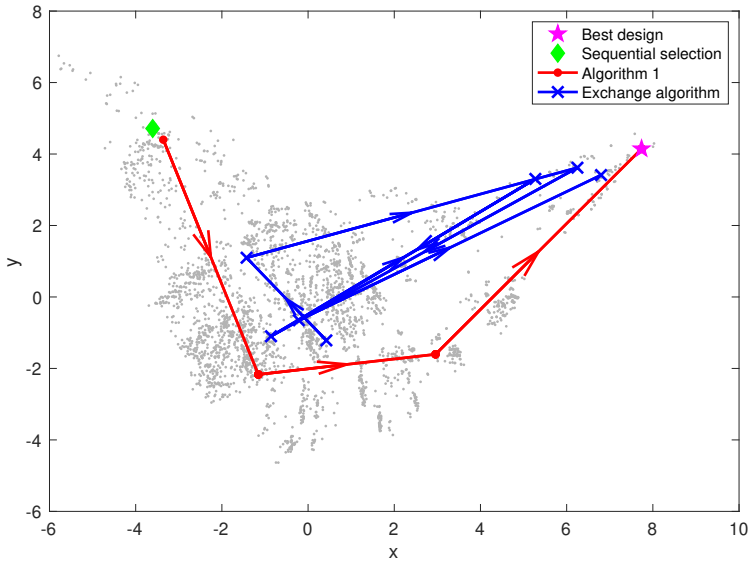
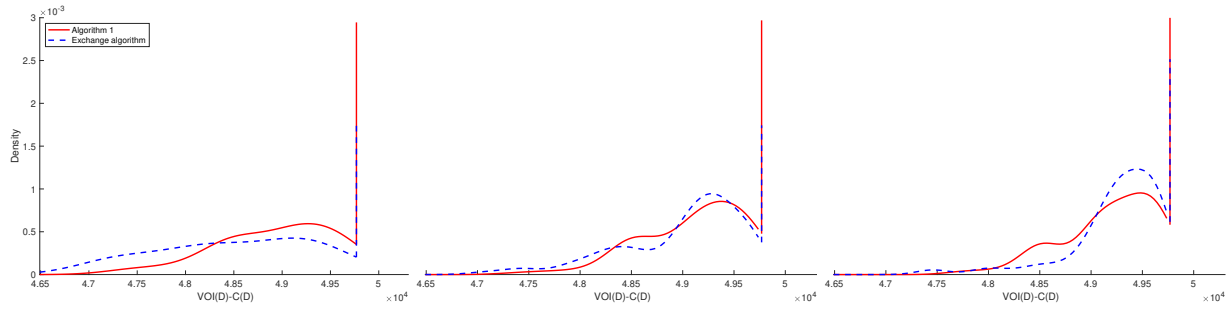


Figure 3: Representation of the spatial Hausdorff distance for the best 3000 designs (grey) in a 2 dimensional Euclidean space using multidimensional scaling technique. The pink star represents the best design and the green diamond the best design from the sequential selection. The red dots and lines represents the path of Algorithm 1 while the blue crosses and lines represents the results from the exchange algorithm.

For further comparison with the exchange algorithm we study the performances over 100 replicate restarts, and observe the maximum score after 250 (Figure 4(a)), 500 (Figure 4(b)), and 800 (Figure 4(c)) VOI evaluations. In Figure 4 we have in red the results of Algorithm 1 and in dashed blue the ones from the

exchange algorithm. We observe that the Bayesian optimisation method gets larger values of I after relatively few iterations because the exchange algorithm struggles with its random structure.



(a) 250 evaluations of I (in €). (b) 500 evaluations of I (in €). (c) 800 evaluations of I (in €).

Figure 4: Comparison of the performances of Algorithm 1 (red) and exchange algorithm (dashed blue) over 100 restarts. The Bayesian optimisation approach is able to get large values of I after few iterations, when the number of evaluations grows the exchange algorithm starts to perform well.

We perform sensitivity analysis to gain insight in the effect of having different input parameters. We focus on high or low values of the profit variance parameters σ^2 and $\mathbf{\Sigma}_\beta$, and on the algorithmic tuning parameter θ_i having separate or common spatial and covariate distance when computing the Hausdorff distance. We study algorithm performance metrics for each combinations of these input factors. Metrics include the highest value for I in the replicate runs, how many times we get a score among the best 100 I values over the replicate starts, and the performance of the sequential selection.

