

Scenarios for valuing sample information in natural resources

Byron K. Williams  | Eleanor D. Brown 

Science and Decisions Center, U.S.
Geological Survey, Reston, VA, USA

Correspondence

Eleanor D. Brown

Email: ebrown@usgs.gov

Funding information

U.S. Geological Survey, Grant/Award
Number: G19AC00146

Handling Editor: Robert B. O'Hara

Abstract

1. Uncertainty is ubiquitous in natural resource systems, science and management. Sample data are obtained in order to reduce uncertainty, thereby increasing knowledge and improving resource management, but sampling always comes at a cost of some sort. Is that cost worthwhile? Analysis of the value of sample information (VSI) addresses this question.
2. In this paper we develop the valuation of sample information in terms of five elements: (a) a *system* whose attributes are the focus of analysis; (b) a range of management *actions* that affect the system's status; (c) *uncertainty* about system status or structure, as characterized by initial (prior) probabilities of possible system states or structures; (d) an *experiment* or other information source that produces new data points and updated (posterior) probabilities; and (e) a *value measure* that is a function of the management action taken, conditional on either the system state or structure.
3. We describe five scenarios for analysing the VSI under uncertainty about system structure and state. Scenarios 1–3 comprise analyses of conditional, expected and optimal expected values of sample information. They focus primarily on choice of management adaptations with new information. Scenarios 4 and 5 involve pre-selected management actions, and are useful for comparing designs of data collection rather than for choosing a management action. These last scenarios expand the framework for VSI to include actions that have been selected independently of the updating of uncertainty.
4. We discuss other extensions of VSI analysis, which include spatial applications, hybrid scenarios, applications involving dynamic systems, and a focus on costs rather than net benefits.
5. Value of sample information analysis holds promise in emerging areas of ecology such as ecological forecasting and the use of remote sensing in conservation, where potential new data from models and satellites can be evaluated in advance, thereby allowing more efficient prioritization of scientific efforts. More generally, VSI can contribute to better ecological understanding and more effective management in a wide range of ecological situations.

KEYWORDS

Bayesian methods, experimentation, prior and posterior probabilities, statistics, uncertainty, value of information

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Published 2020. This article is a U.S. Government work and is in the public domain in the USA. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society

1 | INTRODUCTION

Analysis of the value of information can be an important aid in determining whether further collection of data is warranted (Canessa et al., 2015), given the time and financial resources required. Costs can be contained by well-designed data collection efforts and field expertise. Benefits can be enhanced by a clear understanding of how the data will aid management decisions. Nevertheless, the issue of potential net benefit from new information remains. This issue is captured in the concept of value of information.

The phrase 'value of information' has been around in one form or another since the early 1960s. In their landmark book Raiffa and Schlaifer (1961) provided one of the first treatments of the value of information, coining the name and developing many of its key expressions. Analysis of the value of information occurs frequently in medicine, economics, and other disciplines (Keisler et al., 2014), and its use is growing in conservation decision-making (Bolam et al., 2019). Several variants are now well recognized (Yokota & Thompson, 2004), including both analyses of data that are already known, and prospective analyses of the potential worth of new information.

Most treatments of information value in the environmental or ecological literature concern the expected value of perfect information, rather than the value of sample information (see, e.g. Bolam et al., 2019; Keisler et al., 2014). The value of perfect information is useful in providing an upper bound on the value of information (e.g. Moore et al., 2011). However, perfect information (or even perfect partial information) about ecological systems is never achieved, due to the ubiquitous presence of uncertainty about structure, processes and status of a resource system, no matter how extensive the data collection. In this situation, valuation of sample (or imperfect (Clemen & Reilly, 2014)) information has great potential for assessing the value of actual data.

In this paper, we focus on the value of sample information, with approaches that range from simpler to more complex. We first describe different forms of the value of information, in terms of perfect information, partial information and sample information. We then present a general framework for the value of sample information, discuss key components and variants of the valuation process, and offer some potential extensions.

2 | MEASURES OF THE VALUE OF INFORMATION

The value of information measures the potential for new information (monitoring, research, etc.) to improve understanding of a resource system and thereby change its perceived value. Information value is essentially a comparison of resource value before and after accounting for new information (Howard & Abbas, 2015). Several forms are well established in the literature (Yokota & Thompson, 2004).

The *expected value of perfect information* represents the value added by the elimination of uncertainty. It measures the difference

between an average of optimal values over their likelihoods of occurrence, and optimal value in the presence of uncertainty. The expected value of perfect information is the most prevalent form of the value of information in the environmental literature (Bolam et al., 2019).

The *expected value of partial perfect information* concerns the value added by information that eliminates one of the several sources of uncertainty. There are some examples of the expected value of partial perfect information in the ecological and environmental literature, but not many (e.g. Davis et al., 2019; Maxwell et al., 2015; Moore & Runge, 2012; Williams & Johnson, 2015a).

The *expected value of sample information* expresses the potential gain in value resulting from the collection of less-than-perfect information. It is a comparison of the optimal values expected with and without additional information. Though analyses of sample information are fairly common in medical science (Ades et al., 2004), to date there has been relatively little use in the environmental field.

