Sequential information gathering schemes for spatial risk and decision analysis applications

Jo Eidsvik[1], **Gabriele Martinelli**[2] **and Debarun Bhattacharjya**[3]
1) Department of Mathematical Sciences, NTNU, Norway, (email: joeid@math.ntnu.no)
2) Thomson Reuters, Oslo, Norway
3) IBM T.J. Watson Research Center, New York, US

Corresponding address: Jo Eidsvik (jo.eidsvik@ntnu.no), Department of Mathematical Sciences, NTNU, 7491 Trondheim, NORWAY, phone: +4773590153 / fax: +4773593524

ABSTRACT

Several risk and decision analysis applications are characterized by spatial elements: there are spatially dependent uncertain variables of interest, decisions are made at spatial locations, and there are opportunities for spatial data acquisition. Spatial dependence implies that the data gathered at one coordinate could inform and assist a decision maker at other locations as well, and one should account for this learning effect when analyzing and comparing information gathering schemes. In this paper, we present concepts and methods for evaluating sequential information gathering schemes in spatial decision situations. Static and sequential information gathering schemes are outlined using the decision theoretic notion of value of information, and we use heuristics for approximating the value of sequential information in large-size spatial applications. We illustrate the concepts using a Bayesian network example motivated from risks associated with CO2 sequestration. We present a case study from mining where there are risks of rock hazard in the tunnels, and information about the spatial distribution of joints in the rocks may lead to a better allocation of resources for choosing rock reinforcement locations. In this application, the spatial variables are modeled by a Gaussian process. In both examples there can be large values associated with adaptive information gathering.

Keywords: value of information, spatial risk analysis, spatial statistics, sequential information, adaptive testing, Bayesian networks, Gaussian processes

1 Introduction

With the current trends in machine learning, artificial intelligence and internet of things, there is growing interest in monitoring. By monitoring, or information gathering, we refer to data collected by sensors, surveys or through processing of massive data sets such as by geophysical acquisition or remote sensing. The purpose of gathering such data is to make improved decisions under uncertainty. In several applications, the goal is to monitor a multivariate phenomenon in space or time, and decision makers must choose wisely where or when to gather information. In the current paper we focus on risk and decision analysis applications that involve spatial elements. We discuss one example with CO2 sequestration sites that could be leaking, and by gathering geophysical data one might get valuable information about the sealing properties of the reservoir(s). In another example involving risks of rock hazard, the decision maker must choose the locations for reinforced rock support in a mining tunnel, and borehole data of joint counts can help make these difficult decisions.

These examples, and several other application domains, seem to share at least three *spatial* elements. First, *decisions* are often allocated to geographic coordinates. For instance, the decision maker must select locations for CO2 sequestration, or choose whether to reinforce specific rock formations that could cause severe hazard in the mine. Second, *uncertainties* are characterized by spatial dependence. In terms of the rock hazard risk, the joint frequency in the rocks would be spatially correlated. Third, *information* can be gathered at specific locations. Depending on the application, it may be infeasible or just not worthwhile to acquire such information at all the locations in general. Rather, the budget can often only cover one survey to get an indication about the sealing properties for one of the sequestration sites, and survey data for joint frequency are typically only acquired in a few boreholes.

Oftentimes, information can be collected *sequentially* rather than all at once, i.e. the decision maker can, at any point, determine whether s/he wishes to continue gathering information or go ahead with her/his allocation decision (Miller, 1975). This sequential aspect gives the decision maker flexibility since the information set can grow in different ways. In spatial situations, such an approach may be of particular value to the decision maker because observations at certain locations also provide information about unexplored locations through spatial dependence.

In risk and decision analysis, the decision theoretic notion of *value of information* (Howard, 1966; Raiffa, 1968) provides a formal basis for evaluating information sources that support decision making. The value of information (henceforth referred to as VOI) has been popular in classic decision analysis industries such as oil and gas (Bratvold et al., 2009), but it has also been effectively deployed in domains with a risk analysis bent, including health risk management (Yokota and Thompson, 2004a; Yokota and Thompson, 2004b; Baio, 2012) and environmental applications (Keisler et al., 2014). Topics in environmental network design are similar: Wang and Harrison (2013) study optimal sampling locations for mobile sensors to detect water contaminants, without using a formal decision theoretic setting. Convertino et al. (2015) present monitoring networks for ecosystems, with a value measure tied to biodiversity.

There has been a relatively recent attempt at explicitly modeling spatial dependence for decision-theoretic VOI analysis (Eidsvik et al., 2008; Bhattacharjya et al., 2010; Eidsvik et al., 2015). In this paper, we build upon this framework and study the value of sequential information in spatial risk and decision applications. We are particularly interested in exploring the effect of spatial dependence. Note that while the information gathering is sequential, the underlying decision situation is assumed to be a one-time selection, i.e. the problem of resource allocation is itself static.

Sequential optimization problems are typically solved using dynamic programming (Bellman, 1957; Puterman, 2005), but the 'curse of dimensionality' that is encountered in larger problems requires computational techniques for approximate solutions (Powell, 2011). Some of these methods have been deployed for solving sequential decision problems in applications with dependence, including models such as graphical models or Bayesian networks (BNs) (Krause and Guestrin, 2009; Brown and Smith, 2013; Martinelli et al., 2013a), Markov random fields (Bonneau et al., 2014; Martinelli and Eidsvik, 2014) and Gaussian processes (GPs) (Srinivas et al., 2010). In the rock hazard application, we compare some common heuristics for sequential information gathering.

Although our focus is primarily on dependence among spatial random variables, sometimes called uncertainties in the decision analysis literature, there is some related literature on spatial decision making that studies multi-attribute preference models (Malczewski, 2006; Simon et al., 2014). The formulation presented in this paper can easily incorporate multiple attributes/criteria, as long as the decision maker can provide a value function (Matheson and Howard, 1968; Dyer and Sarin, 1979), which is necessary for VOI analysis. Here we will assume value to be in monetary units, which allows for a buying price interpretation of VOI (Howard and Abbas, 2015). This is often of great practical benefit because it allows the decision maker a direct comparison between the value of an experiment and its cost.

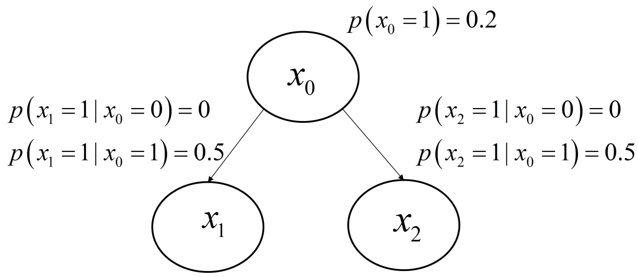
The remainder of the paper is organized as follows. Section 2 presents the notation and basic concepts for formulating decision situations when there is statistical dependence. Section 3 proceeds to VOI analysis, distinguishing between static and sequential information gathering. For gaining insight into the concepts, we use the illustrative running example with a BN model for leaking or sealing reservoir variables at two geographic locations chosen as candidates for CO2 sequestration. Section 4 introduces some heuristic strategies for sequential information gathering in high-dimensional spatial risk and decision analysis applications, and provides computational methods required to approximate the value of these strategies. Section 5 performs a case study of a rock hazard application in a mining tunnel. Section 6 concludes the paper.

2 Spatial Decision Situations

The inherent spatial dependence in many applications is often a consequence of the fundamental 'physics' of the problem. In health and epidemiology, this is a reflection of how human beings interact with each other and with their environment. In the environmental and earth sciences, spatial variability arises from trends caused by eons of geological, physical, chemical, and biological processes. The decision maker's characterization of uncertainties is then best represented by spatial statistics, because properties at a particular geographic location cannot be treated independently of those at other locations. In this section, we describe a formulation of spatial decision situations and present the basic terminology and notation that will be used in subsequent sections. We also introduce the specific models that will be used for a running example about CO2 sequestration as well as the rock hazard application in Section 5.