Table 1 shows the results of this sensitivity analysis, where the top line gives the results for the reference case. The difficult cases seem to be the ones with high prior variability, where the algorithm ends up with

σ^2	$\mathbf{\Sigma}_\beta$	θ	Algorithm 1	% best 100	Sequential selection
Low	Low	Separate	€23 008 (1)	100%	€22 692 (4)
High	Low	Separate	€81 665 (1)	100%	€80 315 (20)
Low	High	Separate	€54 137 (4)	100%	€54 137 (1)
High	High	Separate	€99 639 (33)	20%	€99 020 (172)
Low	Low	Together	€23 008 (1)	100%	€22 692 (4)
High	Low	Together	€81 163 (5)	100%	€80 315 (20)
Low	High	Together	€54 125 (12)	100%	€54 137 (1)
High	High	Together	€99 253 (82)	10%	€99 020 (172)

Table 1: Sensitivity analysis of the main parameters on the performance of the algorithm in the simulation study. The column "Algorithm 1" represents the highest I among the 10 replicates, the column "% best 100" represents how many times we get a score among the best 100 information gain values over the replicate starts; and the column "Sequential selection" highest I of the sequential selection algorithm.

moderate rankings and does not find the optimal solution in 15 iterations. The sequential algorithm also struggles more in these high-variability situations. When there are different levels of prior uncertainty in the spatial correlation and the regression trends, it becomes important to use a model with separate parameters in the kernel for the Hausdorff distance.

5 Examples

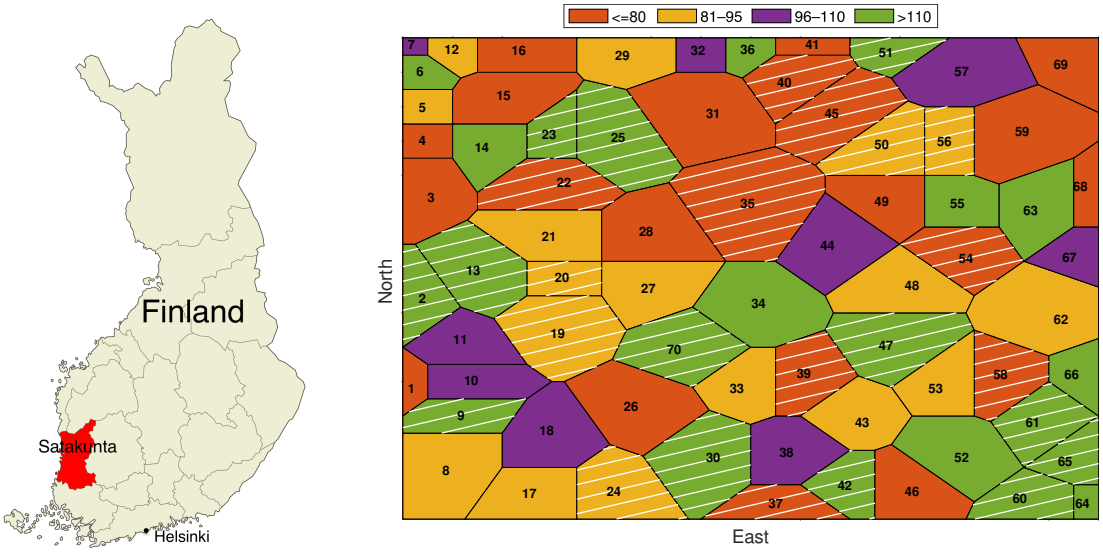
One application of the method is from forestry, where the decision is to choose which parts of a forest to harvest and which parts to spare for conservation. The other one is from petroleum, where the decision is

to choose whether to secure a well or just keep drilling.

5.1 Forestry

In this application the decision maker must choose to conserve forest stands or not. The decision maker is here a governmental institute that has a budget for conservation. The forest stands are owned by private owners who may harvest the timber unless the forest is conserved. In order to conserve a forest stand, the institute must pay a compensation (b_i) to the forest owner. When a forest stand is conserved, the ecological benefit (r_i) is proportional to a biodiversity indicator.