Our emphasis in this paper is on the value of sample information (VSI). Key valuation elements include:

1. a *resource system* characterized by attributes of either the state of the system or its structure and processes. Examples of system attributes include population size or density, population vital rates, spatial distribution, biodiversity and habitat features;
2. *actions* that affect the system's status or condition. Examples include selection of hunting limits, introduction or removal of species, habitat manipulation, contaminant clean-up, adaptations to climate change and regulatory actions;
3. a *value measure* by which to assign value to the outcomes of actions. For example, value could be measured in terms of population survival rate, number of animals, increase in biodiversity, risk abatement, economic profit and opportunity cost;
4. *uncertainty* about resource status or structure, as characterized by the likelihoods of occurrence of possible system states or system structures and processes. Likelihoods can be identified by statistical assessment, elicitation from experts and other means (Johnson et al., 2017; Runge et al., 2011). Before being updated with new data, likelihoods express uncertainty as *prior probabilities* in a *prior distribution*;
5. an *'experiment'* that produces data to update uncertainty and improve understanding of the resource. We use the term 'experiment' broadly to include any source of new information. Experimentation typically produces a new probability distribution for potential data points, which can be combined with the prior distribution to produce *posterior probabilities* in a *posterior distribution*.

We note that if there is more than one experiment under consideration, decision-making will include not only the choice of management action *a*, but also the choice of an experiment *e*. The addition of experimental alternatives factors into the calculation of VSI (see Section 5.3 below).

In what follows, we assume initially that potential states, actions and hypotheses are finite in number. Later we will discuss the challenges posed by continuous variation and approaches to deal with it. We use the following notation:

- resource status is signified by x , with $\{x_1, \dots, x_N\}$ the set of possible resource states;
- a hypothesis about resource structure and function is indexed by k for a set of K alternative hypotheses;
- management action is denoted by a , with $\{a_1, \dots, a_M\}$ the set of actions that could potentially be taken.

3 | UNCERTAINTY

Value of sample information builds on the presence of uncertainty, as expressed by a distribution of prior probabilities, and the potential for its reduction with new information via posterior updating.

3.1 | Sources of uncertainty

Two sources of uncertainty are considered here: uncertainty about the system's structure or processes (e.g. additive vs. compensatory mortality, density-dependent mortality vs. density-independent mortality); and uncertainty about the system's state (e.g. population size, density, distribution). Uncertainty about resource structure or function is represented by the distribution $q = \{q(k)\}$ of hypothesis-specific probabilities $q(k)$. Uncertainty about the resource state is represented by hypothesis-specific probabilities $p(x|k)$ and the marginal distribution $\{p(x)\}$ with probabilities $p(x) = \sum_k p(x|k)q(k)$.

Uncertainty is reduced through experimentation and sampling, which can identify the system's state more precisely and improve understanding of its process or structure. We use the term 'experiment' generically to describe data resulting from experimental trials, field sampling or access to independent data sources. Let e represent an experiment that produces data about the resource, with $\{e_1, \dots, e_n\}$ a set of potential experiments (possibly consisting of a single experiment) and $\{f_e(z|x, k)\}$ the distribution for data z from experiment e with resource status x under hypothesis k . The data distribution $\{f_e(z|x)\}$ accounts only for resource state x , and $\{f_e(z|k)\}$ accounts only for hypothesis k .

We note that other sources of uncertainty, apart from partial observability of the system state and structural uncertainty, can be readily identified. These include partial controllability and environmental variation (Williams, 2011b), as well as other non-physical sources (e.g. linguistic imprecision, subjective judgment and disagreement (Morgan & Henrion, 1990)). We focus here on partial observability and structural uncertainty because of the ubiquity of these uncertainty sources in ecology and natural resources, and the need to address them in management situations.

3.2 | Posterior updating of uncertainty

For uncertainty about the system's state, the prior distribution $\{p(x)\}$ of probabilities of resource states can be combined with data probabilities $\{f_e(z|x)\}$ from experimentation to produce new posterior probabilities. A well-known approach is Bayesian updating (Lee, 1989), with posterior probabilities calculated by Bayes' theorem:

$$p(x|z) = \frac{p(x)f_e(z|x)}{\sum_x p(x)f_e(z|x)} = \frac{p(x)f_e(z|x)}{f_e(z)}.$$

Under uncertainty about the system's structure or process, prior hypothesis probabilities $\{q(k)\}$ can be combined with the data probabilities $\{f_e(z|k)\}$ to produce the posterior probabilities

$$q(k|z) = \frac{q(k)f_e(z|k)}{\sum_k q(k)f_e(z|k)} = \frac{q(k)f_e(z|k)}{f_e(z)}.$$

We note that other methods of posterior updating are possible, including maximum likelihood estimation (Mood et al., 1974) or even the simple replacement of a prior distribution with another distribution from an independent data source. Different updating approaches produce different posterior distributions, and therefore different valuations.

Although updating with discrete distributions can often be handled by summation over discrete states and hypotheses, as above, many natural resource problems involve continuous distributions. Obvious examples include resource states like animal density and vital rates such as survival and reproduction, which can vary across a continuous range of potential values. Continuous distributions present computational challenges. With Bayesian updating, an important simplification that greatly reduces these challenges involves conjugate distributions of the data and prior probabilities (Hobbs & Hooten, 2015), which produce a posterior distribution differing from the prior only in terms of its distribution parameters.

