NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

The Value of Information in Spatial Decision Making

by

Debarun Bhattacharjya, Jo Eidsvik and Tapan Mukerji

PREPRINT
STATISTICS NO. 3/2006

NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2006/S3-2006.pdf
Jo Eidsvik has homepage: http://www.math.ntnu.no/∼joeid
E-mail: joeid@stat.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology,
N-7491 Trondheim, Norway.

The Value of Information in Spatial Decision Making

Debarun Bhattacharjya[1], **Jo Eidsvik**[2] **and Tapan Mukerji**[3]
1) Management Science & Engineering, Stanford University, California
2) Department of Mathematical Sciences, NTNU, Norway
3) Rock Physics Laboratory, Stanford University, California

Abstract

Decisions involving selection of sites over a lateral domain with a spatially correlated distinction of interest are common in several realms. In this paper we use the decision-analytic notion of *value of information* on models common to spatial statistics. We formulate methods to evaluate monetary *values* associated with experiments performed in the spatial decision making context, including the prior value, the value of perfect information, and the value of the experiment. The prior for the spatial distinction of interest is assumed to be a Markov random field where the value at each spatial site belongs to a finite set of states. The likelihood distribution can take any form depending on the experiment one decides to acquire. Typical experiment types are binary registration or a Gaussian measurement at selected spatial sites. We demonstrate how to efficiently compute the value of an experiment for Markov random fields of moderate size. The most computationally demanding task is solving an integral over the result of the experiment under evaluation, which we accomplish using Monte Carlo integration. We explore and compare some measures for the worth of an experiment in our problem context. Our methods are illustrated on two examples. One is relevant to conservation biology, where the downstream decision is the spatial allocation of reserve sites. The other application is motivated by seismic exploration in the petroleum industry. These simple examples demonstrate the complex interplay between the likelihood of the experiment, spatial interaction and the nature of the downstream decision and its associated values. The approach naturally fosters a multi-disciplinary outlook for valuing information in spatial decision making, and stimulates ideas for creative alternatives in decisions related to experimentation.

Keywords: value of information, Markov random field, spatial decision making, experimental design.

1. Introduction

Gathering the right kind and right amount of information is crucial for any decision making process. Auxiliary decision opportunities regarding information gathering are often created when an important decision is to be made in the future. This downstream decision may have a lot at stake and it may be worthwhile to obtain more information before the decision is actually taken, i.e. before an irrevocable allocation of resources. A crucial question to answer is - how much information should one purchase? This is a decision which is related to the well-established concept of value of information, also known as value of clairvoyance, in decision analysis (Howard, 1966; Raiffa, 1968; Matheson, 1990). The value of information (VOI) for a particular information gathering scheme is the maximum monetary amount that a decision maker should be willing to pay to acquire it. The VOI depends on several factors, including the prior probabilities, the "quality" of the test and the decision maker's utility curve.

In this paper we present models that compute the value of information for experiments performed in the context of spatial decision making. We use the phrase *spatial decision making* to refer to decision problems with two important characteristics: 1) the decision generally involves a choice of alternatives over space, for instance, selecting sites; and 2) the distinction of interest is typically spatially correlated. There are several applications that are relevant within this context. Petroleum exploration and production is a natural contender as a possible application: here the distinction of interest is the presence or absence of oil; the presence of oil is spatially correlated in a reservoir, and the decision maker must decide where to drill wells to recover oil and maximize profits. Another application is conservation biology where decisions are made regarding the number of reserve sites and their spatial location. Here the distinction of interest is the presence or absence of a particular species. We illustrate our methods on examples from both these applications.

We briefly discuss some recent work that can be viewed as spatial decision making, but none of which include the full generality that we adopt in this paper. Polasky and Solow (2001) present issues regarding the value of information in conservation biology. They use decision-analytic concepts to investigate the value of information of surveys in the reserve site selection problem. Reserve site selection refers to the selection of sites for establishing biological reserves with the purpose of conserving and nurturing certain species. In their paper, Polasky and Solow show that inferences about value of information in such a setting can often be counter-intuitive. However, they assume that the species incidences are independent and use a value criterion that is not associated with monetary units. Houck and Pavlov (2006) estimate the value of information for electromagnetic surveys in petroleum exploration using the decision-analytic approach. Although their case study is illustrative, they work at the global level of the reservoir by using a simple decision tree formulation. Diggle and Lophaven (2006) describe a Bayesian hierarchical model for geostatistical design of monitoring sites for salinity at sea. Their goal is to minimize spatial prediction variance. With their philosophy it is hard to relate the statistical model to the downstream decision and its associated values, and therefore to compare the worth of an experiment with its cost. In fact, downstream decisions are rarely treated directly in geostatistical situations; the focus is mostly on prediction variance and parameter estimation, see e.g. Müller (2001).

We introduce a decision-analytic formulation to situations that naturally exhibit spatial dependence. We use a categorical Markov random field (MRF) model (Besag, 1974) for the distinction of interest. Other Bayesian prior models can also be used. An advantage with moderate size

MRFs is that recursive assessment of such categorical MRFs (Reeves and Pettitt, 2004) allows us to calculate the marginal probabilities and to sample directly from these models. We are interested in evaluating experiments that are performed to obtain more information about the distinction of interest. We use a monetary value measure to ensure that the value of information corresponds directly to the maximum a decision maker should be willing to spend to purchase an experiment. Our models can easily be extended to deal with other measures of value. For the sake of clarity, we assume throughout the paper that the decision maker is risk-neutral (Raiffa, 1968). In other words, the decision maker is indifferent between a lottery of uncertain monetary prospects and the expected value of the lottery. Our models can incorporate risk aversion or risk seeking behavior if required.

Section 2 develops basic notation for the rest of the paper. In Section 3 we present the model assumptions and equations. This section explains how we integrate spatial statistics with different approaches to valuing information; it is our main contribution in this work. Our focus is on the decision-analytic philosophy but we also mention a formulation using entropy as a measure for valuing information in spatial decision making. In Section 4 we describe an algorithm that couples Monte Carlo simulation and a recursive method for computing the value of information of an experiment. This computational method is crucial for assessing the value of information in practice. The algorithm is demonstrated on several examples in Section 5 to develop critical insights. We begin with a simple hypothetical example from the domain of conservation biology and graduate to a more realistic example from petroleum exploration, motivated by data collected at a reservoir in the North Sea. Finally, we discuss our conclusions and directions for further research in Section 6.

2. Basic Notation

In this section we introduce notation and basic terminology. The word *field* is used to denote the spatial system under consideration. The field lies on a lateral 2-dimensional grid composed of $n_1 \times n_2$ *cells*, i.e. the grid has n_1 rows and n_2 columns with a total of N cells, where $N = n_1 n_2$. Each cell may have identical length and width but this is not essential; the size of the cell is determined by the application and the scale of the study. Besag (1974) refers to this kind of setting as a *lattice system*. We can index the cells in the grid from top to bottom and left to right as $i = 1, 2, \ldots N$. This will enable us to refer to a particular cell by using a single index.

Let $\boldsymbol{x}$ be the random variable over the entire field for the *distinction of interest*, or in other words, the *latent variable* which is of interest to the decision maker. Letters with a subscript i denote the outcome at cell number i, while letters without any subscript refer to the set of outcomes over the entire field. Thus $\boldsymbol{x} = \{x_i : i = 1, 2, \ldots N\}$. Furthermore, we assume that x_i is a categorical random variable that equals any one out of d possible *colors* (or *states*). When $d = 2$, x_i is a Bernoulli random variable. For this work, we will choose $x_i = 1$ as the favorable outcome at cell i. The joint *prior* probability of an outcome $\boldsymbol{x}$ is denoted by $p(\boldsymbol{x})$, while the marginal probability of an outcome x_i is denoted by $p_i(x_i)$. The marginal prior probability of the favorable outcome $x_i = 1$ is then $p_i(1)$.

Let $\boldsymbol{y}$ be the random variable for the result of an *experiment*, which may be conducted in the future to obtain more information about $\boldsymbol{x}$. The experiment is to be performed in cells with indices in the set J, a subset of $\{1, 2, \ldots N\}$. Therefore, $\boldsymbol{y} = \{y_j : j \in J\}$. The experiment need not be

performed at all the cells in the grid. In fact, it may be optimal to perform the experiment only at a few cells, depending on all the parameters and the cost of the experiment. The *likelihood* for the experimental result $\boldsymbol{y}$ given the outcome of the distinction of interest $\boldsymbol{x}$ is $p(\boldsymbol{y}|\boldsymbol{x})$. For a continuous random variable $\boldsymbol{y}$, $p(\boldsymbol{y}|\boldsymbol{x})$ is a density function; it is a probability function if $\boldsymbol{y}$ is a discrete random variable. We inspect both cases with the help of examples, but for presenting our methods we treat $\boldsymbol{y}$ as continuous. The marginal likelihood of the experimental outcome can be obtained by summing out all possible configurations of $\boldsymbol{x}$, i.e. $p(\boldsymbol{y}) = \sum_{\boldsymbol{x}} p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})$. For d-colored categorical fields $\boldsymbol{x}$ of size N, this is a summation over d^N terms and can be computationally demanding.