The study is inspired by data analysed by Eyvindson et al. (2019). The data consist of 70 forest stands (units) of various size from the Satakunta region in southwest Finland (Figure 5(a)). Each stand is classified according to the age class (1: $<= 80$, 2: $81-95$, 3: $96-110$, 4: > 110 years). Figure 5(b) shows the region with the forest stands, which are modified for various reasons, and colors identifying the different age groups. The



(a) The region of interest is located in the southwest of Finland. The red identify the geographic location of the Satakunta region (karttapohja Care, 2010).

(b) Forest stands of different size and age numbered from 1 to 70. Each color identify a different age group, orange: $<= 80$, yellow: $81-95$, violet: $96-110$, green: > 110 years. The hatched regions correspond to the best design.

Figure 5: Study area for forest conservation.

possible alternatives for the decision maker are $\mathcal{A} = \{a_i; i = 1, \ldots, n\}$, where $a_i = \{\text{not conserve}, \text{conserve}\} = \{0, 1\}$ at stand i (Eyvindson et al., 2019). We let x_i be the log-intensity of the number of wood inhabiting fungi at stand $i = 1, \ldots, n$. This number is a commonly used biodiversity indicator. The log-intensities are here modeled with a multivariate Gaussian distribution. The age of the forest stand is treated as a covariate $\boldsymbol{z}$ in the simulation study, so that the mean and the covariance matrix of $\boldsymbol{x}$ can be written as in the (18) and (19). The value function is then

$$\nu(x_i, a_i) = \begin{cases} 0 & a_i = 0, \\ re^{x_i} - b_i & a_i = 1. \end{cases} \tag{22}$$

Designs are constructed to gather information that can assist the decision maker. There are age-dependent inventory costs $C(\text{age}_i)$, so it is important to plan wisely and obtain effective designs at a low overall cost. The measurements of species richness in fungi are defined with a Poisson likelihood function

$$y_i | x_i \sim \text{Poisson}\left(e^{x_i}\right), \tag{23}$$

assuming conditional independence between the stands and constant area for each inventory.

The VOI is here defined by

$$\text{VOI}(D) = \sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{y}_D} \left[\max\{0, \mathbb{E}(re^{x_i} - b_i|\boldsymbol{y}_D)\}\right] - \sum_{i=1}^{n} \max\{0, \mathbb{E}(re^{x_i} - b_i)\}, \tag{24}$$

and the information gain $I(D)$ is obtained as the difference $\text{VOI}(D) - C(D)$ where $C(D)$ denotes the inventory costs. Evangelou and Eidsvik (2017) introduced a method to approximate the VOI in equation (24) based on iterative matrix approximations, linearisation, Gaussian approximations and the Laplace approximation. We use this approximation in this paper. But there is a total of $1.18 \cdot 10^{21}$ possible designs, and it is not feasible to calculate the VOI approximation for all of them to find the optimal design. Instead we use and compare various approximation methods to find efficient designs.

We initiate the Bayesian optimisation by evaluating $m_0 = 50$ random designs of various sizes. At each batch $m = 50$ new evaluations are selected using the EI acquisition function. The results from the Bayesian optimisation are shown in Figure 6. Even though we do not know the optimal solution in this case, the