2.1 Spatial statistical modeling

Spatial and spatio-temporal statistical modeling techniques (Cressie, 1993; Le and Zidek, 2006; Cressie and Wikle, 2011; Chiles and Delfiner, 2012; Banerjee et al., 2014) aim at embedding vari-



	Trap 1 sealing	Trap 1 leaking	Marginal
Trap 2 sealing	0.85	0.05	0.9
Trap 2 leaking	0.05	0.05	0.1
Marginal	0.9	0.1	1

Figure 1: CO2 sequestration example: BN for joint distribution at two reservoir traps.

ables in a framework that allows for multivariate spatial dependence. There has been an increased focus in these fields over the last few years, perhaps due to the remarkable growth and development in positioning equipment as well as computer tools that allow one to register plans and current activities or monitor data sets on a geographical map.

We refer to the most pertinent uncertainties in a decision situation as the *distinctions of interest*. These are denoted $\mathbf{x} = (x_1, \ldots, x_n)$, where the random variables are associated with spatial locations $\mathbf{s}_i$, $i = 1, \ldots, n$, and thus $x_i = x(\mathbf{s}_i)$. The joint probability density or mass function (pdf) for these spatial variables is denoted by $p(\mathbf{x})$.

In the simplest non-trivial situation, there are just two dependent variables $\mathbf{x} = (x_1, x_2)$. Consider an example where x_i is a binary random variable that represents whether the trap of the geological formation of reservoir i will leak; $x_i = 1$ if it leaks and $x_i = 0$ if it is sealing. This setting is relevant for risks associated with CO2 sequestration (Mathieson et al., 2011). Figure 1 shows a BN model of the uncertain trap properties of two reservoirs along with a variable x_0, which is common parent node variable in the BN. It induces dependence in the outcomes for x_1 and x_2 through an underlying geological mechanism where trap leakage at the reservoirs will not occur unless it occurs at the top node (Martinelli et al., 2013b). This sort of dependence is natural in many spatial applications where geological conditions are such that properties cannot propagate unless they occur in an upstream variable (Martinelli et al., 2011). For our example, we will use the following marginal and conditional probabilities, which together construct the joint pdf: $p(x_0 = 1) = 0.2$, $p(x_i = 0|x_0 = 0) = 1$, $p(x_i = 1|x_0 = 1) = 0.5$, $i = 1, 2$, assuming the two trap properties are conditionally independent given the parent outcome and have symmetric properties. The bivariate pdf for the two traps, after summing out the common parent variable, is provided in Figure 1.

Perhaps the most pervasive and widely applicable model for representing spatial variables is the GP (Rasmussen and Williams, 2006; Banerjee et al., 2014), which uses a multivariate Gaussian distribution over spatial locations of interest with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Figure 2 (top) shows realizations of a GP in two spatial dimensions over the unit square. The realizations

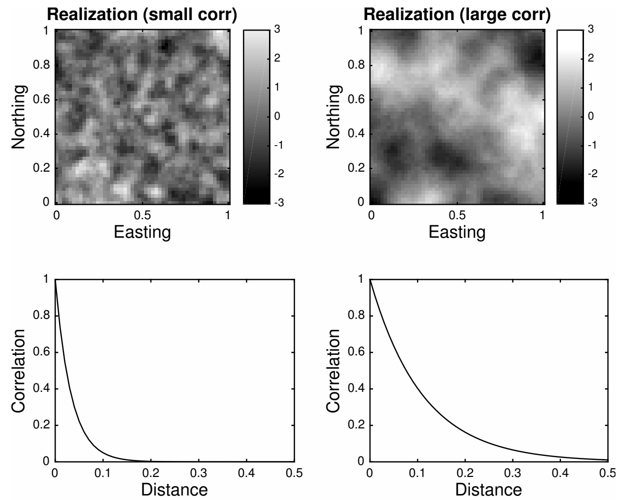


Figure 2: Gaussian process realizations (top) for two different correlation functions (bottom).

are representative of different spatial correlations (left and right). In these displays, the spatial correlation function (bottom displays) is an exponential function of the distance between locations, i.e. the isotropic covariance between variables x_i and x_j with distance $h = |\mathbf{s}_i - \mathbf{s}_j|$ is $\Sigma_{ij} = C(h) = \sigma^2 \exp(-\eta h)$ for parameters σ and η. In the figure, the mean $\mu_i = 0$ for all locations and the variance $\sigma^2 = 1$. The correlation lengths for the two cases are $3/\eta$ equal to 0.1 (left) and 0.33 (right). This is the distance at which the correlation between variables decreases to 0.05 and is therefore a measure of spatial dependence, where a higher value reflects larger correlation.

One of the major benefits of GP models is computational tractability. For instance, there are closed-form expressions for updating the model based on observations. If the outcome x_j of the GP at a site $\mathbf{s}_j$ is observed, the conditional distribution at another site $\mathbf{s}_i$ is Gaussian with mean

$$E(x_i|x_j) = \mu_i + \Sigma_{i,j}\sigma^{-2}(x_j - \mu_j), \tag{1}$$

while the conditional variance is

$$\text{Var}(x_i|x_j) = \sigma^2 - \Sigma_{i,j}\sigma^{-2}\Sigma_{j,i}. \tag{2}$$

These equations are used extensively in spatial interpolation (Kriging). If the two sites are in close proximity, then the cross-covariance $\Sigma_{i,j} \approx \sigma^2$, and the expected value in equation (1) is close to the observation x_j (if $\mu_i = \mu_j$). If the two sites are far away from each other, the cross-covariance is near 0 and the expected value is near μ_i. Similarly, the variance in equation (2) will be close to 0 for nearby sites and increase to σ^2 at large distances, where the observation x_j is not particularly informative of x_i.

The formulas in equations (1) and (2) have natural vector/matrix extensions in multi-site data and prediction, and we will use these for conditioning and for conducting VOI analysis in subsequent sections.

2.2 Spatial decisions

Now let us overlay a decision on the spatial statistical model. We denote alternatives of the decision by $\mathbf{a} = (a_1, \ldots, a_n)$, where $a_i = a(\mathbf{s}_i) \in A_i$ is a choice the decision maker has at spatial location $\mathbf{s}_i$. Note that the spatial scale of the alternatives need not necessarily be the same as that of the uncertain variables, but we assume that this is the case for simplicity of exposition. The decision maker's value from choosing alternative $\mathbf{a}$ when the distinctions of interest are at state $\mathbf{x}$ is $v(\mathbf{x}, \mathbf{a})$. We assume that the decision maker is risk neutral, so decisions are based on expected values. The decision maker hence has the common linear utility function over value. The methods described here for VOI analysis also work for other utility functions, but if a function other than linear or exponential is chosen, then computations must be performed from first principles (Howard and Abbas, 2015) which can be more challenging in general.

The decision is *static* when the decision maker must choose in one go, i.e. without the opportunity of potentially observing outcomes from alternatives at selected locations before making choices at other locations. If the decision situation does not have further constraints, then for a risk neutral decision maker the alternative that maximizes the expected value is optimal, therefore

$$\mathrm{PV} = \max_{\mathbf{a} \in \mathbf{A}} \left\{ \sum_{\mathbf{x}} v(\mathbf{x}, \mathbf{a}) p(\mathbf{x}) \right\}. \tag{3}$$

Here, we assume a discrete sample space for $\mathbf{x}$. An integral would replace summation for a continuous sample space. The notation PV stands for *prior value*, as this is the value to the decision maker prior to any additional information gathering.

The solution could potentially be simplified considerably after considering the distinction between *coupled* and *decoupled* value. A value function is coupled when it depends on several of the spatial locations through a potentially complex function that cannot be separated into different components for each location. This is not uncommon in many spatial applications, for instance where the value calculation involves a solution of a differential equation of diffusion or fluid flow, etc. If the value function decouples, it can be split into different components, and this makes the solution of the optimization of expected value more tractable as it can be done separately for individual locations. The prior value computation then becomes

$$\mathrm{PV} = \sum_{i=1}^{n} \max_{a_i \in A_i} \left\{ \sum_{x_i} v_i(x_i, a_i) p(x_i) \right\}, \tag{4}$$

where v_i denotes the value function at site $\mathbf{s}_i$.