The *posterior* probability of the outcome of the distinction of interest $\boldsymbol{x}$ given an experimental result $\boldsymbol{y}$ is $p(\boldsymbol{x}|\boldsymbol{y})$. The *marginal posterior* probability of an outcome x_i at cell i given an experimental result $\boldsymbol{y}$ is written as $p_i(x_i|\boldsymbol{y})$. As the favorable outcome we use $x_i = 1$, which has marginal posterior probability $p_i(1|\boldsymbol{y})$. In this paper, we analyze the value of such an experiment $\boldsymbol{y}$ and present insights using examples motivated by different types of possible experiments.

3. Model Formulation

We now expand on the terminology and specify model assumptions regarding the probability distributions for $\boldsymbol{x}$ and $\boldsymbol{y}$. The conceptual equations that determine measures for the value of an experiment are also explained in this section. The actual computational issues are postponed until Section 4.

3.1 Prior spatial model for the distinction of interest

Spatial dependence in the distinction of interest $\boldsymbol{x}$ is incorporated through the use of a categorical first-order MRF formulation (Besag, 1974). This implies that the probability for a certain outcome in a given cell, given the outcome in the entire field, depends only on the outcome in the four neighboring cells. The spatial field is thus represented as an Ising model with β as the *interaction parameter*:

$$p(\boldsymbol{x}) = \frac{\exp[\beta \sum_{i\sim j} I(x_i = x_j) + \sum_{i=1}^{N} \alpha_i(x_i)]}{z}. \tag{1}$$

$I(A)$ is an indicator function taking value 1 if A is true and 0 otherwise. The first summation in equation (1) is over all pairs x_i and x_j that are closest neighbors in the grid, and z is a normalizing constant. The *point-wise prior function* $\alpha_i(x_i)$ is a function of the outcome at every cell. It provides the mechanism for including prior information about the outcome at a particular cell, based on expert opinion or previous data. On the other hand β controls the spatial dependence of the latent variable. We assume that β and $\alpha_i(x_i)$ are known a-priori. The special case when $\alpha_i(x_i) = 0$, $\forall x_i, i$ is known as the *uninformative prior* case. In this situation, the marginal probabilities are such that each color is equally likely at every cell.

3.2 Likelihood model for the experiment

The random variable $\boldsymbol{y}$ is the result of an experiment performed in cells with indices in the set J. The decision maker will purchase the experiment only if it is worthwhile to do so. Conducting

the experiment should increase the profitability of the decision by more than its cost. We are ultimately interested in some measures of the worth of this experiment.

We assume that y_j, the result for the experiment at cell j, given the outcome over the entire field $\boldsymbol{x}$, depends only on the marginal outcome x_j. The experimental result at a particular cell is hence conditionally independent of the outcome of the latent variable at other cells, given the outcome at its own cell. This is a reasonable assumption because several experiments and surveys in spatial decision making satisfy this property in practice by providing only local information about sites. From conditional independence, we get:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{j \in J} p(y_j|x_j). \tag{2}$$

Each local likelihood distribution $p(y_j|x_j)$ can be any probability distribution - discrete or continuous, depending on the kind of experiment being performed. We discuss the particular likelihood function used in the context of examples. We show results from a binary experiment (Section 5.1) and a Gaussian experiment (Section 5.2).

3.3 A simple decision-analytic framework

In this section we present a simple decision-analytic framework so that we can explain the concepts easily, and create a framework that is not completely guided by case-specific issues. We make some critical assumptions about the nature of the downstream decision:
i) We assume that the main decision specifically involves a one-time selection of cells from the field. Sequential decisions regarding cell-selection are not considered.
ii) The decision maker's goal is to obtain value from the individual cells. It is thus possible to alienate each cell and think about costs and revenues for each cell separately. In this way we are only concerned with the marginal probabilities of the favorable outcome at each cell. This may not be the case in practice - for instance, value from the main decision may involve joint (global) properties of the entire field, or it may depend on interactions between cells. Crucially, this assumption ensures that the value from the field is equal to the sum of the values from the cells.
iii) The decision for selection of cells is an unconstrained decision problem. The decision maker may choose as many cells as is profitable. The introduction of constraints would entail that the decision at a particular cell could not be made independently of those at other cells.
We present a more general formulation in Section 3.4, where we relax the second and third assumptions.

Say that the cost of selecting cell i is C_i and the revenue gained from observing the favorable outcome at that cell is R_i. No revenue is gained if any other outcome is observed. Note that the outcome for a cell is not ascertained until and unless the cell is selected. As an example, C_i may be the cost associated with drilling a well at cell i and R_i is the corresponding revenue obtained from discovering oil at that cell. We are now equipped with all the notation and assumptions necessary to calculate the value of information and other related measures for an experiment. Firstly, the decision maker can take his/her decision without purchasing the experiment. In that case the expected value from the i^{th} cell V_i is given by:

$$V_i = \max\{0, [R_i \cdot p_i(1) - C_i]\}, \tag{3}$$

where $p_i(1)$, as mentioned earlier, denotes the prior marginal probability of the favorable outcome $x_i = 1$. It is optimal for the risk neutral decision maker to choose the i^{th} cell only if the expected

profit is positive. Hence the value is the maximum of 0 (for the case where the cell is not selected) and the expected profit expression $R_i p_i(1) - C_i$ (for the case where the cell is selected). The favorable outcome is seen when $x_i = 1$ and all other outcomes result in zero revenue. With the assumptions we have described, the *prior value (PV)* or the total value from the field based on the prior alone is the sum of the expected value from the cells:

$$\text{PV} = \sum_{i=1}^{N} V_i = \sum_{i=1}^{N} \max\{0, [R_i \cdot p_i(1) - C_i]\}. \tag{4}$$

What if the decision maker had perfect information about the latent variable? In other words, what if a clairvoyant would be willing to reveal the outcome of the latent variable? How much should the decision maker pay for this information? The i^{th} cell has a favorable outcome with probability $p_i(1)$. If it is indeed favorable, the value obtained is the maximum of 0 (cell is not selected) and $R_i - C_i$ (cell is selected). The decision maker will choose not to select a cell if it is known that the outcome is not favorable, thereby making zero profit.

The *value with free clairvoyance (VFC)* (see e.g. Howard and Abbas, 2006) on the distinction of interest over the whole field is the sum of the expected value (with free clairvoyance on the distinction of interest) obtained at the individual cells. Hence,

$$\text{VFC} = \sum_{i=1}^{N} p_i(1) \cdot \max[0, (R_i - C_i)]. \tag{5}$$

In most situations, presumably the revenues outweigh the costs throughout the field. Therefore $R_i > C_i \; \forall \; i$. This implies that equation (5) can be reduced to $VFC = \sum_{i=1}^{N} p_i(1) \cdot (R_i - C_i)$.

The *value of perfect information (VOPI)* on the distinction of interest, which is the most that the decision maker should pay for perfect information on the distinction of interest, is the increase in profitability from the prior situation to the one where clairvoyance is obtained without cost. Note that this is only true for people who follow the *delta property* (Howard and Abbas, 2006). The delta property is a popular assumption in the decision analysis literature since it enables an analytically tractable method for calculating value of information. The property is satisfied by people who have an exponential or straight-line utility-curve, and since a risk-neutral decision maker has a straight-line utility-curve, s/he satisfies the delta property. Therefore,

$$\text{VOPI} = \text{VFC} - \text{PV}. \tag{6}$$

In both PV and VOPI calculations, the value depends only on the marginal probabilities of observing the favorable outcome since the value from a particular cell does not depend directly on other cells. We call the set created by these marginal probabilities over the grid, the *prior probability map*. The prior probability map is $\{p_i(1) : i = 1, 2, \ldots N\}$. If the experiment is performed, a result $\boldsymbol{y}$ is observed before the main decision is made. The marginal posterior probability of a favorable outcome $x_i = 1$ in the i^{th} cell is $p_i(1|\boldsymbol{y})$. The computation for the conditional value of the i^{th} cell is along the same lines as in equation (3), replacing the prior with the posterior. The expected value V_i' is evaluated with the expectation over the experimental result $\boldsymbol{y}$:

$$V_i' = \int_{\boldsymbol{y}} \max[0, R_i \cdot p_i(1|\boldsymbol{y}) - C_i] p(\boldsymbol{y}) d\boldsymbol{y}. \tag{7}$$

With the assumptions we have described, the *value with the free experiment (VFE)* is the sum of the expected value V_i' over all cells:

$$\text{VFE} = \sum_{i=1}^{N} V_i' = \sum_{i=1}^{N} \int_{\boldsymbol{y}} \max[0, R_i \cdot p_i(1|\boldsymbol{y}) - C_i] p(\boldsymbol{y}) d\boldsymbol{y}. \tag{8}$$

As in the case of clairvoyance on the distinction of interest for a delta person, the value of information for the experiment (or for short, the *value of the experiment (VOE)*) can be computed as the difference between value with the free experiment and value from the prior.

$$\text{VOE} = \text{VFE} - \text{PV}. \tag{9}$$

This is the gain in profitability from performing the experiment and hence this is the maximum that should be spent on purchasing the experiment. In our opinion the VOE is the best measure for the worth of an experiment because, by definition, it indicates how valuable the experiment is to the decision maker in monetary units. However poor the experiment is, one can always choose to ignore the results and end up being as well off as before. For a worthless experiment $VOE = 0$. Also, no matter how good the experiment is, it cannot be better than directly obtaining information about the latent variable since this is the variable that is of ultimate interest to the decision maker. In this way, the VOPI acts as an upper bound on the VOE.