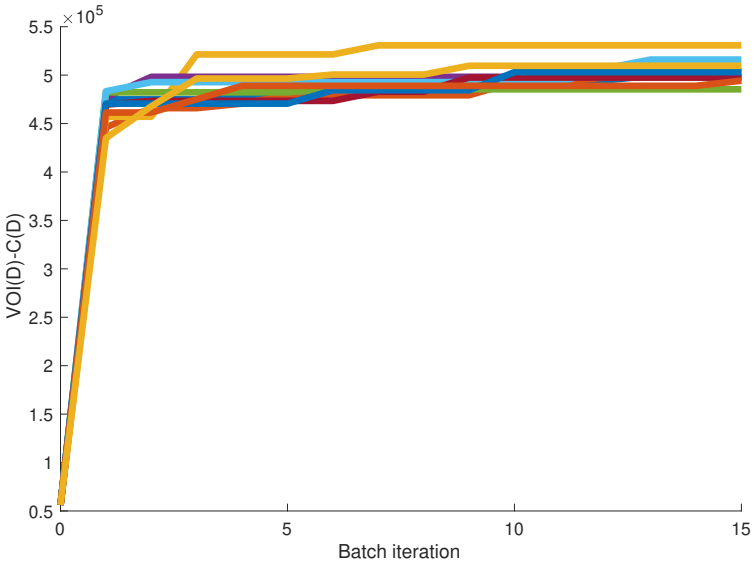


Figure 6: Performance of the Bayesian optimisation algorithm for the application in forest conservation. We do not know the maximum but we notice that the algorithm seems to converge over the 10 replicate restarts.

replicate restarts results shown in Figure 6 seems to indicate that the algorithm finds efficient designs within a few batch iterations. The information gain appears to converge similarly to what we noticed in the simulation example.

In Table 2 we list in descending order the 5 largest values of I obtained running the algorithm, together with the associated design. It is interesting to see how the designs selected have many stands in common, even though the designs are changing in size. This makes it possible to spot the stands that carry more information. The highest value of I corresponds to a rather large design set of $|D| = 26$. This design is illustrated using the hatched areas in Figure 5(b), where we observe that the stands of the design tend to spread and cover both the geographical region and also the various age levels.

Similar to what was done in the simulation study, we also run both the exchange algorithm and the sequential selection method. The exchange algorithm gets a largest replicate information gain of only $I =$€491 760 after 800 evaluations, and is not doing so well in this case. The sequential selection algorithm gives $I =$ €552 930 with 2485 evaluations. The associated design is $D = (1, 2, 3, 4, 5, 8, 15, 16, 19, 24, 26, 32, 40, 41, 44, 47, 49, 57, 59, 60, 66, 6$ In this example the sequential selection algorithm hence performs better than the iterative Bayesian optimisation, at a cost of extra VOI evaluations to find the sequential solution. With this in mind, we added the sequential solution to the evaluations of the Bayesian optimisation method, and continued to run that algorithm. We then achieved slightly larger information gain for designs very similar to the one detected

Design D	$I(D)$
2, 5, 8, 9, 17, 18, 19, 23, 26, 27, 28, 29, 33, 35, 40, 41, 42, 44, 47, 55, 56, 57, 59, 66, 69, 70	€530 740
3, 8, 9, 13, 15, 18, 20, 22, 25, 26, 27, 31, 32, 39, 40, 42, 46, 48, 50, 51, 54, 59, 62, 63, 65	€516 040
1, 2, 3, 8, 15, 18, 19, 22, 23, 25, 26, 27, 29, 31, 35, 39, 40, 41, 51, 55, 56, 59, 67, 70	€509 660
2, 3, 5, 8, 13, 14, 15, 17, 24, 25, 41, 44, 45, 46, 51, 53, 54, 55, 56, 65	€504 470
2, 6, 7, 8, 10, 16, 17, 19, 26, 27, 28, 29, 32, 34, 35, 40, 42, 45, 46, 48, 51, 52, 53, 54, 56, 58, 60, 62, 66, 68, 69	€503 080

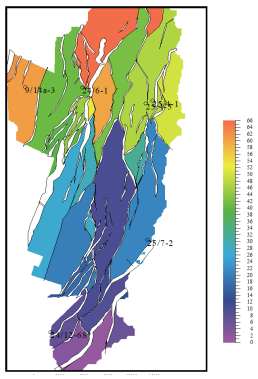
Table 2: The best 5 designs obtained over 10 re-run of the algorithm, listed in descending order.

with the sequential search, but no significant improvement. We hence suspect that the sequential method gives a near optimal solution for this example.