To motivate the setting, let us build upon the BN model in Figure 1 and consider a decision situation pertaining to CO2 sequestration. Suppose a decision maker is considering whether to pump CO2 from her/his factory into the two reservoirs under consideration. Thus, the decision maker has two alternatives: $a_i \in \{0, 1\}$ at each reservoir $i = 1, 2$, representing whether s/he should sequester CO2 into the reservoirs. Recall that the leaking of a reservoir trap is represented by a binary uncertain outcome $x_i \in \{0, 1\}$ at each reservoir $i = 1, 2$. If the decision maker chooses not to sequester at reservoir i ($a_i = 0$), s/he must pay carbon taxes. If s/he chooses to sequester CO2 at reservoir i ($a_i = 1$), s/he faces the risk of paying a hefty fine if there is CO2 leakage. Figure 1 shows that the marginal probability of leakage is $p(x_i = 1) = 0.1$, $i = 1, 2$.

In this application, it is natural to assume that the value decouples because the costs and potential savings from the two reservoirs are additive. If the decision maker chooses not to sequester at reservoir i, s/he must pay a tax of -2 money units, and the values are thus $v_i(x_i, a_i = 0) = -2$, $i = 1, 2$, no matter what x_i is. Sequestration has a fixed cost of injection set to -1 money units. Moreover, one must pay a fine if leakage occurs ($x_i = 1$) in reservoir i. The fine is -17 units for the larger reservoir 2, while it is -7 units for the smaller reservoir 1. Then values are then $v_1(x_1, a_1 = 1) = -7I(x_1 = 1) - 1$ and $v_2(x_2, a_2 = 1) = -17I(x_2 = 1) - 1$. With these values and the previously defined probabilities we use equation (4) to get

$$\begin{aligned} \text{PV} &= \max\{-7 \cdot 0.1 - 1, -2\} \\ &+ \max\{-17 \cdot 0.1 - 1, -2\} = -3.7, \end{aligned} \tag{5}$$

and the decision is to sequester only at reservoir 1. In Section 3 we compute the VOI for this example when the decision maker can perform seismic tests to get information about sealing or leaking reservoirs.

In some spatial decision situations, the decision maker may have the opportunity to allocate resources in a sequential manner, in which case s/he would be able to select among some alternatives after the effect of previous decisions have been revealed. Such decision situations introduce further computational complexities that make them difficult to solve for large problems. We do not consider such situations in the current paper. Here we assume that the decision maker's downstream decision is a static (one-time) allocation of resources. This assumption is natural in the rock hazard application presented in Section 5 because the decision of bolting the rock formation provides no information about the joints in the rock.

3 Value of Information in Spatial Decision Situations

Spatial decision situations allow for spatial information gathering schemes where data is collected at specific spatial locations. Information gathering is an important auxiliary decision because determining how much and what type of data to collect before an allocation of resources could improve the decision maker's value significantly. Supporting data collection is particularly crucial for spatial risk and decision analysis problems because there is a wealth of opportunities for creative experimentation; for instance, the decision maker may benefit from collecting data only over a small region of the entire domain, possibly at a much lower price. Spatial experimental design methods (Le and Zidek, 2006; Muller, 2007; Dobbie et al., 2008) often consider metrics such as entropy and variance reduction. Other information criteria focus on the minimization of prediction errors (Lilleborge et al., 2016) or the uncertainty in excursion set predictions (Azzimonti et al., 2016). Although very useful in several contexts, these criteria do not relate the information to the ultimate decision that it supports. We consider the VOI instead (Eidsvik et al., 2015).

We denote the data obtained from information gathering by $\mathbf{y}$. Information can be *perfect* or *imperfect*. Perfect information refers to observations that tell the decision maker directly about the distinctions of interest, i.e. $\mathbf{y} = \mathbf{x}$, whereas imperfect information refers to data that provide noisy measurements of these variables. We also distinguish between *total* and *partial* information gathering schemes; in the latter type, only a subset of the uncertain variables are measured.

The relationship between the information $\mathbf{y}$ and the distinctions of interest $\mathbf{x}$ is typically modeled through a likelihood function $p(\mathbf{y}|\mathbf{x})$. The marginal pdf of the data is obtained by marginaliz-

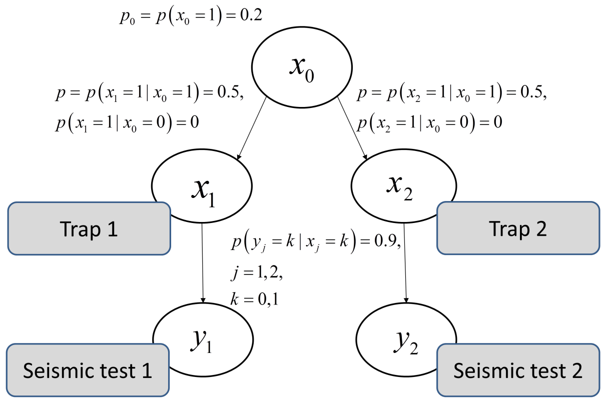


Figure 3: CO2 sequestration example: BN model for the joint distribution at two reservoir traps, along with bottom nodes indicating the results of seismic tests.

ing over all possible outcomes for the uncertain variables $\mathbf{x}$, i.e.

$$p(\mathbf{y}) = \sum_{\mathbf{x}} p(\mathbf{y}, \mathbf{x}) = \sum_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}).$$

The posterior distribution of the distinctions of interest is given by Bayes' rule:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}, \mathbf{x})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}.$$

Returning to the CO2 sequestration example, the bottom nodes in Figure 3 represent the results of seismic tests that provide (imperfect) information about the reservoir traps. The result of a seismic test for a reservoir is conditionally independent of all other variables given the outcome of the trap for the corresponding reservoir. The decision maker could consider purchasing partial information, i.e. seismic results for only one of the reservoirs. The seismic tests are assumed to have binary outcomes where the results either indicate a closed or open structure, which is indicative of the reservoir trap sealing or leaking. The likelihood model is defined by $p\left(y_j = k | x_j = k\right) = 0.9$, $k = 0, 1$, and for reservoirs $j = 1, 2$. This implies that at both reservoirs the chance of observing an open test result for a leaking reservoir equals that of observing a closed test result for a sealing reservoir. If we observe an open result from a seismic test at reservoir 1, the posterior probability of leaking becomes $p(x_1 = 1|y_1 = 1) = \frac{0.9 \cdot 0.1}{0.9 \cdot 0.1 + 0.1 \cdot 0.9} = 0.5$, while $p(x_2 = 1|y_1 = 1) = 0.28$. This is a significant increase from the prior probability of 0.1. The opposite test result gives $p(x_1 = 1|y_1 = 0) = 0.012$, while $p(x_2 = 1|y_1 = 0) = 0.06$. Because of these differences, the partial test could be quite informative as it might change the downstream decision.

Moving to the continuous example, consider a prior Gaussian pdf, $p(\mathbf{x}) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, for the distinction of interest, where the length n mean vector is $\boldsymbol{\mu}$ and the $n \times n$ covariance matrix is $\boldsymbol{\Sigma}$. Assume that one measures this GP at m locations, i.e. $y_j = y(\mathbf{s}_j) = x(\mathbf{s}_j) + N(0, \tau^2)$, $j = 1, \ldots, m$, where the measurement noise variance is τ^2. By extensions of equation (1) and (2) we have:

$$E(x_i|\mathbf{y}) = \mu_i + \boldsymbol{\Sigma}_{i,y}\boldsymbol{\Sigma}_y^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \tag{6}$$

$$\text{Var}(x_i|\mathbf{y}) = \Sigma_{ii} - \mathbf{\Sigma}_{i,y}\mathbf{\Sigma}_y^{-1}\mathbf{\Sigma}_{y,i}, \tag{7}$$

where the marginal pdf of the data is $p(\mathbf{y}) = N(\boldsymbol{\mu}_y, \mathbf{\Sigma}_y)$. The cross-covariance between x_i and $\mathbf{y}$ is denoted $\mathbf{\Sigma}_{i,y}$. Because the noise terms in the measurements are assumed independent, this cross-covariance equals that of respective sites in the GP for $x(\mathbf{s})$. Equations (6) and (7) are used in Section 4 and 5.