This naturally leads to another measure of the experiment: the *chance of knowing (COK)* (Howard and Abbas, 2006). It can be shown that for a risk-neutral decision maker,

$$\text{COK} = \text{VOE}/\text{VOPI}, \quad 0 \leq \text{COK} \leq 1. \tag{10}$$

The chance of knowing can be explained as follows: Consider a lottery where with a probability p, a clairvoyant will provide perfect information on the distinction of interest for no charge and with probability $1 - p$ will provide no information. The probability p that makes the decision maker indifferent between obtaining the experiment for free and playing this lottery is the chance of knowing for that experiment. A good experiment would require a higher probability for a person to be indifferent, whereas for a poor experiment a smaller value of p would suffice. To summarize, the COK is a number between 0 and 1 that rates the worth of an experiment in a certain context.

VOE provides an actual monetary value on information, and COK presents a quick and intuitive way to compare different experiments with each other and also with information on the distinction of interest. In Section 4 we describe how the equations presented in this section can be solved.

3.4 A general decision-analytic framework

We will now show a more general framework where we relax two of the assumptions presented in Section 3.3. Rather than treating cells separately, the joint distribution of the distinction of interest is considered. Furthermore, the constrained case needs special attention since the action (where to select sites) depends on the outcome of the distinction of interest.

Let $f(a, \boldsymbol{x})$ be the value derived from a realization of the field $\boldsymbol{x}$ when action a is taken. a is one of the actions in the set of alternatives A, or $a \in A$. The prior value, before any information is revealed, is the value derived from the optimal course of action based on the prior on $\boldsymbol{x}$. It is

optimal for the risk-neutral decision maker to choose the action that maximizes the expected value. Thus,

$$\text{PV} = \max_{a \in A} \Big[\sum_{\boldsymbol{x}} f(a, \boldsymbol{x}) \cdot p(\boldsymbol{x}) \Big]. \tag{11}$$

For the case of perfect information on the distinction of interest, the realization $\boldsymbol{x}$ is known to the decision maker before the optimal action is chosen. The value of perfect information is given by the difference between the expected value with free clairvoyance on the distinction of interest and the prior value:

$$\text{VOPI} = \sum_{\boldsymbol{x}} \max_{a \in A} [f(a, \boldsymbol{x})] \cdot p(\boldsymbol{x}) - PV. \tag{12}$$

Note that the order of summation and maximization is reversed for the VOPI equation. As before, the value of the experiment is the difference between the expected value given result $\boldsymbol{y}$ and the prior value:

$$\text{VOE} = \int_{\boldsymbol{y}} \max_{a \in A} \Big[\sum_{\boldsymbol{x}} f(a, \boldsymbol{x}) \cdot p(\boldsymbol{x}|\boldsymbol{y}) \Big] p(\boldsymbol{y}) d\boldsymbol{y} - PV. \tag{13}$$

The COK can be obtained from equation (10).

To motivate how this general formulation might be preferred to our simplified version presented in Section 3.3, consider the petroleum example again. For a comprehensive study at a suitable scale, the area of each cell would not be large. Deviated wells could drain oil from adjacent cells and therefore the cost for an adjacent cell would be less if its neighbor is already selected. Value from the field would also depend on interaction factors such as permeability and the flow of oil in the reservoir. Another issue is that of constraints in selecting sites; the budget for a field puts a limit on the number of wells that can be drilled. All these aspects must be modeled intelligently in the spatial decision making context, because the price we pay for generality is a severe increase in computational intensity. For the simple framework presented in Section 3.3, the VOE calculation of equation (7) involves an integral over the experiment which is approximated and summed over all N cells. In addition to this calculation, the general framework also requires a summation over $\boldsymbol{x}$, as seen in equations (11) and (13). It is hard to provide general guidelines for the value function and the actions as they would remain case-specific.

3.5 Another criterion: Entropy

In this section we discuss entropy as another criterion for evaluating experiments. We do not provide numerical examples, but merely define the concept and compare it with our decision-analytic approach. The notion of entropy, introduced from *information theory*, see e.g. Ash (1965), has been used to measure the reduction in uncertainty of $\boldsymbol{x}$ on observing the outcome $\boldsymbol{y}$ of an experiment. The entropy is defined by

$$H(\boldsymbol{x}) = -\sum_{\boldsymbol{x}} p(\boldsymbol{x}) \log p(\boldsymbol{x}), \tag{14}$$

which can be constructed sequentially as

$$H(\boldsymbol{x}) = H(x_N) + H(x_{N-1}|x_N) + \ldots + H(x_1|x_2, \ldots, x_N). \tag{15}$$

This sequential formulation is used when computing entropy, see Appendix A. The expected reduction in entropy from the experiment $\boldsymbol{y}$ is

$$\text{EMI} = H(\boldsymbol{x}) - \int_{\boldsymbol{y}} H(\boldsymbol{x}|\boldsymbol{y})p(\boldsymbol{y})d\boldsymbol{y}, \tag{16}$$

where EMI is the *expected mutual information* between $\boldsymbol{x}$ and $\boldsymbol{y}$.

Both VOE and EMI can be used as measures for valuing information. However, there is a wide gulf between the philosophies of the two measures. For instance, EMI provides a sense of how much uncertainty can be reduced by performing an experiment, but it cannot directly imply how much the decision maker should pay for it. It is hardly surprising that the decision-analytic notion of VOE is tied inexorably with decisions and the preferences of the decision maker. VOE is a more complete measure for valuing information and also more difficult to obtain. According to Howard and Abbas (2006), an experiment should be conducted when it is: (i) *relevant* to the distinction of interest, (ii) *material* to the decision that brings value and (iii) *economic* for the decision maker. A material experiment is one that can affect the decision, i.e. the action chosen by the decision maker is not identical for different outcomes of the experiment. By the decision-analytic philosophy, information from experimentation may reduce uncertainty but is not *valuable* until it can change the decision. An economic experiment in our context is one that costs less than the VOE. Entropy based parameters such as EMI only address aspects of relevancy of the experiment, without addressing the other two requirements. An experiment with zero mutual information will be irrelevant for the distinction of interest. Mutual information measures may be used as a guide in designing the most relevant experiment (e.g. Mukerji et al, 2001). The three requirements of an experiment being relevant, material and economic are intertwined - for example, an immaterial experiment will not be economic because its value is 0. Ascertaining the value from a decision to be made in the future is often a difficult task; yet it is crucial if measures like VOE are to be estimated.

4. Computational Issues

The joint distributions for the distinction of interest and for the experimental result are over the entire field with N cells and are likely to be high dimensional. Solving equation (7) analytically may not be possible in general, so we use Monte Carlo simulation by generating realizations of the experimental result. The same holds for several of the expressions in Section 3. The Monte Carlo simulation is shortly described as follows: We generate M independent and identically distributed (i.i.d.) random samples $\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots, \boldsymbol{y}^M$ from $p(\boldsymbol{y})$. For the m^{th} sample, let $w(\boldsymbol{y}^m) = \max[0, R_i \cdot p_i(1|\boldsymbol{y}^m) - C_i]$. Now we can approximate equation (7) as

$$V_i' \approx \frac{1}{M}\sum_{m=1}^{M} w(\boldsymbol{y}^m). \tag{17}$$

An exposition on Monte Carlo methods can be found in Liu (2001). In general, a high dimensional integral of a function $w(\boldsymbol{y})$ over a region D, can be approximated as follows:

$$\int_D w(\boldsymbol{y})p(\boldsymbol{y})\,d\boldsymbol{y} \approx \frac{1}{M}\sum_{m=1}^{M} w(\boldsymbol{y}^m), \quad \boldsymbol{y}^m \sim p(\boldsymbol{y}), \quad m = 1, \ldots, M, \quad \text{i.i.d.} \tag{18}$$

The Monte Carlo error is not large if a sufficient number of realizations are generated.