To illustrate, assume a normal distribution for the system state, $x \sim N(\mu, \sigma^2)$, and a lognormal distribution for the data, $z \sim \text{lognormal}(x, \sigma_0^2)$. Because the lognormal and normal distributions are conjugate (Hobbs & Hooten, 2015), Bayesian updating produces another normal distribution, $(x|z) \sim N(\mu_1, \sigma_1^2)$, with

$$\mu_1 = \left(\frac{\mu_0}{\sigma_0^2} + \frac{\ln(z)}{\sigma^2} \right) / \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} \right),$$

and

$$\sigma_1^2 = \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} \right)^{-1}.$$

The conjugate relationship thus allows one to avoid more burdensome analytical or numerical methods that otherwise would be needed to

determine a posterior distribution (see Appendix). Exact expressions of information value for this uncertainty structure are found in, for example, Bickel (2008) and Bhattacharjya et al. (2013).

4 | VALUATION

A value function representing resource valuation includes arguments that reflect the factors influencing value. In particular, value can be expressed in terms of management actions as well as attributes of the resource system like abundance, diversity and spatial extent. In addition, it may reflect hypothesized structures and processes of the resource system, such as species interactions, community structure, environmental influences and movement patterns.

In general, resource value can incorporate any combination of the foregoing factors. For example, a value function $V(a, x)$ denotes the influence of management a and resource state x , for a known or agreed-upon hypothesis (no need for the index k). Alternatively, $V_k(a)$ denotes the influence of management a under hypothesis k , for a known or agreed-upon system state (no need for the state variable x). Finally, $V_k(a, x)$ denotes the influence of management a and resource state x under hypothesis k .

A generic expression of value accounts for the costs of experimentation as well as the net benefits of management:

$$V_k(a, x, e, z) = B_k(a, x) - C(e, z),$$

with returns $R_k(a, x)$, costs $C_k(a, x)$ and net benefits $B_k(a, x) = R_k(a, x) - C_k(a, x)$ for action a , along with costs $C(e, z)$ for experiment e . Because environmental returns and costs often are not expressed in the same units of value, it may be necessary to express returns in units of economic value. For example, it may be possible to describe the biological benefit of population augmentation in terms of dollars, for comparison with the dollar cost of the enhancement. More generally, returns and costs often can be described in terms of 'utility' (Clemen & Reilly, 2014). Several variations of valuation can be distinguished:

- The resource structure is known but its state is uncertain. Then the structural index k in $V_k(a, x, e, z)$ can be treated as a fixed parameter, and value can be represented by $V(a, x, e, z) = B(a, x) - C(e, z)$.
- The resource state is known but structure is uncertain. Then state x can be treated as a fixed parameter, and value can be represented by $V_k(a, e, z) = B_k(a) - C(e, z)$.
- The value function omits costs and records only returns, $V_k(a, x) = R_k(a, x)$. Then the focus of decision-making is on increasing the returns to management actions. This is the usual way of measuring value in an analysis of the value of information; overall costs are typically considered only after determining marginal value based on returns (see Section 7.3 for further discussion).
- The value function may omit returns and record only costs, $V_k(a, x, e, z) = -[C_k(a, x) + C(e, z)]$. In this case the decision focus is on choices that can reduce costs.

Because the value function just described is influenced both by actions as well as uncertainty factors (state x and structure k), it may be thought of as a bivariate function, in which the action variable and uncertainty distribution are at least partially independent. In this context one can hold either one or the other variable constant, and investigate not only how average values vary with actions for a given uncertainty distribution, but also how average values vary with uncertainty distributions for a particular action. The latter approach offers an opportunity to expand the concept of value of information itself, as discussed in Sections 5.4 and 5.5.

5 | SCENARIOS FOR THE VALUE OF SAMPLE INFORMATION

In this section, we describe five scenarios for the value of sample information that cover a range of complexity. Each scenario is described in terms of the common elements listed in Section 2, that is, a resource system characterized by attributes of its state or structure, uncertainty about the resource system, management action(s), a value function and new 'experimental' data. To simplify notation, we initially assume uncertainty about the resource system state, as represented by a prior distribution $\underline{p} = \{p(x)\}$ and a value function $V(a, x)$. We discuss uncertainty about system structure in a later section.

The first three scenarios (1–3) represent standard metrics of the value of sample information—conditional value, expected value and maximum expected value of sample information, as described in, for example, Raiffa and Schlaifer (1961). The last two scenarios (4–5) represent logical extensions of the standard mathematical framework.

In scenarios 1–3, we consider analyses of values that involve the collection of data to inform future management actions. This is the purview of traditional VSI, where decision-making is directly influenced by the data that are collected (Figure 1). The posterior values, which depend on actions that are directly influenced by new data, are compared to prior values that are determined in the absence of new data. As shown below, one option is to collect the data first, determine optimal actions with the data in hand, and do a posterior assessment of value (scenario 1). Another is to do a pre-posterior assessment that accounts for stochastic data results, before any data are actually collected (scenarios 2–3). With either option, the point is to use data, whether actual or potential, as a basis for future decision-making.