3.1 Static information gathering

A static information gathering scheme is one where all the information $\mathbf{y}$ is available to the decision maker simultaneously. The *posterior value* is defined as the value of the situation when the information is available for free, before the downstream spatial decision is made. Since the information is not actually available before it is observed, the posterior value for a risk-neutral decision maker is computed by averaging the maximum expected value over all possible data outcomes. Assuming that the value is decoupled, the posterior value and the VOI are

$$\text{PoV} = \sum_{\mathbf{y}} \sum_{i=1}^{n} \max_{a_i} \left\{ \sum_{x_i} v_i(x_i, a_i) p(x_i|\mathbf{y}) \right\} p(\mathbf{y}),$$

$$\text{VOI} = \text{PoV} - \text{PV}, \tag{8}$$

since the VOI is the difference between the posterior and prior values for a risk-neutral decision maker. The posterior value is computed from maximum conditional expected values, for all data outcomes.

Since the posterior value is an expectation over the data, $\text{PoV}(\mathbf{y})$ is also used for posterior value in the literature, like it is done for expectation and variance with respect to random variable $\mathbf{y}$. Notably, the posterior value and the VOI depend on the design of the data, i.e. where the data is collected. When the decision maker can gather information $\mathbf{y}_j$ according to a particular design j, we will denote the resulting posterior value and VOI by $\text{PoV}(j)$ and $\text{VOI}(j)$. The notation will be clarified in the context of sequential information gathering below.

For a test to be worthwhile, its VOI in equation (8) should be greater than its buying price P. The price of gathering data according to design j is denoted P_j. The decision maker could conduct VOI analysis to explore and compare various potential information gathering schemes.

3.2 Sequential information gathering

Sequential information gathering is possible when the decision maker can decide about subsequent tests after only obtaining partial information. For instance, after performing one test, the decision maker could continue with more testing or choose to stop. The potential tests could be conducted at a single location or over a set of locations. The different tests could potentially also be done using different equipment, perhaps through a laboratory test with negligible error (perfect information), or using just a perfunctory on-site evaluation (imperfect information). In what follows we use letters j, k, and so on, to indicate various kinds of test designs that are done sequentially, and the test results are likewise denoted $\mathbf{y}_j$, $\mathbf{y}_k$, etc.

The sequential posterior value for data $\mathbf{y}_j$ is defined as the value of the situation when the decision maker can observe $\mathbf{y}_j$ for free, and then subsequently have the opportunity for sequential information gathering of further data $\mathbf{y}_k$ at a price P_k, before making the downstream decision. This value and the associated VOI can be obtained using the following set of equations:

$$\begin{aligned}
\text{PoV}_{\text{seq}}(j) &= \sum_{\mathbf{y}_j} \max \left\{ \begin{array}{l} \text{Stop}(j), \\ \max\limits_{k \neq j} \{\text{CV}(k|j)\} \end{array} \right\} p(\mathbf{y}_j), \\
\text{Stop}(j) &= \sum_{i=1}^{n} \max_{a_i} \{E[v_i(x_i, a_i)|\mathbf{y}_j]\}, \\
\text{CV}(k|j) &= \text{Cont}(k|j) - P_k, \\
\text{Cont}(k|j) &= \sum_{\mathbf{y}_k} \max \left\{ \begin{array}{l} \text{Stop}(j,k), \\ \max\limits_{l \neq j,k} \{\text{CV}(l|j,k)\} \end{array} \right\} p(\mathbf{y}_k|\mathbf{y}_j), \\
\text{VOI}_{\text{seq}}(j) &= \text{PoV}_{\text{seq}}(j) - \text{PV}.
\end{aligned} \tag{9}$$

Equation (9) represents an adaptive testing scheme, where one starts by gathering data $\mathbf{y}_j$ and then continues with more testing if further data gathering is worthwhile. Static information gathering with design j, denoted PoV(j), corresponds to always stopping after the first test in the top row of equation (9), i.e. Stop(j). In the sequential information gathering case, the second test is the one which optimizes the continuation value (denoted CV) among all the remaining tests, hence the inner maximization in the top row equation in (9). This CV again allows for the opportunity of stopping or continuing testing after the second test, and so on. The scheme is adaptive because the ordering of data designs and the number of tests depend on the actual data values gathered in the sequence.

Note that there could also be other stopping rules such as constraints for the number of tests, or the decision maker may not allow tests with accumulated prices exceeding a specified budget. The price P_j could in some applications indicate a bound on the additional information value, where the decision maker stops collecting data when no tests can add value exceeding this bound.

In the CO2 sequestration example, sequential information gathering implies that the decision maker can perform seismic testing at only one reservoir, and then continue testing the other if it is worthwhile to do so. For some outcomes of the first test, the decision maker might stop testing, whereas s/he might continue with a test at the other reservoir for other first-test results. Considering this adaptivity in information gathering, the value of sequential testing would always be larger or equal to that of static testing. Since there are only two binary data sources in the CO2 sequestration example, one can compute the optimal sequential information gathering strategy defined by equation (9) exactly.

Figure 4 demonstrates the notion of sequential information gathering in this example using a (partial) decision tree. Here, the price of seismic testing is set to $P_1 = P_2 = 0.3$ monetary units. (Below we compare testing options over price ranges.) The value numbers indicated in Figure 4 are the expected values when one conducts no further testing or with (continued) testing, at a price that is subtracted. As an example, assume one starts with a test at reservoir 1, and the test result is closed: $y_1 = 0$. Then $p(x_1 = 1|y_1 = 0) = 0.012$ and $p(x_2 = 1|y_1 = 0) = 0.06$. With no further testing, and using the monetary values of equation (5), it is then optimal to sequester at reservoir 1 because $-7 \cdot 0.012 - 1 = -1.09 > -2$, while it is optimal to avoid sequestration at reservoir 2 because $-17 \cdot 0.06 - 1 = -2.02 < -2$. The overall value is then $-1.09 - 2 = -3.09$. The

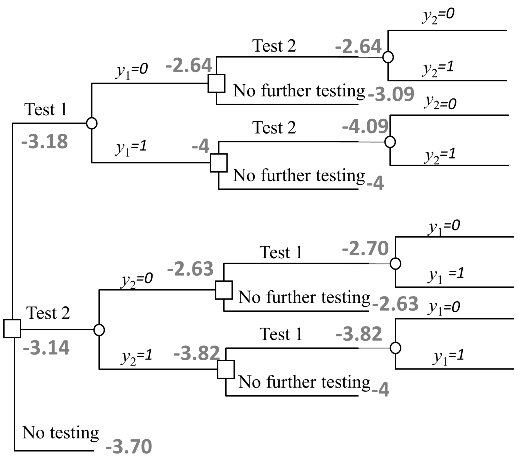


Figure 4: CO2 sequestration example: Illustration of sequential information gathering using a partial decision tree.

value of continued testing is obtained by averaging over the second test result, conditional on the first result ($y_1 = 0$), and subtracting the price P_2. This expected continuation value is -2.64. It is thus optimal to continue testing with this outcome of the first test, because $-2.64 > -3.09$. If the result of the first test is open: $y_1 = 1$, it is optimal to stop testing as this has value -4, while the expected value of continued testing is -4.09.

Figure 5 shows decision regions for adaptive information gathering in the CO2 sequestration example, indicating the optimal information gathering scheme over potential prices of the two seismic tests. This figure shows three displays. The middle one represents the model with the probabilities specified in Figure 3, while the others represent minimal (top) and maximal (bottom) correlation between the two reservoir traps, while keeping $p(x_i = 1|x_0 = 0) = 0$, $i = 1, 2$, and maintaining the marginal probabilities of sealing/leaking in Figure 1. The minimal correlation case is obtained by setting $p(x_0 = 1) = 1$ and $p(x_i = 1|x_0 = 1) = 0.1$, $i = 1, 2$. The maximum correlation case is obtained by setting $p(x_0 = 1) = 0.1$ and $p(x_i = 1|x_0 = 1) = 1$, $i = 1, 2$. All displays have the same value functions, as described above. The decision regions are constructed by computing the VOI for different tests, and then subtracting the price of relevant tests. The case denoted 'Both' means that seismic tests are at both reservoirs simultaneously is optimal, without any sequentiality. 'Only 1', or 'Only 2', means that testing at only one reservoir is optimal, and there is nothing to gain by adaptive testing at the other. 'Seq 1->2', or 'Seq 2->1', means that there is clearly a gain in adaptive testing, and the decision maker should consider sequential information gathering, when prices are in this region. In these situations it is beneficial to continue testing for one outcome of the first test, while it is optimal to stop for the other outcome.