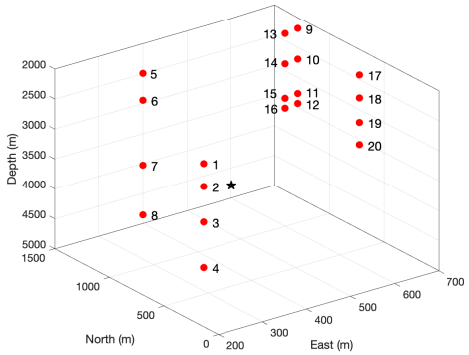
5.2 Petroleum drilling risks

This example concerns decision making during offshore drilling operations in the oil and gas industry (Lothe et al., 2019a). We study a drilling situation in the Alvheim oil field located in the central part of the North Sea, on the Norwegian continental shelf (Figure 7(a)). The field is divided in 68 compartments (Figure 7(b)) separated by faults. The circles in Figure 7(b) represent wells, and Figure 7(c) highlights these to indicate the decision location (black star) and the potential data gathering location (red dots) in a 3 dimensional plot. In the following we describe this decision situation and the opportunities for data gathering to make improved decisions.

(a) Geographical location of Alvheim. The rectangle indicates the position of the field.



(b) Map view of the oil field, circles indicate the locations of wells for data gathering. Different colors are used to identify different geological compartments (Lothe et al., 2019b).



(c) 3d view of the location of the measurements in red and the decision site in black star.

Figure 7: The study area is an offshore oil and gas field in the central part of the North Sea.

Drilling operations at the Alvheim field are characterised by the risk of overpressure, which occurs when the pore pressure in the rock exceeds the hydrostatic pressure. To prevent big hazards, the drilling mud pressure must be calibrated. We study a specific layer, located at about 3700 m depth, as marked in black in Figure 7(c). This layer is composed of mainly shale rocks and believed to be at drilling risk. The decision maker, which is the petroleum company in this case, must decide if it is safe enough to just keep drilling, or if they should set casing to strengthen the well because of a high risk of blowout. The

alternatives are $\mathcal{A} = \{0, 1\} = \{\text{keep drilling, set casing}\}$. To set casing is an expensive operation and it will reduce the borehole diameter, so the decision maker is interested in trying to postpone this operation, if not necessary because of very high risk. The value function of the decision is $\nu(x, a = 0) = -c_0(\text{LB}(x) - x)$ and $\nu(x, a = 1) = -c_1(\text{LB}(x) - x)$, where x is the unknown pore pressure variable and LB is the lower bound of the mud weight drilling window. Here, c_0 is the cost when one keeps drilling, while c_1 is the additional cost of casing. We note that the costs stretch the value functions so that it becomes more valuable to set casing instead of continued drilling for some values of pore pressure. Critically, the mud weight is used during the drilling of a well to exert a pressure on the borehole wall and avoid well collapse, and it is difficult to make decisions when the pore pressure is not known. Please keep in mind that there are a number of other parameters that would also affect LB, but pore pressure is an important parameter that is always taken into consideration during drilling operation (Moos et al., 2004).

Figure 7(c) (red) shows the possible measurement location. The design will entail any combination of these locations, and the measurements gathered at the design locations will be informative of the pore pressure where they are made and at other locations via the statistical model formulation. There are 5 wells where accurate measurements of pore pressure can be gathered in 4 different layers. The cost of data acquisition is assumed to be the same for each well. However it will be cheaper to obtain more information from the same well, since the tools for gathering the measurements have been already been placed into the well.

Based on seismic data from the region along with geological simulations of pressure build-up and release (Borge, 2000; Lothe, 2004; Paglia et al., 2019) we fit an initial multivariate Gaussian model for the pore pressure at the well location and at neighboring wells. The pore pressure measurements $\boldsymbol{y}_D$ in neighboring wells will then be indicative of the pore pressure in the target well via correlations. These modeling assumptions simplify the computation of the conditional expectation of pore pressure in the well of interest, given data obtained in the other wells. But the main challenge in the current setting is the expectation of the nonlinear value function and its integration over all possible data. The expectations required for the PV (expression (1)) and the PoV (expression (3)) are here calculated with numerical approximations of the integrals. This entails computing the value function ν over discretised levels of pore pressure. The LB is then obtained with a spline interpolation. This approximations then become

$$\text{PV} = \max_{a \in \mathcal{A}} \{\mathbb{E}\left(\nu(\boldsymbol{x}, a)\right)\} \approx \max_{a \in \mathcal{A}} \left\{ \sum_j \nu(x_j, a) p(x_j) \Delta x_j \right\}, \tag{25}$$