Alternatively, a *retrospective* valuation involves management actions that have been pre-selected, independent of any new data collection (scenarios 4–5). Though valuation again compares posterior values based on new data to prior values in the absence of new data, the approach differs from a prospective assessment in that management actions are not influenced by the data (Figure 3). Instead, the perspective shifts to experimental design. A retrospective assessment looks back at decisions already made, and assesses value based on the impact the new data have on the posterior probabilities. The difference between posterior values based on new data and prior values in the absence of new data is useful for guiding the design of data collection, rather than for selecting management

Scenarios 1–3

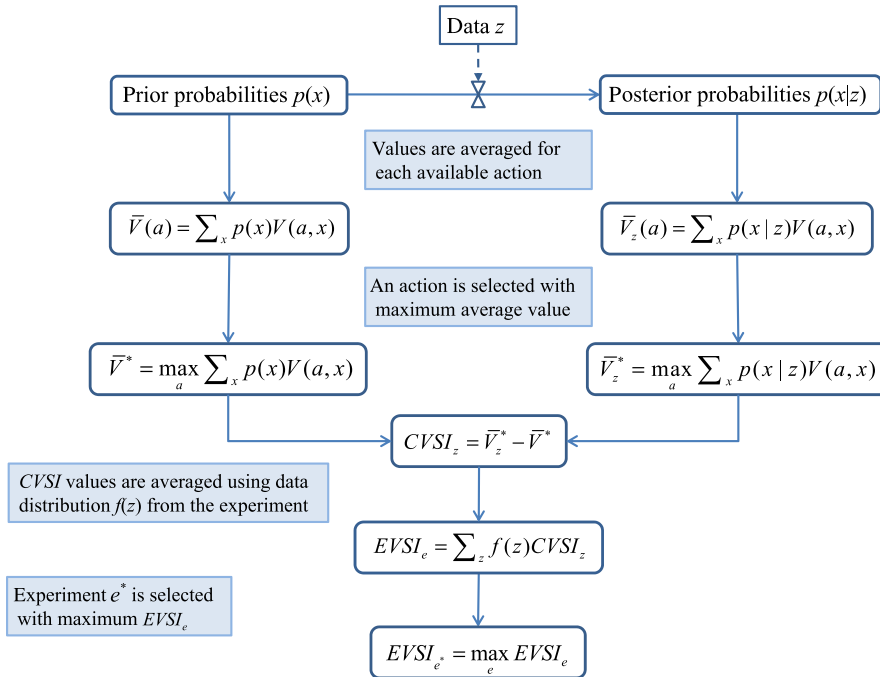


FIGURE 1 Computation of the value of sample information metrics in scenarios 1–3. **Scenario 1:** Assumed prior uncertainties are used to compute average values for each possible action. Posterior uncertainties based on new data z are used to compute average values for each posterior action. Optimal averages for prior and posterior distributions are compared to produce $CVSI_z$. **Scenario 2:** $EVSI_e$ is the average of $CVSI_z$ values over the distribution of data z for experiment e . **Scenario 3:** $EVSI_e^*$ is the optimal $EVSI_e$ value over the range of experiments

actions. Although this approach is uncommon in ecology (but see Bal et al., 2018), it occurs more frequently in medical science (e.g. Willan & Eckermann, 2010; Willan & Pinto, 2005).

5.1 | Scenario 1—Conditional value of sample information

In scenario 1, average values are determined for the prior and posterior probability distributions representing uncertainty, and actions are selected that maximize average values for both (Figure 1). Thus, an average value $\bar{V}(a)$ is calculated for each of the available actions under prior uncertainty, and the action with maximum average value is identified:

$$\bar{V}^* = \bar{V}(a^*) = \max_a \bar{V}(a).$$

Similarly, average values $\bar{V}_z(a)$ are calculated for the available actions under posterior uncertainty, and again the action with maximum average value is identified,

$$\bar{V}_z^* = \bar{V}_z(a_z^*) = \max_a \bar{V}_z(a),$$

where the subscript refers to the data point z upon which uncertainty updating is based. The difference between these two maximum average values is the *conditional value of sample information*

$$CVSI_z = \bar{V}_z^* - \bar{V}^*,$$

(conditional on the particular data point z (Raiffa & Schlaifer, 1961)). This measure is in effect a comparison of the highest value achievable

with the additional information z , versus the highest value achievable without it.

A simple example of scenario 1 is the initiation of a new survey of the extent and severity of climate change in a region, followed by the use of survey results to refine a climate adaptation strategy that was previously thought to be optimal on the basis of earlier information. The difference between the optimal valuations expresses the gain in value from the new survey information.