Consider now the middle display in Figure 5: when both seismic tests are inexpensive, it is optimal to do both tests at once. As the price of a seismic test at the first reservoir increases, it becomes optimal to test the other reservoir first, with the option of sequential testing at the first reservoir. The price configuration $P_1 = P_2 = 0.3$, shown in Figure 4, is in the region where sequential testing starting at reservoir 2 is best. If the price of reservoir 1 gets very large, it is optimal to test only at reservoir 2. Note that sequential information gathering would always be

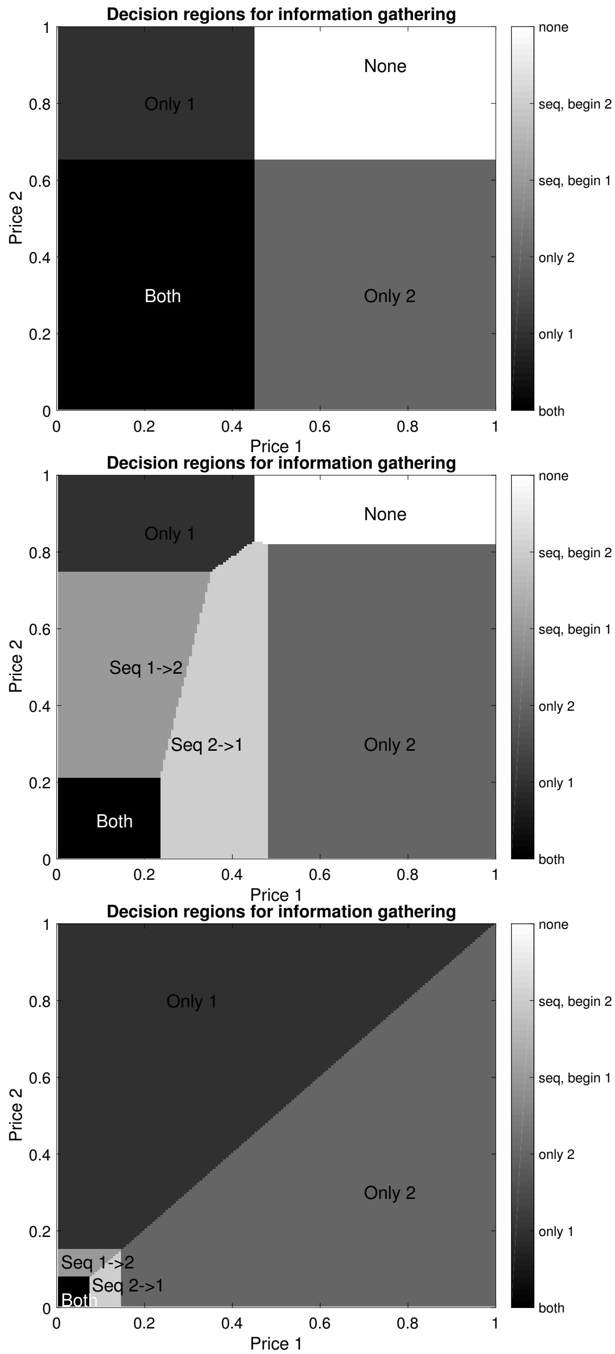


Figure 5: CO2 sequestration example: Decision regions with the optimal information gathering schemes. Top: minimum correlation. Middle: medium correlation (reference case). Bottom: maximum correlation.

better or equal to performing 'One' or 'Both' in this situation, so to compare static and sequential information gathering in Figure 5 we show 'Only 1', 'Only 2' or 'Both' when this value is the same as the sequential strategy.

In the situations with more extreme correlations (top and bottom displays of Figure 5), there is little additional value of sequential testing. Here the optimal information gathering strategy is almost always to acquire one test, both tests or none. This could be important in practice - if the dependence is very large or small, there is less value in sequential testing over static tests.

4 Heuristic Strategies

Calculating the values associated with sequential information gathering entails a series of maximizations and summations (or integrals when the sample space is continuous), such as in equation (9). Computations grow combinatorially with the number of possible tests, as should be evident from the decision tree in Figure 4. Thus, exact evaluation is only possible in small-size problems, where one can store the computations from all strategies. For large-size problems, approximations of the expected values are necessary (Powell, 2011). The values of the resulting strategies are sub-optimal. These techniques sacrifice optimality for tractability.

In this section, we present some common heuristic strategies that approximate the sequential forward computations required to calculate the continuation value at any decision epoch. It is our intention to explore information gathering schemes that are simple, yet have the ability to exploit dependence in spatial risk and decision analysis applications.

For GP models like the one deployed in the application in Section 5, there exist partly closed-form expressions for the posterior values under some working assumptions, as has been noted in previous spatial VOI literature (Bhattacharjya et al., 2013). These expressions are also effective in the selection of sequential tests. Consider again a Gaussian variable $\mathbf{x} = (x_1, \ldots, x_n)$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Assume the decoupled value function (equation (4)) is defined by $v_i(x_i, 0) = 0$ and $v_i(x_i, 1) = x_i$ for the two alternatives $a_i \in \{0, 1\}$, $i = 1, \ldots, n$. Then, Bhattacharjya et al. (2013) show that the posterior value of a static information gathering scheme for Gaussian data $\mathbf{y}_j$ is

$$\sum_{i=1}^{n} \int \max\{0, E(x_i|\mathbf{y}_j)\}\, p(\mathbf{y}_j) d\mathbf{y}_j = \sum_{i=1}^{n} [\mu_i \Phi(\mu_i/r_i) + r_i \phi(\mu_i/r_i)], \tag{10}$$

where $r_i = \sqrt{\boldsymbol{\Sigma}_{i,j} \boldsymbol{\Sigma}_{\mathbf{y}_j}^{-1} \boldsymbol{\Sigma}_{j,i}}$, $\boldsymbol{\Sigma}_{\mathbf{y}_j}$ is the covariance matrix of the data $\mathbf{y}_j$, and $\boldsymbol{\Sigma}_{i,j}$ is the cross-covariance between x_i and data $\mathbf{y}_j$. Furthermore, $\phi(z)$ and $\Phi(z)$ are the pdf and cumulative distribution function of the standard normal distribution. We will use equation (10) and related results for computing the value for three heuristic strategies for sequential information gathering.

An important reason for computational infeasibility in large-scale sequential problems is that one must compare all the remaining potential tests for the exact solution. In one class of heuristics, the order of potential tests is fixed before the sequential decision problem is solved; we refer to these as *fixed-order* heuristics. The computational advantage in these heuristics is that the determination of the order can be done offline, before the sequential problem is solved. Importantly, during the online solution of the problem, the computation only needs to consider whether or not to continue to the next test, and the potential test is pre-determined in the fixed order.

A second aspect of computational infeasibility for large-scale sequential problems is that the dynamic program grows quickly when several subsequent decision epochs need to be considered. In our heuristics, we approximate this away in our computations through a working assumption that at most one more test will be done at any decision epoch. This implies that the maximization in the continuation value in equation (9) is conducted by only comparing whether to stop at any given epoch or perform one more test and then stop. It is often necessary to approximate the sequential problem in this fashion, neglecting the sequential nature of the problem beyond a few subsequent epochs; this approach is common in approximate dynamic programming (Martinelli et al., 2013a; Goodson et al., 2017). For special value functions one can derive bounds for these heuristics, using for instance submodular properties (Golovin and Krause, 2011; Chen et al., 2015).