In our simple framework from Section 3.3 we first generate i.i.d. realizations $\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^M$ of MRFs in equation (1). We next draw realizations of the experiment $\boldsymbol{y}^1, \boldsymbol{y}^2, \ldots, \boldsymbol{y}^M$. These are marginally from the distribution $p(\boldsymbol{y})$. For each of these conceptual datasets $\boldsymbol{y}^m$ we compute the marginal posterior for the favorable outcome denoted by $p_i(1|\boldsymbol{y}^m)$. We use the recursive forward and backward techniques to 1) draw realizations $\boldsymbol{x}^m$, and 2) compute the marginal $p_i(1|\boldsymbol{y}^m)$. We will refer to the two methods as *RecGenerate* and *RecCompute* respectively, to specify where exactly they are used in the main algorithm. The recursive method itself is based on Reeves and Pettitt (2004) and outlined in Appendix A.
The algorithm for computing the value of an experiment is as follows:

1. Find the marginal prior probability for $\boldsymbol{x}$ using *RecCompute*.
2. Solve equations (3) through (6) to find the PV and the VOPI.
3. Generate a realization $\boldsymbol{x}$ from the prior with *RecGenerate*. .
4. Generate a conditional realization $\boldsymbol{y}$ of the experimental result from the likelihood of the experiment, given the realization $\boldsymbol{x}$. The general form of the likelihood is depicted in equation (2).
5. With the current realization of the experiment $\boldsymbol{y}$, use *RecCompute* to evaluate the marginal posterior probability of the favorable outcome $x_i = 1$, denoted as $p_i(1|\boldsymbol{y})$, and compute the associated value. This is done for all cells $i = 1, \ldots, N$.
6. Repeat steps 3. - 5. a total of M times, and approximate the integral for V_i', $i = 1, \ldots, N$, in equation (7) with the average value from the simulations, shown in equation (17).
7. Solve equations (8) through (10) to get VOE and COK.

Crucial tasks in steps 1. - 7. of the algorithm are *RecGenerate* and *RecCompute*, using recursive computing on the field of size $n_1 n_2 = N$. The recursive method presented in Appendix A is of order $O(N)$, but for each step of the recursion we need to evaluate and store terms of size d^{n_1}; this is the computer memory intensive part of the algorithm. Therefore the smallest grid dimension n_1 should not be too large (say not more than $n_1 = 10$ for $d = 3$). In our case, with only one favorable outcome, the most efficient way is to use $d = 2$ and marginalize over all non-favorable categories. However, such a marginalization would not be natural from a modeling perspective. For example, in the petroleum exploration case, it is easier for experts to consider three categories: oil sand, brine sand and shale, and assign prior probabilities and likelihoods for each class. Also, more than one outcome can be favorable in general.

Note that the algorithm above only applies to the simple framework of Section 3.3. The computational advantage of this model is that the marginal posterior probability $p_i(1|\boldsymbol{y})$ of the favorable outcome is easily calculated by the recursive algorithm. For the more general setting of Section 3.4, one may require a function of the joint distribution over the entire field $p(\boldsymbol{x}|\boldsymbol{y})$, and this is not easy to obtain in general. A more complicated and time consuming Monte Carlo method would be necessary for the general case. The algorithm of choice would typically depend on the particular situation.

5. Examples

5.1 Example from conservation biology

Consider a region of land under scrutiny in a conservation biology project, modeled as a 3 by 3 grid. The decision maker is interested in selecting sites from the grid to set up natural wildlife reserves with the goal of conserving an endangered species. However, there is uncertainty regarding the presence of the species. If the decision maker selects a cell, she must pay the cost C (assumed to be the same across cells) for construction of the natural reserve. If the species is present at a cell that she chooses, she obtains revenue R (again assumed to be the same across cells). The decision maker will not know for sure whether the species is present or not until the cell has been selected. Let $\boldsymbol{x}$ be a random variable for the presence or absence ($d = 2$) of the species at all cells in the grid. The categorical outcome x_i is 1 if the species is present at cell i and 0 if the species is absent. The decision maker is interested in the value of information for different kinds of surveys.

In this example we analyze experiments that have binary results. We also assume for now that if the experiment is purchased, it will be performed at all cells in the grid. One can imagine a survey where a team explores every cell and indicates whether they believe the species is present or not, for each cell. The survey result $y_i = 1$ implies that the team believes the species is present at cell i and $y_i = 0$ suggests otherwise. The experimental results need not be accurate. Say that the accuracy of the test at every cell is the same and is denoted by γ. This accuracy is defined by the following likelihood equation:

$$p(y_i = k|x_i = k) = \gamma; k = 0, 1; \forall i. \tag{19}$$

The likelihood may be different conditioned on whether the latent variable x_i is 1 or 0. *Sensitivity* and *specificity* have been the terms used in the literature to denote these likelihoods. For simplicity we will combine these such that there is only one parameter. A value of γ close to 1 indicates a good test, i.e. a test with high accuracy. We choose to model the field's spatial prior as an uninformative prior, i.e. we have no prior point-wise information on the field, and believe that there is an equal chance of species presence or absence in every cell. There is some spatial correlation for $\boldsymbol{x}$, determined by the interaction parameter β. Furthermore, $C = 1$ monetary unit (so that revenue can now be written in units of C).

First we assume that the cell-selection is unconstrained, i.e. the decision maker can select as many cells as is profitable. Sensitivity analysis on the parameters β and γ can provide insights into general trends. Figure 1 shows VOE as a function of these parameters in three plots, for revenue $R = 2, 5$ and 10 monetary units from left to right respectively. Let us analyze trends within each plot to begin with. Firstly, the accuracy is a critical parameter. The curve for γ = 0.9 shows the highest VOE. In all three graphs, γ = 0.5 has $VOE = 0$ as the experiment provides no information about the distinction of interest. The VOPI is a horizontal line and has the same value (4.5 units in this case) for all β and for all $R \geq 2$. It only depends on the marginal probability of success in each cell, which is 0.5 in the case of the uninformative prior, and does not depend on β. Note that the VOE increases as β increases. This is because the chance that the entire grid will either contain the species at all cells or in no cells becomes higher as β increases. The experiment becomes more valuable as β increases because it can tell you about a possible jackpot (all cells favorable) or prevent a huge loss (all cells unfavorable). As there are no constraints on the number of cells

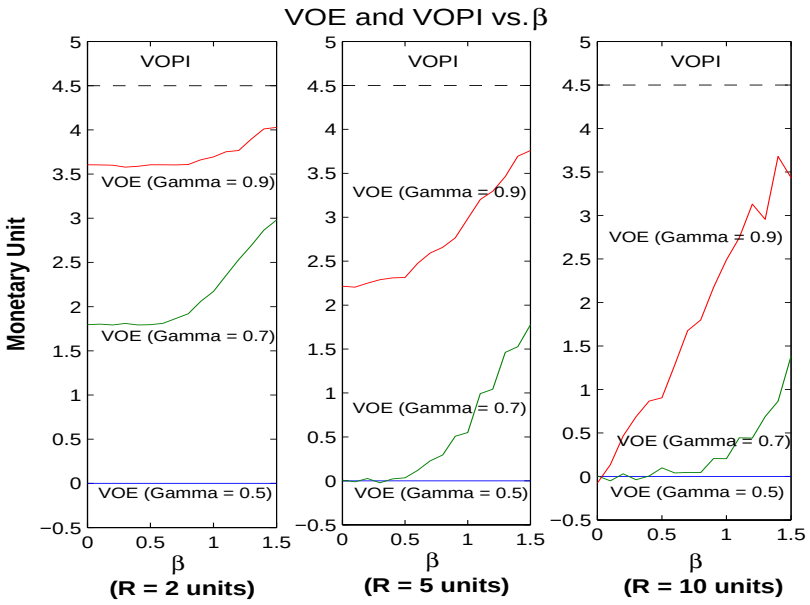

Figure 1: Sensitivity analysis in the unconstrained case. There are 3 plots of VOE and VOPI vs. β: (left) revenue $= 2$ units; (middle) revenue $= 5$ units; and (right) revenue $= 10$ units. Each plot shows VOPI (dashed line) and VOE for tests with accuracy $\gamma = 0.5, 0.7$ and 0.9 (solid, colored lines).

that can be selected, the decision maker is free to choose all the cells or none; for large values of β, this all-or-nothing policy is optimal. Therefore in the unconstrained case for the uninformative prior, the experiment can really make a difference for large β. The spikes in the curves are due to Monte Carlo error, and are most notable for the plot of $R = 10$. We use $M = 25000$ Monte Carlo simulations in this example. VOE is not very sensitive to β for lower values, which is useful to know if one is unsure about the interaction parameter or cannot spare much time for estimating it.