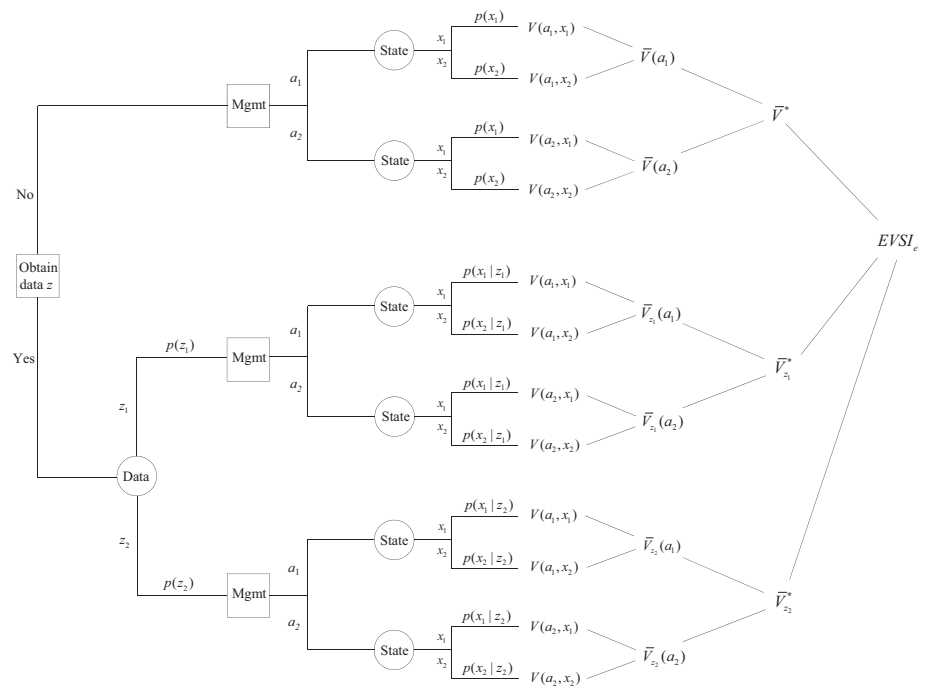
Scenario 1 is an example of a *posterior assessment*, in that it is based on specific data z to calculate posterior probabilities. The reference to specific data makes the resulting conditional value of sample information useful for valuation with data known already. However, such a conditional valuation is not particularly useful for determining the benefit of new information before any data are collected. To make the latter determination, we need an analysis that considers the full set of possible data points, as described in scenarios 2 and 3.

5.2 | Scenario 2—Expected value of sample information

Scenario 2 builds directly on scenario 1, by recognizing a range of data-specific $CVSIs$ produced in scenario 1 and calculating an average $CVSI$ over the data values (Figure 1). The result is the *expected value of sample information*,

$$\begin{aligned} EVSI_e &= \sum_z f_e(z) CVSI_z \\ &= \sum_z f_e(z) (\bar{V}_z^* - \bar{V}^*) \\ &= \sum_z f_e(z) \bar{V}_z^* - \bar{V}^*, \end{aligned}$$

FIGURE 2 Decision tree for calculation of $EVSI_e$ involving state uncertainty, with two states, two actions and two experimental outcomes. Boxes represent actions; circles represent factors that are subject to uncertainties



where the subscript e on the expectation indexes the experiment producing $f_e(z)$. The sequence of calculations for $EVSI_e$ is highlighted in Figure 2 in terms of decision trees (Clemen & Reilly, 2014), for a problem with two states, two decisions and two experimental outcomes.

An example of scenario 2 might involve a wildlife monitoring plan, and corresponding data distribution, designed to reduce uncertainty about population growth, so as to enhance population sustainability while at the same time allowing hunting. $EVSI_e$ measures the enhancement that is expected to result from the monitoring effort.

In contrast to scenario 1, scenario 2 is a *pre-posterior assessment* that does not require data to be known in advance. By averaging across potential data points, an expected value of sample information can be calculated before any new data are actually collected. Thus, $EVSI_e$ allows one to consider the potential benefit of new information in deciding whether to proceed with data collection.

5.3 | Scenario 3—Maximum expected value of sample information

Scenario 3 is also a pre-posterior valuation, which builds directly on scenario 2 by considering the expected value of sample information for multiple experiments, each generating its own data distribution $f_e(z)$ (Figure 1). A comparison across experiments allows one to identify the experiment with *maximum expected value of sample information*:

$$\begin{aligned} EVSI^* &= \max_e EVSI_e \\ &= \max_e \sum_z f_e(z) CVSI_z, \end{aligned}$$

with the optimal experiment

$$e^* = \arg \max_e EVSI_e.$$

It is through such a comparative evaluation of different experiments that the expected value of sample information can play a role in sampling and experimental design.

An example of scenario 3 is the selection of a sampling design for climate change mitigation involving several potential activities. Each sampling design and mitigation action is expected to produce a distinct distribution of carbon dioxide emissions, and a comparison of $EVSI_e$ values across designs allows one to choose the design that on average will yield the optimal reduction of carbon dioxide emissions.

Each of the foregoing scenarios (1–3) involves the comparative evaluation of actions in determining optimal valuations, with optimal actions a^* and a_z^* that are based on the distributions $p(x)$ and $p(x|z)$. That is, the actions selected are influenced by prior and posterior uncertainties. A further topic for consideration is the influence of uncertainty when actions are selected independently, as in scenarios 4 and 5.