The first two heuristics that we consider here are both fixed-order heuristics; the difference between these heuristics lies in how the fixed order is determined. The first heuristic is the **naive** selection strategy for tests. This scheme is established by first ranking the possible tests $j = 1, \ldots, m$ based on the difference between the static VOI(j) of test $\mathbf{y}_j$ and the price P_j of the test. This order is denoted by $j_{(1)}$, $j_{(2)}$, etc. At each stage, the decision maker will either stop testing, or continue with the next pre-detemined test. In practice, even though the order of tests is fixed, the number of tests actually done will depend on the realized data. This is because the values for the stopping or continuation options change with the data. To evaluate the strategy we 'play the game' of running the approach over simulated data samples. This can be used to study the depth of testing, and it also gives an estimate of the VOI. Say, for realization b, one either stops after the first test $\mathbf{y}^b_{j_{(1)}}$, with expected values defined by equation (6) for data $\mathbf{y}^b_{j_{(1)}}$. Alternatively, one performs test two $\mathbf{y}^b_{j_{(2)}}$ if the continuation value

$$\mathrm{CV}^b(j_{(2)}|j_{(1)}) = \sum_i [\mu^b_{i|1}\Phi(a^b_{i,2|1}) + r_{i,2|1}\phi(a^b_{i,2|1})] - P_{j_{(2)}}, \tag{11}$$

is larger than the value of stopping. Here, $a_{i,2|1} = \mu^b_{i|1}/r_{i,2|1}$, and $\mu^b_{i|1}$ and $r_{i,2|1} = \sqrt{\mathbf{\Sigma}_{i,2|1}\mathbf{\Sigma}^{-1}_{2|1}\mathbf{\Sigma}_{2,i|1}}$ are defined from the joint Gaussian distribution of x_i and $\mathbf{y}_{j_{(2)}}$, given $\mathbf{y}_{j_{(1)}}$ (see equation (6) and (7)). If the decision maker conducts the second test, the same stop or continue procedure is done for the third test $\mathbf{y}_{j_{(3)}}$, and so on.

A potential way to include more of the dependence in the spatial model is to rank the tests based on the added inclusion of one more test in the information gathering scheme. This is done in the second heuristic which we refer to as the **naive-expand** selection strategy. We start with the best test (just like in the naive scheme), and next choose the test which increases the VOI the most when performed together with the first selected test. If we again let $j_{(1)}$ be the index of the best single test, the second test is chosen according to:

$$j_{(2)} = \mathrm{argmax}_j\{\mathrm{VOI}(j_{(1)}, j) - P_j\}. \tag{12}$$

Subsequent tests, 3, 4, etc, are selected in a similar fashion, by computing the VOI jointly with all previous tests in the order. Note that the path of this strategy will in general be different from the naive scheme. If there is significant correlation between the tests, this strategy will more effectively avoid tests that carry almost the same information as previous test(s). This strategy is also evaluated by Monte Carlo sampling, which at every stage works in a similar stop or continue procedure as for the naive strategy in equation (11).

These fixed-order heuristics are easy to evaluate, and could likely be extended further, but they are not adaptive in the sense that their ordering is pre-determined and does not change with the data that are gathered. The third heuristic is the **myopic** strategy, which is a well-known adaptive approach in sequential decision problems. This strategy is not a fixed-order heuristic because the order of the remaining tests may change depending on the observed test results. At any decision epoch, all the observations till date are used to update the decision maker's beliefs about the remaining uncertainties, and the decision maker chooses an alternative assuming that this will be the final choice s/he will make. In our information gathering formulation, the decision maker can choose the test to perform next, if it is profitable to perform a test at all.

To evaluate the myopic strategy we again 'play the game' of applying Monte Carlo simulation. But unlike the naive strategies this heuristic must select among all possible continuation tests, or stopping, since the myopic strategy does not give a pre-determined sequence of potential tests. We sample data variables and find the next candidate for data gathering (or stop) by maximizing the increase in posterior value for every sample. Similar to the previous heuristics, assume that data $\mathbf{y}^b_{j_{(1)}}$ have been collected in the best first testing alternative $j_{(1)}$. The next stage test selection for this sample is determined by

$$\begin{aligned} \text{Stop}^b(j_{(1)}) &= \sum_i \max\left\{0, E(x_i|\mathbf{y}^b_{j_{(1)}})\right\}, \\ \text{CV}^b(j|j_{(1)}) &= \max_{j \neq j_{(1)}} \left\{\text{Cont}^b(j|j_{(1)}) - P_j\right\}, \\ \text{Cont}^b(j|j_{(1)}) &= \sum_i [\mu^b_{i|1}\Phi(a^b_{i,j|1}) + r_{i,j|1}\phi(a^b_{i,j|1})], \end{aligned} \tag{13}$$

with similar notation as in equation (9) and (11).

The values of the three different strategies are obtained by averaging value results over the Monte Carlo samples.

5 Rock Steady: A Rock Hazard Application

In this section, we study a mining rock hazard application (Karam et al., 2007; Ellefmo and Eidsvik, 2009; Zetterlund et al., 2011). Figure 6 shows $n = 52$ tunnel locations in a mine. The tunnel locations (marked by x) are at risk of rock fall, which could incur a large cost for the mining company. They must decide whether to prevent rock fall via extensive bolting, or to avoid this costly operation at the risk of rock hazard that occurs at a random rate. The frequency of joints in the rocks are critical for the stability.

The mining company is considering information gathering schemes that could help them make a better rock reinforcement decision. We study the VOI of joint frequency measurements in boreholes for this application. A total of 30 possible boreholes have been designed (plotted in Figure 6), and data acquired in these boreholes will be indicative of the joint frequency at the locations where data are collected. Because of the spatial dependence, this information will also propagate to be informative about the joint frequency in the vicinity of the boreholes.

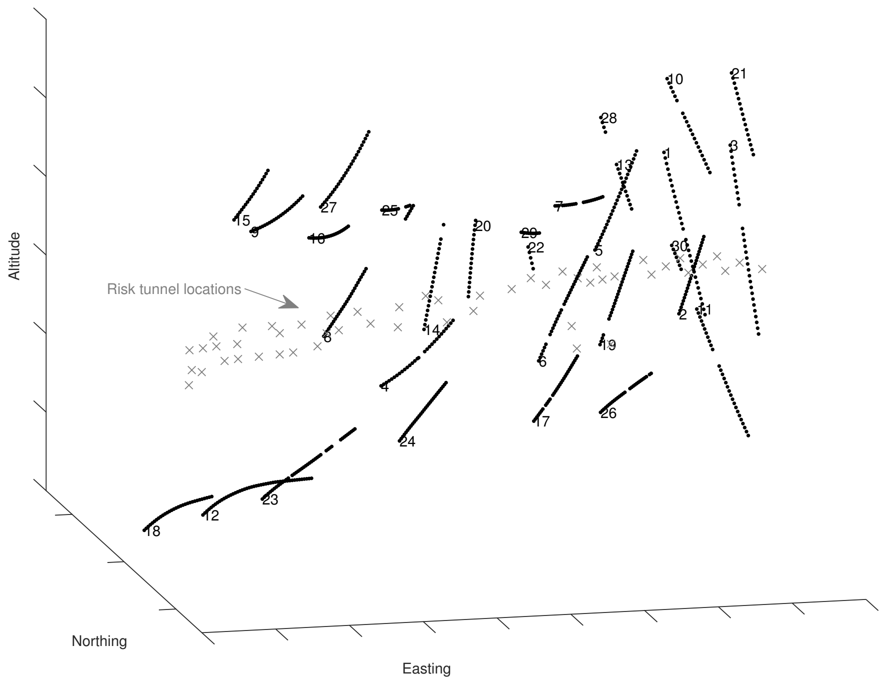


Figure 6: Risk tunnel locations and borehole design of 30 boreholes for joint frequency data.

5.1 Modeling

We use a GP to model the joint frequency in the mine, with pdf $p(\mathbf{x}) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the distinctions of interest $\mathbf{x} = (x_1, \ldots, x_{52})$. Adjusted from earlier studies with joint frequency data (Ellefmo and Eidsvik, 2009), we assign a fixed mean $\mu_i = 35$ for all sites, variance $\sigma^2 = 10^2$, and reference correlation length $3/\eta = 300$. Sensitivity results to other correlation lengths are also discussed. We assume that the data are imperfect measurements of the joint frequency where the boreholes are drilled. The measurement noise terms are assumed to be independent and Gaussian distributed with variance $\tau^2 = 0.1^2$. It is worth noting that the joint frequency data are counts and could be modeled by Poisson distributions (Ellefmo and Eidsvik, 2009; Eidsvik et al., 2015; Evangelou and Eidsvik, 2017) having a GP as latent intensity. For rather large counts, which is the case here, the GP is a reasonable model approximation.

One must further specify costs for stabilizing the rock mass through bolting and costs for rock fall, if it occurs. Here, the decision situation takes the form: $\max\{-C_b, -C_f E(x_i)\} = -C_b + \max\{0, C_b - C_f E(x_i)\}$, $i = 1, \ldots, n$. Note that this decision rule is of a similar form as that of equation (10), and for the VOI analysis, we can use results from Section 4 since the prior and posterior of $\mathbf{x}$ is Gaussian. We assume the cost parameters are $C_f = 1$ and $C_b = 30$ monetary units.