Now observe the differences between each of the plots in Figure 1. Consider the case of γ = 0.7 and $\beta = 0$ in the graphs for $R = 2$ and $R = 5$, and compare them. For this choice of γ and β, $VOE = 0$ for $R = 5$ whereas for $R = 2$ VOE is almost 2 units. Experiments do not automatically become more valuable when the decision situation is more lucrative. When $R = 2$, there is a chance for the experimental result to change the main decision, and in this way the experiment is valuable as it is able to affect the decision. On the other hand if $R = 5$ (and $\beta = 0$), it is worthwhile to select all cells no matter what the experiment has to say. This is a fundamental issue in the decision-analytic approach to valuing information.

What if there are constraints on the number of reserve sites that can be selected? We use ideas from both Section 3.3 and Section 3.4 to solve the equations for the situation of constraints in cell-selection. We retain the assumption that the cells act as separate units, while introducing another parameter: k, or the maximum number of cells that can be selected (based on a budget). Once we have a particular probability map (prior or posterior), we can maximize profits from the field by choosing the best k prospects, if profitable. Even in a simple example with only 9 cells, trends in VOE when the problem involves both spatial dependence and constraints in cell-selection can be

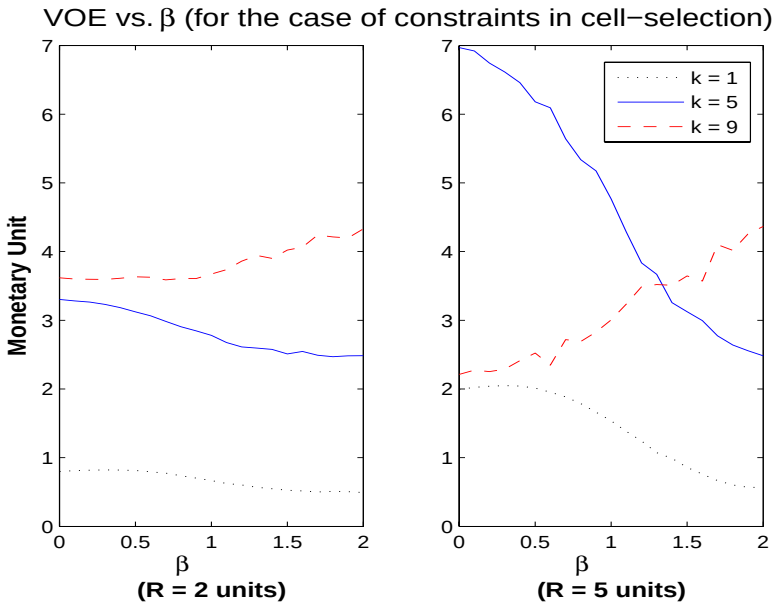

Figure 2: Sensitivity analysis in the constrained case. The accuracy of the experiment is $\gamma = 0.9$. Two plots of VOE vs. β: (left) revenue $= 2$ units; and (right) revenue $= 5$ units. Each plot shows VOE for 3 values of k (the maximum number of cells that can be selected): $k = 1$ (dotted line), $k = 5$ (solid line) and $k = 9$ (dashed line).

counter-intuitive. Figure 2 demonstrates results from sensitivity analysis on β for a test with an accuracy of $\gamma = 0.9$. We compare the VOE for $k = 1, 5$ and 9 for $R = 2$ (left) and $R = 5$ (right) in Figure 2.

The graph on the left is more in accordance with our initial reaction; VOE is highest when $k = 9$ and lowest when $k = 1$. The case with $k = 9$ corresponds to the unconstrained case as there are $N = 9$ cells in the field. As we observed in Figure 1, VOE appears to increase in the unconstrained situation. However, for $k = 5$ and to some extent also for $k = 1$, VOE seems to decrease as β increases. This is even more prominent in the graph on the right. We explain this tendency as follows. The experiment conducted over the entire field has a certain facet that can be relatively more valuable when there is both low spatial dependence and a limit on the number of cells that can be selected. It tells you *which* cells are likely to be favorable, thereby guiding the decision maker about the choice of site location. There is a little more leeway for the decision maker regarding selection of an appropriate location when there is high spatial dependence. Thus VOE can be relatively smaller when β is high. There are other seemingly unusual trends in the plot for $R = 5$. The VOE for curve $k = 5$ appears to be much higher than the other two curves for smaller β. It is indeed possible for the experiment to be more valuable in the situation when there are constraints, as opposed to when there are none. In the realistic situation of an informative prior on a large grid, the interaction between the parameters can become even trickier to understand. Simple examples can highlight some of the ideas that should be kept in mind while gathering information. High likelihood experiments are useful, but they should be able to affect the downstream decision.

5.2 Example from petroleum exploration

We next use an example from the petroleum industry to demonstrate how our formulation can be employed for real-world applications. It is not a detailed case study by itself, but can give a sense of how an actual case study would be performed. The case we are considering is a lateral two dimensional reservoir domain that is a candidate for oil exploration. Decisions have to be made regarding which data attributes to analyze and over which spatial area. More specifically we evaluate the value of different attributes obtained from seismic reflection data (Avseth et al, 2005). We are interested in analyzing the value of information for two attributes of amplitude versus offset (AVO) seismic data. If these attributes are purchased, the decision maker will have to pay to conduct AVO processing and analyses. The example is motivated by previous work on AVO seismic data from the Glitne field in the North Sea, for example Eidsvik et al. (2004) and Avseth et al. (2005). These studies predicted the spatial distribution of oil sands based on AVO seismic attributes. The results obtained using just one of the AVO attributes (called R_0, normal-incidence reflectivity) were quite different from using both (called (R_0, G), normal-incidence reflectivity and gradient with offset), and it is hence of interest to study the value of different seismic attributes following the decision-analytic framework outlined earlier. Furthermore, we compare the value of partial tests to that of acquiring AVO attributes along the entire field. We have chosen to postpone details regarding the model and implementation of this rather complex seismic example to Appendix B.

The focus of attention is a $n_1 = 5$ by $n_2 = 20$ grid ($N = 100$); a part of a reservoir that may be the lobe of a turbidite system and hence of main interest for exploration. The distinction of interest is the rock and fluid composition which takes on any one of $d = 3$ colors: oil sand ($x_i = 1$), brine sand ($x_i = 2$) or shale ($x_i = 3$). The only profitable outcome is that of oil sand. For the MRF prior model in equation (1) we use $\alpha_i(x_i) \neq 0$, representing an informed prior based on previous knowledge and expert geologic opinion. The experimental result y_j, $j \in J$ is a continuously distributed random variable. With an experiment that measures only one AVO attribute, R_0, the measurement at site j is $y_j = R_{0,j} \in \mathcal{R}$, whereas a situation where both AVO attributes are acquired is indicated by $\boldsymbol{y}_j = (R_{0,j}, G_j)' \in \mathcal{R}^2$. The experimental results are assumed to be Gaussian distributed with likelihoods given in Appendix B.

Figure 3 shows the AVO seismic data from a 5×20 grid from the North Sea (Eidsvik et al, 2004). This data is somewhat upscaled from the original dataset and each gridcell is of size $100^2\, m^2$. The domain in Figure 3 is hence of size 500m $\times$ 2000m. For parameter estimation of interaction parameter β we assume that this field is similar to the one under consideration and use the data for this purpose. We estimate β in equation (1) based on these AVO seismic data and fixed values of $\alpha_i(x_i)$. The maximum likelihood estimate of β is computed by evaluating the marginal likelihood $p(\boldsymbol{y})$ for a set of β values, see Appendix A. In Figure 4 we display $p(\boldsymbol{y})$ as a function of β, and see that the maximum likelihood estimate is about 0.9. In the following calculations we treat this parameter as fixed.