5.4 | Scenario 4—Value of sample information with pre-selected actions

In scenarios 4 and 5 the emphasis shifts from management to data collection. With scenarios 1–3, the focus was primarily on management actions, and the adaptation of management with new information. In the next two scenarios, the emphasis is on data collection design. Thus, scenarios 4 and 5 involve pre-selected management actions that are not influenced by the data (Figure 3). Valuation again compares posterior values based on new data and prior values in the absence of new

Scenarios 4–5

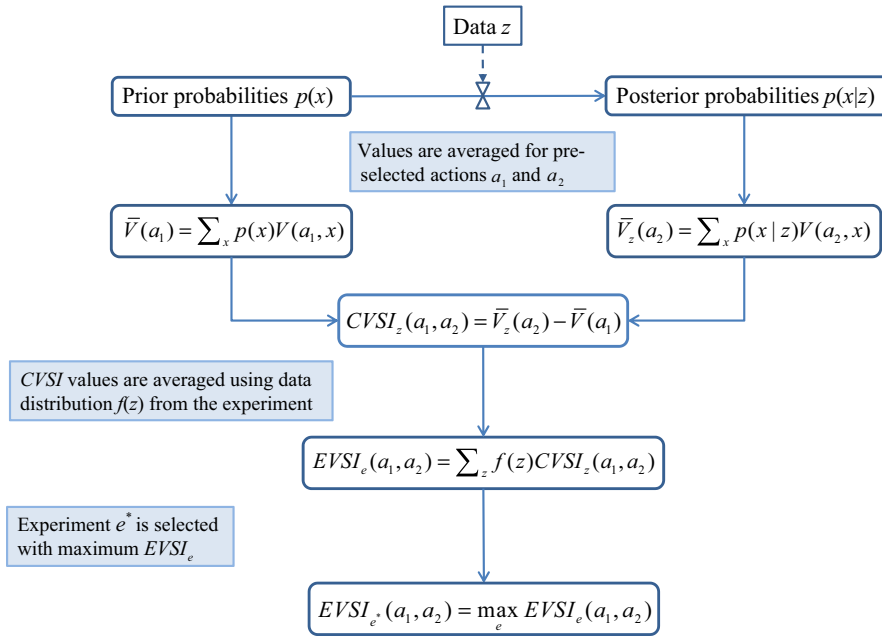


FIGURE 3 Computation of VSI metrics in scenarios 4 and 5. **Scenario 4:** Assumed prior uncertainties are used to compute average value for action a_1 . Posterior uncertainties based on new data z are used to compute average value for action a_2 . Prior and posterior average values are compared to produce $CVSI_z(a_1, a_2)$, $EVSI_e(a_1, a_2)$ and $EVSI_{e^*}(a_1, a_2)$. **Scenario 5:** Computations are identical with those for scenario 4, except that prior and posterior actions are identical: $a_1 = a_2 = a$

data, but the difference between values is useful for guiding the design of data collection rather than selecting management actions.

As with the previous scenarios, calculations in scenario 4 begin with a distribution of prior probabilities and combination with new data to produce posterior probabilities. The average prior valuation based on pre-selected action a_1 and prior distribution $p(x)$ is

$$\bar{V}(a_1) = \sum_x p(x)V(a_1, x),$$

and the average posterior valuation based on pre-selected action a_2 and posterior distribution $p(x|z)$ is

$$\bar{V}_z(a_2) = \sum_x p(x|z)V(a_2, x).$$

Because the actions are pre-selected, maximization across actions is no longer required. Then the difference between values,

$$CVSI_z(a_1, a_2) = \bar{V}_z(a_2) - \bar{V}(a_1),$$

measures the value of the sample information given the two actions. Averaging $CVSI_z(a_1, a_2)$ over the distribution $f_e(z)$ produces a pre-posterior assessment of value for experiment e , and maximizing over experiments identifies an optimal data collection design e^* (Figure 3).

An example of scenario 4 might be new federal regulations for a wildlife refuge that mandate a change in fire management, followed by monitoring of the outcome. In this instance, the prior and posterior actions consist of the pre-selected fire management strategies, the value function measures net biological productivity and the 'experiment' consists of monitoring productivity after the change in fire management. The comparison of prior and posterior values reflects the change in knowledge due to the additional monitoring, as well as the effect of substituting one management action for another.

Variants of scenario 4 are frequently used in land and natural resource programmes, in which new agency leadership initiates a change in management along with follow-up monitoring for effectiveness. Any of a number of management situations (Grafton et al., 2010; Ryder, 2018; Silvy, 2012) entail decisions to change permits and other controls regulating fish and wildlife populations, in response to public pressure, patterns of land and water use or political directives.

The value function in scenario 4 operates like any of the other value functions in scenarios 1–3, with action and state drivers a and x in $V(a, x)$. The only difference is that the action drivers, which can differ between pre- and post-valuations just as they do in scenarios 1–3, are given a priori and are not subject to optimal selection. Obviating the selection of actions affects the actual value produced, but does not change the meaning of the value function itself. The same value functions that apply in scenarios 1–3 also can apply to scenario 4.

The use of two pre-selected actions in scenario 4 means that $CVSI_z(a_1, a_2)$ expresses not only the influence of new data, but also the influence of a change in management. To highlight the marginal value of the data alone, one can use identical prior and posterior actions.