The boreholes have lengths between 5 and 48 samples, but most of them have $25 - 35$ measurements, see Figure 6. The price of acquiring and processing data is assumed to be proportional

Table 1: Length of borehole, VOI and average posterior prediction variance, for seven boreholes.

Borehole	1	2	3	4	5	6	7
Length	30	30	30	25	29	33	27
VOI	5.05	3.28	3.02	3.02	2.97	2.65	2.65
Ave.Pred.Var	9.69	9.79	9.81	9.81	9.82	9.83	9.83

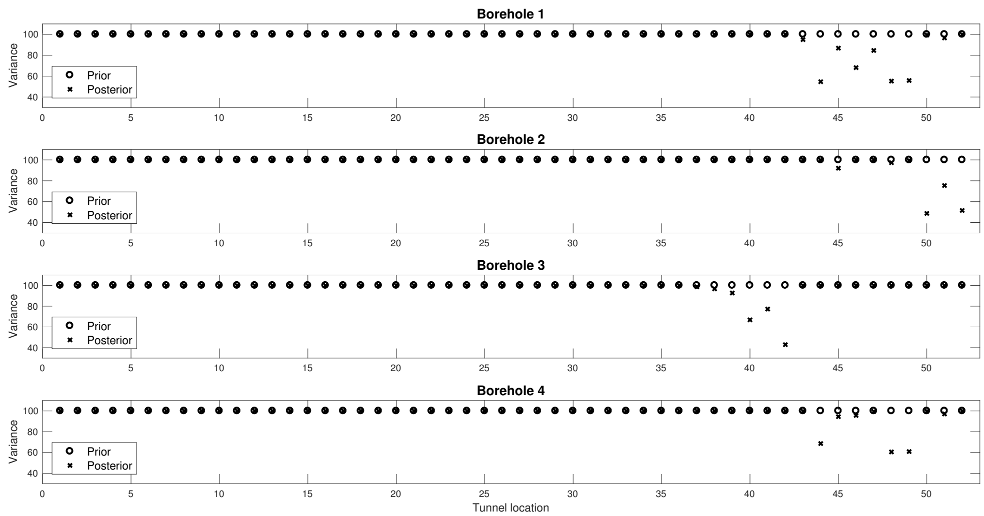


Figure 7: Variance reduction obtained by data in single boreholes 1, 2, 3 and 4.

to the lengths of the boreholes, and we set price $P_j = 0.05m_j$, where m_j is the number of samples in borehole j.

5.2 VOI analysis

Table 1 shows initial analysis of the planned borehole data, presenting the lengths of boreholes, the static VOI of a single borehole, and the average prediction variance at the 52 risk sites when data is collected in a borehole. The borehole numbers $(1, \ldots, 30)$, which are shown in Figure 6, are here defined according to the ranks of single borehole VOI results. Borehole 1 is the most valuable because it is close to several of the identified tunnelling locations. Assuming no further opportunities for information gathering, one should gather data in this borehole if the price of data acquisition and processing is less than 5.05. Borehole 1 also gives the smallest average prediction variance in the conditional distribution.

Figure 7 shows the prior and posterior variance at all 52 risk locations, when we have conditioned on data in a single borehole, see equation (7). Data in borehole 4 (bottom row in Figure 7) clearly reduce the uncertainty at some tunnel locations. However, if one is in the setting sequential information gathering, where one would already have collected data in borehole 1 and 2, the additional uncertainty reduction achieved by borehole 4 is likely to be much smaller than

Table 2: VOI for different strategies, and for smaller (top) and larger (bottom) correlation. Monte Carlo sample size is 1000. The 90 percent uncertainty intervals are based on bootstrapping replicate value results.

Corr. range: 225m				
	Static	Naive	Naive-expand	Myopic
VOI (1)	3.6	5.9 (5.4-6.3)	13.5 (12.7-14.2)	17.0 (16.3-17.8)
Comp.Time (sec)	0.1	3	7	100
Ave.Depth(1)	1	2.4 (2.3-2.5)	6.5 (6.4-6.6)	8.7 (8.6-8.8)
Corr. range: 300m				
	Static	Naive	Naive-expand	Myopic
VOI (1)	5.1	8.0 (7.3-8.5)	18.6 (17.6-19.7)	25.3 (24.1-26.4)
Comp.Time (sec)	0.1	3	7	100
Ave.Depth(1)	1	2.4 (2.3-2.5)	5.6 (5.5-5.7)	10.2 (10.1-10.4)

in the bottom row of Figure 7 because borehole 4 provides similar information as 1 and 2. The naive-expand strategy would recognize this by the forward expansion of the next best boreholes in the fixed ordering of sequential tests. Borehole 4, which is ranked fourth in the naive strategy, thus gets a poor rank for the naive-expand strategy.

We next run the strategies for computing the VOI of sequential testing starting with borehole 1. The results are summarized in Table 2. Having the opportunity of sequential testing is clearly beneficial in this case, since the VOI increases with the complexity of the strategies. Moreover, the value is larger for 300 m spatial correlation range (bottom) compared with 225 m (top); when the correlation range is larger, data gathering in boreholes provide more information at all risk sites. It should be useful to incorporate a workflow that can gauge such sensitivity effects of model parameters on the VOI. The computation time is very small for the static and the fixed-order strategies. It is larger for the myopic strategy, where a significant time is spent on computing the optimal order for the different data samples. Such numbers of course depend on the implementation, software and hardware, but they are nevertheless indicative of the relative computer times. The average depths (i.e. number of tests performed) are clearly larger for the naive-expand strategy compared with the naive. It gets even larger for the myopic strategy. This average depth does not seem to depend so much on the spatial correlation length 225 m or 300 m. We study sensitivity to more extreme correlations below.

Figure 8 shows the testing stages for the three strategies (left to right). This is illustrated for two Monte Carlo samples (top-bottom). For the naive and naive-expand strategies, the paths are predetermined, but the number of stages depends on the realized tests in the Monte Carlo sample. The naive strategy (left plots) tends to stop after testing at borehole 1 and 2. For some data samples, it continues testing over more stages, like in the lower left display. The naive-expand strategy (middle plots) usually takes more steps than the naive strategy before it stops testing (see Table 2). For the myopic strategy (right plots), the paths depend on the Monte Carlo samples.

We now study the effect of extreme spatial dependence conditions. When the correlation range is extremely large; 30 km, the VOI of static testing in borehole 1 is 81. In this case the value of sequential information is approximated to 88 (naive), 93 (naive-expand) and 94 (myopic). For such very large dependence, knowing the outcome in the first borehole is enough in most cases; thus it is rarely worthwhile conducting more tests, and the additional value of sequential testing is small.

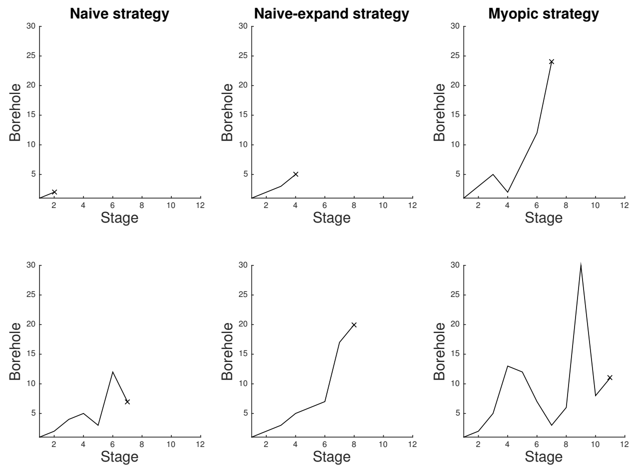


Figure 8: Borehole data strategies for two Monte Carlo realizations of tests.

In the opposite situation with a smaller correlation range of 125 m, the VOI of static testing in borehole 1 is 1.2. The value of sequential information is now approximated to 1.2 for both naive and naive-expand strategies, which never perform more than 1 test for the 1000 Monte Carlo samples. For the myopic strategy, the approximated value is 1.4, and the average depth of this strategy is now 2.1. Thus there is little benefit from a sequential strategy in this situation as well. The largest additional value of sequential testing tends to occur for moderate to high spatial dependence, when the information in one borehole can influence the posterior pdf at other decision locations, but there is still remaining uncertainty so more data could be useful.