Now recall that we do not have the extracted AVO attributes for the reservoir. We will compute the value of information for some experimental configurations to identify which attribute(s) to extract. We do, however, suppose that geologic knowledge is present. We assess a realization of the field from a hypothetical geologist, along with the confidence in the assessment using a likelihood function at all cells. We obtain the point-wise function $\alpha_i(x_i)$ from geologic expert opinion. This knowledge assigns high $\alpha_i(1)$ for central cells and the South-East flank, larger $\alpha_i(2)$

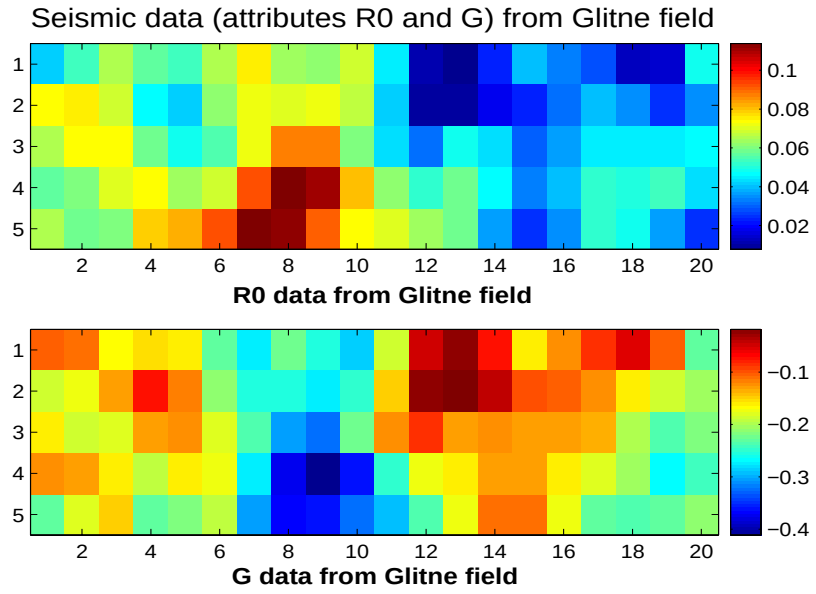

Figure 3: Seismic data from the Glitne reservoir in the North Sea. The display shows two seismic attributes R_0 (top) and G (below). Each grid cell is 100m $\times$ 100m and the domain covers an area of 500m $\times$ 2000m.

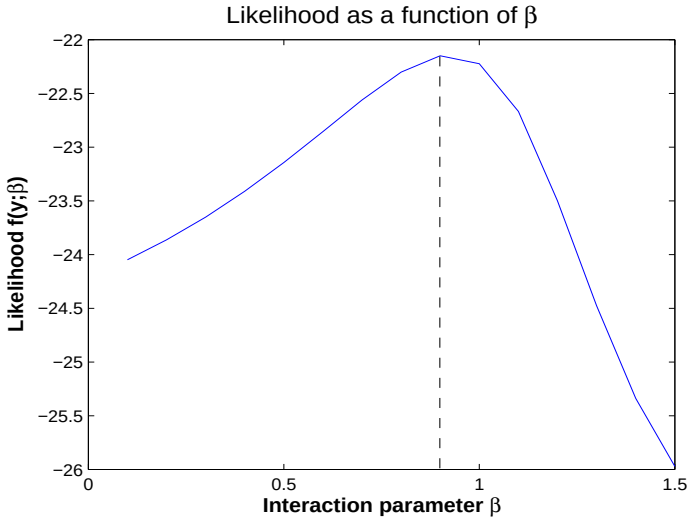

Figure 4: Maximum likelihood estimation of β. The graph is obtained by evaluating the marginal likelihood $p(\boldsymbol{y})$ using the Glitne dataset for several β values between 0 and 1.5. The maximum is near 0.9.

and $\alpha_i(3)$ in other areas. In Figure 5 we show the marginal prior probabilities of the favorable oil sand.

We demonstrate the value of information for the full test, i.e. extracting seismic attributes at every grid cell $i = 1, \ldots, N$, and for two partial tests covering only parts of the domain. The

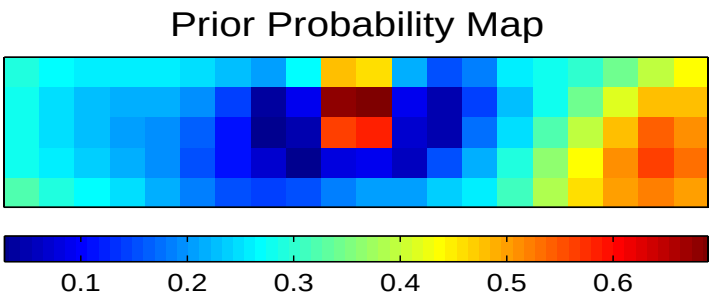

Figure 5: Prior probability of oil sand. The image shows $p_i(1)$, $i = 1, \ldots, 100$. Note that oil sand is more likely in the central parts and in the South-East where the pointwise prior terms $\alpha_i(1)$ are larger.

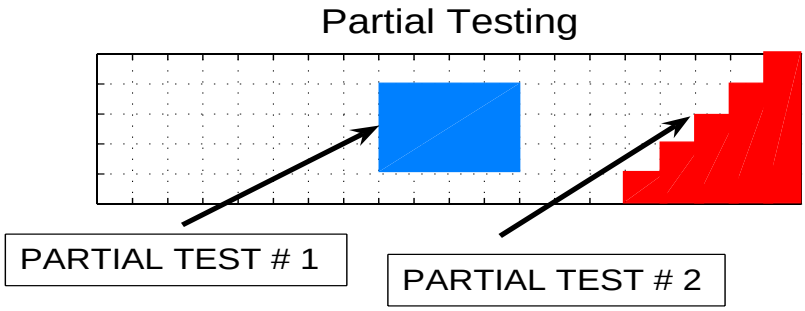

Figure 6: Cell locations for two kinds of partial tests. Partial test 1 covers 12 cells in the central part (shaded blue), while partial test 2 covers 15 cells in the South-East (shaded red). In comparison the full test covers all 100 cells.

two partial tests are shown in Figure 6. Partial test 1 covers the central parts of the domain, while partial test 2 mostly covers the south-eastern parts. These two spatial domains are believed to be the most likely candidate areas for oil, as specified by the informative prior. For each of the tests we consider using only the R_0 seismic attribute, or both (R_0, G) seismic attributes. Altogether this entails six testing configurations.

For evaluating the decision-analytic expressions we use the algorithm at the end of Section 4. We briefly illustrate how we compute the VFE for this example. First we generate $M = 10000$ realizations of the MRF $\boldsymbol{x}$, two of these realizations are illustrated in Figure 7 (top). Next, test results are generated for each realization of the MRF. For each of the six configurations the test result will be generated from different likelihood functions depending on the particular spatial design and on the seismic attributes. In Figure 7 (middle) we illustrate the test results of R_0 for a full test conditional on the MRF realizations in Figure 7 (top). Finally, based on the test results, we compute the marginal posterior probability of the favorable state oil at each gridcell. These marginals are used when computing the VFE in equation (17). In Figure 7 (bottom) we show the marginal probabilities of oil for each of the two test results in Figure 7 (middle). Note that the

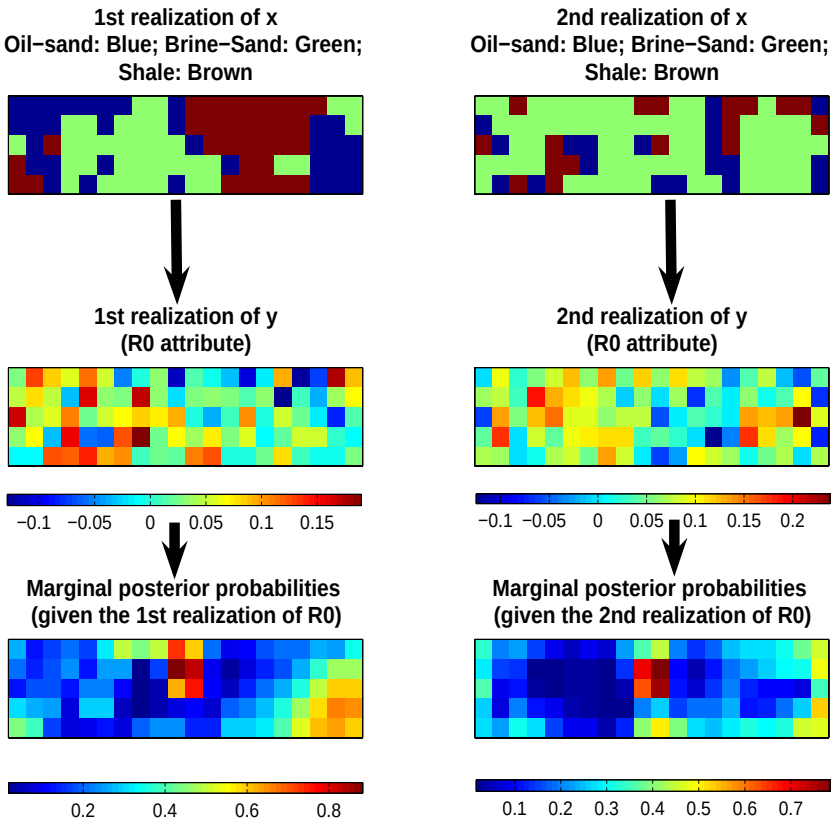

Figure 7: Realizations. Top: Two realizations of the field $\boldsymbol{x}$ from the prior $p(\boldsymbol{x})$. Middle: Two associated realizations of data $\boldsymbol{y}$ which in this case is the zero-offset attribute R_0. Below: The marginal posterior probabilities $p_i(1|\boldsymbol{y})$, $i = 1, \ldots, 100$, of oil sand corresponding to each dataset $\boldsymbol{y}$. For computing value of information we generate $M = 10000$ realizations of data and compute the marginals for each.

marginal probability of oil sand in Figure 7 (left, bottom) is small in the North-West, even though the realization Figure 7 (left, top) shows oil sand in that part. This occurs since oil sand and shales are almost indistinguishable based on R_0 data alone (see Appendix B), and because the shale in this region of the field is more likely a priori.