$$\text{PoV} = \mathbb{E}_{\boldsymbol{y}_D} \left[\max_{a_\in \mathcal{A}} \{\mathbb{E}(\nu(\boldsymbol{x}, a) | \boldsymbol{y}_D)\} \right] \approx \sum_{\boldsymbol{y}_a} \max_{a \in \mathcal{A}} \left\{ \sum_j \nu(x_j, a) p(x_j | \boldsymbol{y}_D) \Delta x_j \right\} p(\boldsymbol{y}_D) \Delta_{\boldsymbol{y}_D}, \tag{26}$$

where Δx_j and $\Delta_{\boldsymbol{y}_D}$ denote the distances between two consecutive discretised levels of x and $\boldsymbol{y}_D$ respectively.

Figure 8 shows the performance of Algorithm 1 for this application. As in the first example we do not know the design with the largest information gain value, but we observe a convergence of the method towards larger information gain as more batches are evaluated.

Design (D)	$I(D)$
5, 6, 8, 14	€9 115 100
6, 7, 8, 15	€9 115 100
5, 7, 8,11	€9 115 100
3, 5, 6, 7, 8, 19, 20	€9 114 900
3, 5, 6, 7, 8, 9, 10, 12	€9 114 600

Table 3: The largest 5 designs obtained over 10 restarts of the algorithm, listed in descending order.

Table 3 shows the five largest values of I in descending order, together with the associated design. Once one starts gathering data at a specific well, acquiring more data at other depth is relatively cheap. This implies that the cost effective designs suggest to explore more then one depth for a single well.

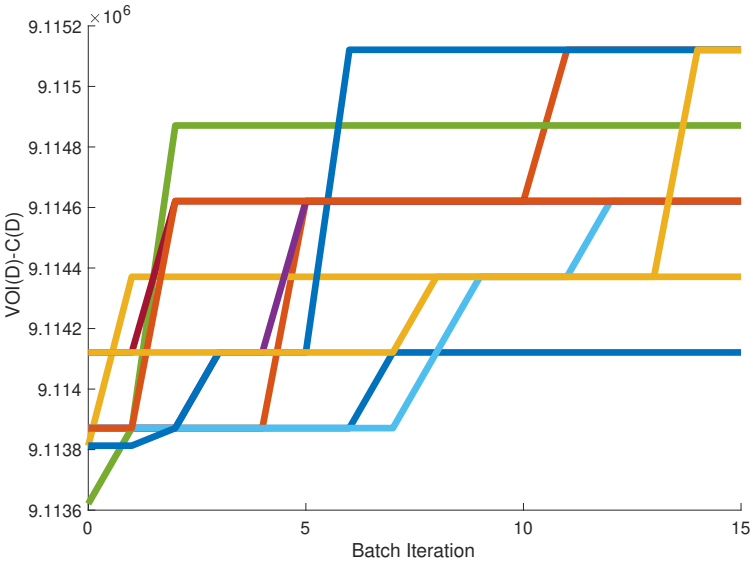


Figure 8: Performance of the Bayesian optimisation algorithm for an application in drilling of a well. Without knowing the point of maximum we can notice over the 10 restarts of the algorithm to converge to a large value of VOI-C.

In this case the sequential selection method gave $I =$€9 114 900 with 210 iterations, which is rather good but not as high as the best one achieved with the Bayesian optimisation method. The exchange algorithm obtained $I =$€9 114 600 with 800 iterations.

6 Closing remarks

The main purpose of this study is to develop an algorithm that can assist a decision maker in choosing a spatial design configuration for collecting information. The methodology has its applications in earth sciences, where data are often distributed over a spatial domain. We illustrated the approach by presenting one example from the forestry and one from petroleum.

We have adopted the Hausdorff distance to model dissimilarities between designs, and demonstrated its use in examples. We believe that, depending on the field where the methodology is applied, other metrics can work as well. The Jaccard distance described in (11) and the metrics introduced by Fujita (2013), see (12), can be valid alternatives to the Hausdorff distance. Say, the Jaccard distance could work in situations where we are not too interested in spatial distance between designs, but in a more machine learning oriented context where one must select the appropriate number of sets for training, and because data may come from different sources it is important to guide the active learning wisely (Settles, 2012). The developed methodology could be also applied in subset selection problems such as the selection of individuals in epidemiological follow-up studies (Reinikainen et al., 2016) or in genotyping (Karvanen et al., 2009).