6 Conclusions

In this paper, we studied sequential information gathering schemes for spatial risk and decision analysis applications. Sequential testing implies that a decision maker has the opportunity to continue testing after the results of former tests are revealed, if it is worthwhile to do so. It has been our intention to provide tutorial explanation of concepts and interpretation of results, under different spatial models.

We learned that statistical dependence does indeed have a bearing upon the suitability of sequential testing strategies. In the CO2 sequestration example, the additional value of sequential testing is large for moderate to high correlations, while it is negligible for very low and very high correlations. In the rock hazard application, we further learned that the value of sequential testing, much like static testing, depends on the spatial correlation and on the possible spatial designs of the tests, and that a wisely chosen (sequential) strategy can provide the decision maker with a significant value gain. We recommend a workflow studying the sensitivity to different correlations in the spatial model.

For large-size problems, like with our rock hazard example, there is no tractable optimal solution to the dynamic program involved in value of sequential information calculations. Here, we used common heuristics such as the myopic approach that has been shown to be easy to implement and is very popular in practice. The myopic approach is truly adaptive in our case: the sequential data gathering scheme will depend on the observed data. Note that the heuristic strategies presented here could be improved upon, at the expense of computational time.

References

Azzimonti D, Bect J, Chevalier C, Ginsbourger D (2016) Quantifying uncertainties on excursion sets under a Gaussian random field prior. SIAM/ASA Journal Unc Quant 4:850–874

Baio G (2012) Bayesian Methods in Health Economics. CRC Press

Banerjee S, Carlin BP, Gelfand AE (2014) Hierarchical modeling and analysis of spatial data (2nd Ed). CRC Press

Bellman RE (1957) Dynamic Programming. Princeton Univ Press

Bhattacharjya D, Eidsvik J, Mukerji T (2010) The value of information in spatial decision making. Math Geosc 42:141–163

Bhattacharjya D, Eidsvik J, Mukerji T (2013) The value of information in portfolio problems with dependent projects. Dec Anal 10:341–351

Bonneau M, Gaba S, Peyrard N, Sabbadin R (2014) Reinforcement learning-based design of sampling policies under cost constraints in Markov random fields: Application to weed map reconstruction. Comp Stat & Data Anal 72:30–44

Bratvold RB, Bickel JE, Lohne HP (2009) Value of information in the oil and gas industry: Past, present, and future. SPE: Reserv Eval & Eng 12:630–638

Brown DB, Smith JE (2013) Optimal sequential exploration: Bandits, clairvoyants, and wildcats. Operat Res 60:262–274

Chen Y, Javdani S, Karbasi A, Bagnell JA, Srinivasa, S, Krause A (2015) Submodular surrogates for value of information. Proc. of 29th AAAI Conf on Artif Intel 3511-3518

Chiles JP, Delfiner P (2012) Geostatistics: Modeling Spatial Uncertainty (2nd Ed). Wiley

Convertino M, Munoz-Carpena R, Kiker GA, Perz SG (2015) Design of optimal ecosystem monitoring networks: hotspot detection and biodiversity patterns. Stoch Environ Res and Risk Assess 29:1085–1101

Cressie N (1993) Statistics for Spatial Data. Wiley

Cressie N, Wikle CK (2011) Statistics for Spatio-Temporal Data. Wiley

Dobbie MJ, Henderson BL, Stevens DL (2008) Sparse sampling: Spatial design for monitoring stream networks. Stat Surveys 2:113–153

Dyer JS, Sarin R (1979) Measurable multiattribute value functions. Operat Res 27:810–822

Eidsvik J, Bhattacharjya D, Mukerji T (2008) Value of information of seismic amplitude and CSEM resistivity. Geophysics 73:R59–R69

Eidsvik J, Mukerji T, Bhattacharjya D (2015) Value of Information in the Earth Sciences. Cambridge Univ Press

Ellefmo SL, Eidsvik J (2009) Local and spatial joint frequency uncertainty and its application to rock mass characterization. Rock Mech and Rock Eng 42:667–688

Evangelou E, Eidsvik J (2017) The value of information for correlated GLMs. J of Stat Plan and Infer 180:30–48

Golovin D, Krause A (2011) Adaptive submodularity: Theory and applications in active learning and stochastic optimization. J of Artif Intel Res 42:427–486

Goodson JC, Thomas BW, Ohlmann JW (2017) A rollout algorithm framework for heuristic solutions to finite horizon stochastic dynamic programs. Eur J of Operat Res 258:216–229

Howard R (1966) Information value theory. IEEE Trans on Sys Sci and Cyb 2:22–26

Howard R, Abbas A (2015) Foundations of Decision Analysis. Prentice Hall

Karam KS, Karam JS, Einstein HH (2007) Decision analysis applied to tunnel exploration planning I: Principles and case study. J of Construct Eng and Manag 133:344–353

Keisler JM, Collier Z, Chu E, Sinatra N, Linkov I (2014) Value of information analysis: The state of application. Environ Sys and Dec 34:3–23

Krause A, Guestrin C (2009) Optimal value of information in graphical models. J of Artif Intel Res 35:557–591

Le ND, Zidek JV (2006) Statistical Analysis of Environmental Space-Time Processes. Springer-Verlag

Lilleborge M, Hauge R, Eidsvik J (2016) Information gathering in Bayesian networks applied to petroleum prospecting. Math Geosc 48:233–257

Malczewski J (2006) GIS-based multicriteria decision analysis: A survey of the literature. Int J of Geogr Inf Sci 20:703–726

Martinelli G, Eidsvik J, Hauge R, Forland MD (2011) Bayesian networks for prospect analysis in the North Sea. AAPG Bull 95:1423–1442

Martinelli G, Eidsvik J, Hauge R (2013a) Dynamic decision making for graphical models applied to oil exploration. Eur J of Operat Res 230:688–702

Martinelli G, Eidsvik J, Sinding-Larsen R, Rekstad S, Mukerji T (2013b) Building Bayesian networks from basin modeling scenarios for improved geological decision making. Petr Geosc 19:289–304

Martinelli G, Eidsvik J (2014) Dynamic exploration designs for graphical models using clustering with applications to petroleum exploration. Knowl Based Sys 58:113–126

Matheson J, Howard R (1968) An introduction to decision analysis. In R. Howard and J. Matheson (Eds) The Princip and Applic of Dec Anal, Vol. I. Strat Dec Group 17–55

Mathieson A, Midgely J, Wright I, Saoula N, Ringrose P (2011) In Salah CO2 storage JIP: CO2 sequestration monitoring and verification technologies applied at Krechba, Algeria. Energy Proc 4:3596–3603

Miller AC (1975) The value of sequential information. Manag Sci 22:1–11

Muller W (2007) Collecting Spatial Data. Springer-Verlag

Powell WB (2011) Approximate Dynamic Programming: Solving the Curses of Dimensionality (2nd Ed). Wiley

Puterman ML (2005) Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley

Raiffa H (1968) Decision Analysis: Introductory Lectures on Choices under Uncertainty. Addison-Wesley

Rasmussen CE, Williams C (2006) Gaussian Processes for Machine Learning. MIT Press

Simon J, Kirkwood CW, Keller LR (2014) Decision analysis with geographically varying outcomes: Preference models and illustrative applications. Operat Res 62:182–194

Srinivas N, Krause A, Kakade S, Seeger M (2010) Gaussian process optimization in the bandit setting: No regret and experimental design. Proc of the 27th Int Conf on Mach Learning.

Wang H, Harrison KW (2013) Bayesian approach to contaminant source characterization in water distribution systems: adaptive sampling framework. Stoch Environ Res and Risk Assess 27:1921–1928

Yokota F, Thompson K (2004a) Value of information literature analysis: A review of applications in health risk management. Med Dec Making 24:287–298

Yokota F, Thompson K (2004b) Value of information analysis in environmental health risk management decisions: Past, present, and future. Risk Anal 24:635–650

Zetterlund M, Norberg T, Ericsson LO, Rosen L (2011) Framework for value of information analysis in rock mass characterization for grouting purposes. J of Construct Eng and Manag 137:486–497