We assume a cost of $C = \$2$ million for drilling a well at a cell, and a revenue of $R = \$5$ million if a well discovers oil sand at a cell. Note that these are costs and revenues per cell and are obtained from back-of-the-envelope calculations. They would be obtained from management and petroleum engineers in a real-world case study. We use the volume of a cell, the assumed porosity and fraction of recoverable oil and the price of oil to estimate the revenue.

Results of the tests and attribute selection schemes are displayed in Table 1. The specific course of action regarding experimentation depends on the cost of each configuration, which would depend on factors such as whether the reservoir is offshore or not, etc. It seems likely that in this case, purchasing both attributes over the entire field may be most beneficial. The VOE and COK values are high in our example, particularly when both the attributes (R_0, G) are acquired over the

Table 1: Value of experiment (VOE) and chance of knowing (COK) for seismic AVO attributes R_0 and (R_0, G) and for different experimental configurations; complete test, partial tests 1 and 2.

	Complete test		Partial test 1		Partial test 2	
	VOE (million \$)	COK	VOE (million \$)	COK	VOE (million \$)	COK
R_o	4.91	0.07	0.48	0.007	2.11	0.03
(R_0, G)	35.94	0.51	1.93	0.03	10.61	0.15

entire field. This is because decisions regarding several cells in the field can be strongly affected by the experimental result. The field itself seems to be very lucrative, and has a prior value of around $\$12$ million. We would expect COK values to be much lower in practice because the magnitude of revenues and costs typically result in a very high VOPI in the petroleum industry. In our example, VOPI is around $\$70$ million. Depending on the costs involved, partial test 2 may be preferred over partial test 1 - the numbers indicate that intelligent experimental positioning can be extremely valuable. Also, again depending on the costs, a partial test with both attributes can be better than a complete test with only one attribute. The decision-analytic approach naturally encourages creative alternatives for decisions related to experimentation. There are often several alternatives regarding the kind and specific location of the experiment, and these can be discussed at length by the stakeholders.

6. Conclusions and Future Research

In this paper, we propose a decision-analytic approach to valuing experiments performed in situations that naturally exhibit spatial dependence. We incorporate dependence by modeling the system as a two dimensional grid, and by treating the joint prior distribution of the distinction of interest as a Markov random field. The experimental decision is based on the value of information in our framework. We illustrate our methodology with the help of two examples. Our example from conservation biology indicates that spatial dependence can play a significant role in determining the value of an experiment. The petroleum exploration example is a demonstration of how our framework can be applied in real-world case studies. The results suggest that intelligent experimental design can add substantial value to the decision situation. Moreover, decisions regarding the choice and location of experiments should be analyzed by a multi-disciplinary taskforce. Our methodology inspires a collaborative effort by aggregating the experimental likelihood, prior spatial model and downstream decisions and associated values into a single meta-model.

Decision trees and similar tools have previously been used to compute the monetary value of information in spatial decision making. These techniques aggregate distinctions to a global level and therefore excessively simplify several important aspects of the decision problem, including the alternatives, experimental likelihood and value assessments. Employing a formulation with a grid incorporates spatial dependence, maneouvres the spatial decision making problem to a more realistic front, and encourages creative alternatives for experimental design.

There are several possible directions for further research. One direction involves statistical issues. For instance, other prior models like Gaussian models can be investigated. Hierarchical Bayesian models could also be captured within our framework, such as a prior on β, but we have

not pursued this idea here. From the computational point of view, algorithms for larger grid sizes should be explored. In our framework the recursive method works best for moderate size grids, say $n_1 < 10$. Block updating using Gibbs sampling is a possible way to estimate value of information for larger grids. Other simulation techniques could also be considered; specific techniques may possibly be suitable for specific models. Finally, the decision-analytic assumptions can be relaxed or modified in future work. We currently focus on the marginal probabilities in our simple framework. In general, the value from the field can depend on the joint distribution of the distinction of interest. Sequential decisions and even sequential experimentation are other avenues that can be examined. Decisions regarding experimentation should be subjected to intensive analysis in all domains, particularly in realms where there is the additional complexity of spatial dependence.

Acknowledgements

This report is based on work done at NTNU, Trondheim, Norway, from June to Sept. 2006, in collaboration with Prof. John Weyant (GCEP; Management Science and Engineering, Stanford University) and the Rock Physics Laboratory, Stanford University. We had enlightening early discussions with Prof. Ross Shachter and Thomas Seyller (Management Science and Engineering, Stanford University) and Joshua Goldstein (Interdisciplinary Program in Environment and Resources, Stanford University).

References

Ash, R. B., 1965, Information theory, Dover Publ. Inc.

Avseth, P., Mukerji, T., and Mavko, G., 2005, Quantitative seismic interpretation, Cambridge University Press.

Besag, J., 1974, Spatial interaction and the statistical analysis of lattice systems, Journal of the Royal Statistical Society, Series B, **36**, 192-225.

Diggle, P., and Lophaven, S., 2006, Bayesian geostatistical design, Scandinavian Journal of Statistics, **33**, 53-64.

Eidsvik, J., Avseth, P., Omre, H., Mukerji, T., and Mavko, G., 2004, Stochastic reservoir characterization using prestack seismic data, Geophysics, **69**, 978-993.

Houck, R. T., and Pavlov, D. A., 2006, Evaluating reconnaissance CSEM survey designs using detection theory, The Leading Edge, **25**, 994-1004.

Howard, R., 1966, Information value theory, IEEE Transactions on Systems Science and Cybernetics, **2(1)**, 22-26.

Howard, R., and Abbas, A., 2006, The foundations of decision analysis (tentative title), Manuscript in Progress.

Liu, J. S., 2001, Monte Carlo strategies in scientific computing, Springer-Verlag New York, Inc.

Matheson, J. E., 1990, Using influence diagrams to value information and control, In R. M. Oliver and J. Q. Smith, editors, Influence diagrams, belief nets and decision analysis, 25-48, John Wiley and Sons, New York.

Mukerji, T., Avseth, P., Mavko, G., Takahashi, I., and Gonzalez, E., 2001, Statistical rock physics; combining rock physics, information theory, and geostatistics to reduce uncertainty in seismic reservoir characterization, The Leading Edge, **20**, 313-319.

Müller, W. G., 2001, Collecting spatial data, Springer-Verlag, Heidelberg.
Polasky, A., and Solow, A. R., 2001, The value of information in reserve site selection, Biodiversity and Conservation, **10**, 1051-1058.
Raiffa, H., 1968, Decision analysis, Addison-Wesley.
Reeves, R., and Pettitt, A. N., 2004, Efficient recursions for general factorisable models, Biometrika, **91**, 751-757.

Appendix A: Recursive computations for Markov random fields (MRFs).

Let $\boldsymbol{x} = (x_1, \ldots, x_N)$ be a categorical MRF on the two dimensional regular grid. Here, $N = n_1 n_2$ is the total number of grid nodes, n_1 is the shorter direction (say North) and n_2 is the longer direction (say East). Assume that each $x_i \in \{1, \ldots, d\}$, $i = 1, \ldots, N$, i.e. the categorical values can take d possible colors. Suppose the grid nodes are numbered sequentially from North-West so that x_1 is the categorical value in the North-West, x_{n_1} in the South-West, $x_{n_1(n_2-1)+1}$ in the North-East, and x_N in the South-East. See Figure 8 for an example of a grid and the indexing of cells.

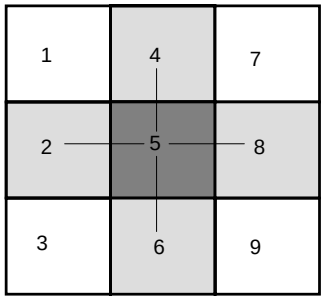

Figure 8: Illustration of a 3×3 grid and the indexing of cells. For an Ising model the full conditional probabilities at cell 5 depend only on the outcomes at the four nearest neighbors (shaded colors).