It is possible to extend the study considering more challenging probability distributions for data, where the computation of VOI becomes more difficult. The combination of the VOI analysis with Bayesian optimisation techniques gives us an efficient way to find satisfactory data gathering scheme. With the Bayesian optimisation we move the problem of evaluating VOI to that of computing EI, which requires less computational effort. The total number of evaluation of VOI is considerably reduced. In situations where computing the information gain is computationally demanding or the number of alternatives to explore is too large, the developed methodology reduces the time of computation.

Acknowledgements

Jacopo Paglia's and Jo Eidsvik's work are supported by the KPN project 255418/E30: "Reduced uncertainty in overpressures and drilling window prediction ahead of the bit (PressureAhead)", of the Norwegian Research Council and the DrillWell Centre (AkerBP, Wintershall, ConocoPhillips and Equinor).
Juha Karvanen's work is supported by Grant number311877 "Decision analytics utilizing causal models and multiobjective optimisation" (DEMO), of the Academy of Finland.

References

Abbas, A. E. and Howard, R. A. (2015). *Foundations of decision analysis*. Pearson Higher Ed.

Bhattacharjya, D., Eidsvik, J., and Mukerji, T. (2013). The value of information in portfolio problems with dependent projects. *Decision Analysis*, 10(4):341–351.

Binois, M., Huang, J., Gramacy, R. B., and Ludkovski, M. (2019). Replication or exploration? Sequential design for stochastic simulation experiments. *Technometrics*, 61(1):7–23.

Borg, I., Groenen, P. J., and Mair, P. (2018). *Applied multidimensional scaling and unfolding*. Springer.

Borge, H. (2000). *Fault controlled pressure modelling in sedimentary basins*. PhD thesis, Norwegian University of Science and Technology.

Bouneffouf, D. (2016). Exponentiated gradient exploration for active learning. *Computers*, 5(1):1.

Brochu, E., Cora, V. M., and De Freitas, N. (2010). A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*.

Diggle, P. and Lophaven, S. (2006). Bayesian geostatistical design. *Scandinavian Journal of Statistics*, 33(1):53–64.

Dobbie, M. J., Henderson, B. L., Stevens Jr, D. L., et al. (2008). Sparse sampling: spatial design for monitoring stream networks. *Statistics Surveys*, 2:113–153.

Drovandi, C. C., Holmes, C., McGree, J. M., Mengersen, K., Richardson, S., and Ryan, E. G. (2017). Principles of experimental design for big data analysis. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 32(3):385.

Drovandi, C. C., McGree, J. M., and Pettitt, A. N. (2013). Sequential monte carlo for Bayesian sequentially designed experiments for discrete data. *Computational Statistics & Data Analysis*, 57(1):320–335.

Dubuisson, M.-P. and Jain, A. K. (1994). A modified Hausdorff distance for object matching. In *Proceedings of 12th international conference on pattern recognition*, volume 1, pages 566–568. IEEE.

Eidsvik, J., Martinelli, G., and Bhattacharjya, D. (2018). Sequential information gathering schemes for spatial risk and decision analysis applications. *Stochastic environmental research and risk assessment*, 32(4):1163–1177.

Eidsvik, J., Mukerji, T., and Bhattacharjya, D. (2015). *Value of information in the earth sciences: Integrating spatial modeling and decision analysis*. Cambridge University Press.

Evangelou, E. and Eidsvik, J. (2017). The value of information for correlated GLMs. *Journal of Statistical Planning and Inference*, 180:30–48.

Eyvindson, K., Hakanen, J., Mönkkönen, M., Juutinen, A., and Karvanen, J. (2019). Value of information in multiple criteria decision making: an application to forest conservation. *Stochastic Environmental Research and Risk Assessment*.

Eyvindson, K. J., Petty, A. D., and Kangas, A. S. (2017). Determining the appropriate timing of the next forest inventory: incorporating forest owner risk preferences and the uncertainty of forest data quality. *Annals of Forest Science*, 74(1):2.