5.5 | Scenario 5—Value of sample information with a single pre-selected action

Scenario 5 involves a system under fixed (or no) management, in which new information about the system's status alters the measurement of its economic or other value, without the influence of a change in management. In this context, $a_1 = a_2 = a$ and the value of sample information represents a change in value occasioned

TABLE 1 Examples of analysis of the value of sample information, adapted from the environmental literature to show the five valuation elements described in Section 2

Authors	Resource system	Potential actions	State or structural uncertainty	Value function	'Experimental' data
Canessa et al. (2015)	Frog population managed for long-term sustainability	Translocate a cohort of frogs (a_1), or not translocate (a_2)	Presence (k_1) or absence (k_2) of chytrid disease at the translocation site	$V_k(a)$ is population size after 10 years	Results of test for chytrid disease at the translocation site
Williams and Johnson (2015b)	Wildlife habitat subject to prescribed burns	Single burn (a_1) or 2 burns on successive weeks (a_2)	No rainfall (k_1) or rainfall in the second week (k_2)	$V_k(a)$ is combination of burn costs and wildlife impacts	Weather forecast information from local weather service
Williams and Johnson (2015b)	Managed predator/prey system	Predator control (a_1) and habitat improvement (a_2)	Low (x_1), medium (x_2) or high (x_3) prey population	$V(a, x)$ is prey density in response to management	Mark-release-recapture of individuals
Canessa et al. (2015)	Terrapin population managed for survival	Release of age-3 cohort into the wild (immediate (a_1), 1-year delay (a_2), 2-year delay (a_3))	Post-release survival (same as (k_1), greater than (k_2), or less than (k_3) wild survival)	$V_k(a) = S_{31}S_{32}S_{33}$ is a 3-year survival rate of released 3-year-old individuals	Release and tracking of survival of 10 radio-tagged 3-year-old turtles
Costello et al. (2010)	System of K marine habitat patches	Different levels of steady-state fish harvest $a = (a_1, \dots, a_K)$ in the patches	Dispersal rates $D = \{D_{ij}\}$ among patches	Weighted combination of profit from harvest and expected steady-state fish biomass	New distribution is given for the dispersal rates
Sahlin et al. (2011)	Habitat that is subject to invasion by non-indigenous species	Allow invasion (a_1) or prevent invasion (a_2) by a species	Native (x_1) or non-native (x_2) status of species from a pool of invasive species	Loss $L(a, x)$ associated with invasiveness x and action a	Screening for indigenous versus non-indigenous status
Clemen and Reilly (2014) and Pearlman et al. (2019)	Potential disaster (hurricane, volcanic eruption) threatening lives and property	Partial evacuation (a_1) or complete evacuation (a_2)	Path of hurricane or lava inundation (k_1 or k_2)	Loss $L_k(a)$ associated with the level of evacuation a and path k of storm or lava inundation	New information from remote sensing about storm path or pattern of inundation
Appendix, this paper	Wildlife population that is subjected to hunting	Specification of an annual harvest rate a	Annual survival rate with a continuous beta distribution	$V(a, s)$ is the amount of harvest expected with hunting rate a and survival rate s	Release and tracking of the survival of n radio-tagged individuals
Costello et al. (1998)	Salmon populations subjected to El Niño weather events	Level of annual catch	Annual survival and recruitment parameters	Optimal value of future benefits from commercial and recreational fishing	Forecasts of El Niño events
Grantham et al. (2009)	Potentially protected habitats for the Proteaceae family in South Africa	Extent and duration of survey effort over the course of a multi-year survey	Rates of habitat loss and establishment of new protected areas	Proteaceae retention at the end of 20 years	Results over the course of survey effort
Dakins et al. (1996)	Winter flounder <i>Pseudopleuronectes americanus</i> exposed to polychlorinated biphenyls (PCBs) in New Bedford, MA	Amount of dredging in the harbour to remove PCBs	Predicted PCB in flounder tissue, based on a model with random inputs	Loss function that includes dredging costs, fishery closure and re-start remediation costs	Samples of flounder, with tissue PCB measured for each fish

(Continues)

TABLE 1 (Continued)

Authors	Resource system	Potential actions	State or structural uncertainty	Value function	'Experimental' data
McDonald and Smith (1997)	Fish population that is subject to catch management for sustainability	A range of sustainable fish biomass targets	Three models producing different catch amounts as a function of sustainable stock	Amount of sustainable catch	Future observations of escapement and catch
Taylor et al. (1993)	Chemical compound with uncertain carcinogenic potential	Control of exposure and no control of exposure; posterior actions also include follow-up bioassay	Carcinogenic potency as a function of uncertain maximum tolerated dose (MTD)	Loss of function that includes cost of control and social cost of lives lost	MTD data from laboratory animal tests; potency for a given level of MTD, based on bio-assays
Forsberg and Guttormsen (2006)	Commercial harvest of farm-raised salmon	Immediate or delayed harvest	Price of fish harvested at a future date	Sales revenue given by harvested biomass time price per unit biomass	Prediction of future market conditions from expert opinion or forecasting model
Back et al. (2007)	Contaminated land that is investigated for remediation	Design and intensity of sampling	Amount of contamination	Combined estimator precision and sampling costs	Sample-based parameter estimates

by auxiliary information alone. Because new information (but not new management) is used to change the value, this scenario is essentially an efficiency analysis (Back et al., 2007). It may also be thought of as an assessment of the loss in value if fewer data were available.