The MRF probability function is written as

$$p(\boldsymbol{x}) = \frac{\exp[\beta \sum_{i \sim j} I(x_i = x_j) + \sum_i \alpha_i(x_i)]}{z} = \frac{h(\boldsymbol{x})}{z}, \tag{20}$$

where $i \sim j$ means all neighboring pairs, β and $\alpha_i(l)$, $i = 1, \ldots, N$, $l = 1, \ldots, d$ are model parameters. This is the simplest MRF and is known as the Ising model, where the neighbors of an interior node i are defined by $i + 1$, $i - 1$, $i - n_1$, and $i + n_1$, see Figure 8. If node i is on the edge or is a corner node, some of these neighbors vanish. The normalizing constant z is given by

$$z = \sum_{x_1=1}^{d} \ldots \sum_{x_n=1}^{d} h(\boldsymbol{x}). \tag{21}$$

Note that this normalizing constant is a function of the model parameters. Location specific parameters $\alpha_i(l)$ can be functions of data y_i via the likelihood term, i.e. $\alpha_i(x_i) = \log p(y_i|x_i)$,

$x_i \in \{1, \ldots, d\}$. The density in equation (20) is then a posterior $p(\boldsymbol{x}|\boldsymbol{y})$, and the normalizing constant depends both on parameters and the data. The marginal likelihood of data is given by

$$p(\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{x}|\boldsymbol{y})} = \frac{\prod_j p(y_j|x_j)h^1(\boldsymbol{x})/z^1}{h^2(\boldsymbol{x})/z^2} = \frac{z^2}{z^1}, \tag{22}$$

where $h^1(\boldsymbol{x})$ and z^1 are defined by $\alpha_i(x_i)$ terms including only prior information, while $h^2(\boldsymbol{x})$ and z^2 are defined by both prior information and the $\log p(y_i|x_i)$ likelihood terms. Hence, the functional expressions depending on $\boldsymbol{x}$ cancel and the marginal likelihood equals the ratio of the normalizing constants in posterior and prior. For parameter estimation of β we evaluate the marginal likelihood for a set of parameter values and find the maximum likelihood over this set.

The probability function for $\boldsymbol{x}$ in equation (20) can be written sequentially as:

$$\begin{aligned} p(\boldsymbol{x}) &= p(x_1|x_2, \ldots, x_n)p(x_2|x_3, \ldots, x_N) \ldots p(x_{N-1}|x_N)p(x_N) \\ &= p(x_1|x_2, \ldots, x_{1+n_1})p(x_2|x_3, \ldots, x_{2+n_1}) \ldots p(x_{N-1}|x_N)p(x_N) \\ &= \frac{h(x_1|x_2, \ldots, x_{1+n_1})h(x_2|x_3, \ldots, x_{2+n_1}) \ldots h(x_{N-1}|x_N)h(x_N)}{z}, \end{aligned} \tag{23}$$

where we use the Markov property, but choose to condition on all buffer variables in the sequential line up. The buffer is of length n_1 in this case. It gets shorter in the last (easternmost) column. The terms in expression (23) are defined by $h(x_1|x_2, \ldots, x_{1+n_1}) = \exp\{\beta[I(x_1 = x_2) + I(x_1 = x_{1+n_1})] + \alpha_1(x_1)\}$ for the first location, then goes on for x_2, x_3 and so on, until $h(x_{N-1}|x_N) = \exp\{\beta I(x_{N-1} = x_N) + \alpha_{N-1}(x_{N-1})\}$ and $h(x_N) = \exp\{\alpha_N(x_N)\}$.

We first illustrate a method for recursive *forward* computation of the normalizing constant z in equation (21). This method follows Reeves and Pettitt (2004), and z is computed by summing out one variable at a time. The recursion starts by

$$z_1(x_2, \ldots, x_{n_1+1}) = \sum_{x_1=1}^{d} h(x_1|x_2, \ldots, x_{1+n_1}), \tag{24}$$

since x_1 is only involved in the first term of the sequential formulation in equation (23). The recursive calculation continues by using the equation that for general $i \leq N - n_1$

$$z_i(x_{i+1}, \ldots, x_{i+n_1}) = \sum_{x_i=1}^{d} h(x_i|x_{i+1}, \ldots, x_{i+n_1})z_{i-1}(x_i, \ldots, x_{i+n_1-1}). \tag{25}$$

The terms in $(x_{i+1}, \ldots, x_{i+n_1})$ takes one value for every buffer configuration, and with d colors we have d^{n_1} possible configurations. As the buffer length gradually decreases in the last column, the number of possible configurations gets smaller, and at the final step N we calculate

$$z = z_N = \sum_{x_N=1}^{d} h(x_N)z_{N-1}(x_N), \tag{26}$$

which is the normalizing constant in equation (21).

We next demonstrate a recursive *backward* sampling algorithm for drawing $\boldsymbol{x}$ from the probability function in equation (20). The value of x_N is sampled from probability vector

$$\begin{aligned} p_N(x_N) &= \frac{1}{z}\sum_{x_1=1}^{d}\cdots\sum_{x_{N-1}=1}^{d} h(\boldsymbol{x}) \\ &= \frac{1}{z}h(x_N)z_{N-1}(x_N), \quad x_N \in \{1,\ldots,d\}. \end{aligned} \tag{27}$$

where the sequential normalizing constants evaluated in equation (25) and (26) are used. We continue in this manner; generating x_{N-1} conditional on the sample of x_N from probability vector

$$\begin{aligned} p(x_{N-1}|x_N) &= \frac{\sum_{x_1=1}^{d}\cdots\sum_{x_{N-2}=1}^{d} h(\boldsymbol{x})}{zp(x_N)} \\ &= \frac{h(x_{N-1}|x_N)z_{N-2}(x_N,x_{N-1})}{z_{N-1}(x_N)}, \quad x_{N-1} \in \{1,\ldots,d\}, \end{aligned} \tag{28}$$

and so on for all $i = N-2,\ldots,1$.

We finally present a *backward* evaluation scheme for the marginal probabilities $p_i(x_i)$, $i = N,\ldots,1$. The marginal for x_N is given directly in equation (27). For $N-1$ we first arrange the joint density $p(x_{N-1},x_N)$ and then sum out x_N:

$$\begin{aligned} p_{N-1}(x_{N-1}) &= \frac{\sum_{x_1=1}^{d}\cdots\sum_{x_{N-2}=1}^{d}\sum_{x_N=1}^{d} h(\boldsymbol{x})}{z} \\ &= \frac{\sum_{x_N=1}^{d} h(x_N)h(x_{N-1}|x_N)z_{N-2}(x_N,x_{N-1})}{z}, \quad x_{N-1} \in \{1,\ldots,d\}. \end{aligned} \tag{29}$$

For general node i this evaluation consists of a backward construction of the joint probability for the buffer of length n_1. The marginal for node i is then evaluated by summing over all buffer configurations $x_{i+1},\ldots,x_{i+n_1}$ for each possible value of $x_i \in \{1,\ldots,d\}$. A similar backward formula is used when computing the *entropy* defined by the sequential formula in equation (15).

Appendix B: Implementation issues for the petroleum exploration example

The two amplitude versus offset (AVO) seismic attributes that are used here are denoted as R_0 and G. The first attribute relates to the reflectivity of a seismic wave at zero angle, while the latter is associated with the reflectivity of a seismic wave as a function of angle (offset angle between emitted and received signal). If one acquires only R_0, there is no information about the variation of the reflected response as a function of angle. The seismic reflectivity is connected to the rock and fluid composition of the subsurface. These relationships are quite well known in rock physics, see e.g. Avseth et al. (2005), and are typically modeled as a hierarchy of physical attributes such as pressure and shear wave velocity and density.

In Eidsvik et al. (2004) statistical relationships for each level of the hierarchy of physical attributes was specified. We use these statistical relationships in a Monte Carlo setting to fit a Gaussian likelihood model to the AVO seismic data, conditional on oil sand, brine sand and shale. For the case with only R_0 the likelihood equals

$$p(y_j|x_j) = \text{Normal}[\mu_1(x_j), 0.06^2], \quad y_j = R_{0,j}, \quad j \in J, \tag{30}$$

where $\boldsymbol{\mu}_1 = (0.03, 0.08, 0.02)$. For the case with both attributes R_0 and G, the likelihood equals

$$p(\boldsymbol{y}_j|x_j) = \text{Normal}\left\{[\mu_1(x_j), \mu_2(x_j)]', \begin{bmatrix} 0.06^2 & -0.007 \\ -0.007 & 0.17^2 \end{bmatrix}\right\} \quad \boldsymbol{y}_j = (R_{0,j}, G_j)', \quad j \in J, \quad (31)$$

where $\boldsymbol{\mu}_2 = (-0.21, -0.15, 0)$. The off-diagonal entry in the covariance matrix in equation (31) corresponds to a correlation between R_0 and G of about -0.7. In these equations we use $x_j \in \{1, 2, 3\}$; 1 corresponds to oil sand, 2 is brine sand, and 3 is shale.