Frazier, P. I. (2018). A tutorial on Bayesian optimization. *arXiv preprint arXiv:1807.02811*.

Fujita, O. (2013). Metrics based on average distance between sets. *Japan Journal of Industrial and Applied Mathematics*, 30(1):1–19.

García-Ródenas, R., García-García, J. C., López-Fidalgo, J., Ángel Martín-Baos, J., and Wong, W. K. (2020). A comparison of general-purpose optimization algorithms for finding optimal approximate experimental designs. *Computational Statistics & Data Analysis*, 144:106844.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition.

Huan, X. and Marzouk, Y. M. (2013). Simulation-based optimal Bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317.

Huttenlocher, D. P., Rucklidge, W. J., and Klanderman, G. A. (1992). Comparing images using the Hausdorff distance under translation. In *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 654–656. IEEE.

Kangas, A., Kangas, J., and Kurttila, M. (2008). *Decision support for forest management*, volume 16. Springer.

karttapohja Care (2010). Public domain. `https://commons.wikimedia.org/w/index.php?curid=8954495`. Accessed December 2019.

Karvanen, J., Kulathinal, S., and Gasbarra, D. (2009). Optimal designs to select individuals for genotyping conditional on observed binary or survival outcomes and non-genetic covariates. *Computational Statistics & Data Analysis*, 53(5):1782–1793.

Levandowsky, M. and Winter, D. (1971). Distance between sets. *Nature*, 234(5323):34–35.

Lothe, A. E. (2004). *Simulations of hydraulic fracturing and leakage in sedimentary basins*. PhD thesis, University of Bergen.

Lothe, A. E., Cerasi, P., Aghito, M., et al. (2019a). Digitized uncertainty handling of pore pressure and mud-weight window ahead of bit: North sea example. *SPE Journal*. `https://doi.org/10.2118/189665-PA`.

Lothe, A. E., Grøver, A., Roli, O. A., Leirdal, G., and Kristiansen, T. G. (2019b). Uncertainty modelling of minimum horizontal stresses and pore pressures in deeply buried grabens. what's next in modelling? In *81st EAGE Conference & Exhibition*. EAGE.

Min, D., Zhilin, L., and Xiaoyong, C. (2007). Extended Hausdorff distance for spatial objects in GIS. *International Journal of Geographical Information Science*, 21(4):459–475.

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

Mondal, S. and Chatterjee, R. (2019). Quantitative risk assessment for optimum mud weight window design: A case study. *Journal of Petroleum Science and Engineering*, 176:800–810.

Moos, D., Peska, P., Ward, C., Brehm, A., et al. (2004). Quantitative risk assessment applied to pre-drill pore pressure, sealing potential, and mud window predictions from seismic data. In *Gulf Rocks 2004, the 6th North America Rock Mechanics Symposium (NARMS)*. American Rock Mechanics Association.

Overstall, A. M. and Woods, D. C. (2017). Bayesian design of experiments using approximate coordinate exchange. *Technometrics*, 59(4):458–470.

Paglia, J., Eidsvik, J., Grøver, A., and Lothe, A. E. (2019). Statistical modeling for real-time pore pressure prediction from predrill analysis and well logs. *Geophysics*, 84(2):ID1–ID12.

Reinikainen, J., Karvanen, J., and Tolonen, H. (2016). Optimal selection of individuals for repeated covariate measurements in follow-up studies. *Statistical methods in medical research*, 25(6):2420–2433.

Royle, J. A. (2002). Exchange algorithms for constructing large spatial designs. *Journal of Statistical Planning and Inference*, 100(2):121–134.

Ryan, E. G., Drovandi, C. C., McGree, J. M., and Pettitt, A. N. (2016). A review of modern computational algorithms for Bayesian optimal design. *International Statistical Review*, 84(1):128–154.

Settles, B. (2012). Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1):1–114.

Weaver, B. P., Williams, B. J., Anderson-Cook, C. M., Higdon, D. M., et al. (2016). Computational enhancements to Bayesian design of experiments using Gaussian processes. *Bayesian Analysis*, 11(1):191–213.