As before, scenario 5 involves updating a distribution of prior probabilities with new data to produce a distribution of posterior probabilities (Figure 3), followed by the calculation of average values

$$\bar{V}(a) = \sum_x p(x) V(a, x),$$

and

$$\bar{V}_z(a) = \sum_x p(x|z) V(a, x).$$

The value of the information is simply the difference between these average values,

$$CVSI_z(a) = \bar{V}_z(a) - \bar{V}(a).$$

Thus, the gain in value for scenario 5 depends on the data used for updating the prior probabilities, but not on any management change. Because the value comparison assumes known data z , the comparison describes a posterior assessment. Averaging $CVSI_z(a)$ over the distribution of potential data points produces a pre-posterior assessment of value for experiment e , and maximizing over experiments identifies an optimal design e^* (Figure 3). In the absence of management per se, $V(a, x)$ is effectively the univariate function $V(x)$, with actual values produced solely as a result of new information that informs x .

Many variants of scenario 5 use new monitoring to sample the condition and productivity of natural resources that are under stable management. A specific example might involve monitoring wildlife habitat on a natural preserve with remote sensing. In this case there is new data collection, perhaps mandated by political authorities, but no change in management actions. VSI is then a comparison of prior and posterior values reflecting the acquisition of new data. Though there is no active management on the preserve (action a consists of an absence of management), better understanding of the distribution of habitats, and recognition of the value of their ecosystem services (National Ecosystem Services Partnership, 2014) with new data, can by themselves lead to a change in perceived value of the preserve. $CVSI_z(a)$ measures the change in value resulting from new sensor information z , and $EVSI_{e^*}(a)$ is used to guide decisions about the data collection from the sensor.

Other applications of scenario 5 measure value in terms of design criteria related to the costs of experiments. For example, Back et al. (2007) analysed the value of information to determine an optimal sampling design for investigating contaminated lands that are subject to remediation, and measured value in terms of estimator precision (see Table 1). In medical applications, Willan and Pinto (2005) and Willan and Eckermann (2010) used this type

of value of information analysis to optimize design of clinical trials by relating treatment efficacy of different experimental designs to their cost.

Scenario 5 is the simplest of the five scenarios in our spectrum, in that it focuses exclusively on the effect of data in reducing uncertainty and changing value. The contrast of values nonetheless produces a value of sample information that is useful for designing a data-collection effort, even though both prior and posterior values involve the same (or no) management action. Though valuation in this scenario focuses on data collection per se, we emphasize that in ecological applications, the value function is usually tied to conservation objectives rather than to statistical performance as such.

6 | UNCERTAINTY ABOUT SYSTEM STRUCTURE

Thus far we have considered *VSI* in terms of uncertainty about system state. The treatment of uncertainty about system structure follows the same logic as with system state, except the reference to state x in the foregoing development is replaced by a reference to hypothesis k . The following steps produce scenarios for the five variants of sample information under uncertainty about system structure rather than system state.

1. For each action a , average the values $V_k(a)$ over the prior probabilities $\{q(k)\}$ to get the prior valuation $\bar{V}(a) = \sum_k q(k) V_k(a)$.
2. For each action a with data z , average the values $V_k(a)$ over the posterior probabilities $\{q(k|z)\}$ to get the posterior valuation $\bar{V}_z(a) = \sum_k q(k|z) V_k(a)$.
3. Maximize the valuation in step 1 over the potential actions a to get $\bar{V}^* = \max_a \bar{V}(a)$; maximize the valuation in step 2 to get $\bar{V}_z^* = \max_a \bar{V}_z(a)$.
4. Compute the difference of maximum valuations in step 3 to get the conditional value of sample information $CVSI_z = \bar{V}_z^* - \bar{V}^*$.
5. Average the $CVSI_z$ values in step 4 with $f_e(z)$ to get the expected value of sample information $EVSI_e = \sum_e f_e(z) \bar{V}_z^* - \bar{V}^*$.
6. Maximize $EVSI_e$ in step 5 over the possible experiments, to get $EVSI^* = \max_e EVSI_e$.

These expressions of *VSI* for structural uncertainty differ from analogous expressions for state uncertainty only in the use of structure probabilities. After replacing the probabilities $\{p(x)\}$ for state uncertainty with $\{q(k)\}$ for structural uncertainty, all calculations for the scenarios under structural uncertainty can be seen as identical to those under state uncertainty.

7 | EXTENSIONS OF THE FRAMEWORK

There are a number of extensions for the scenario framework we described earlier, including hybrids that combine features of different

scenarios; spatial applications; the incorporation of system dynamics; valuation based on costs rather than net benefits; trade-offs among system attributes; and other extensions.

7.1 | Hybrid scenarios

There are other possibilities for *VSI* in addition to the five scenarios already discussed. Two possibilities are hybrids of scenarios 1 and 4:

HYBRID(a₁) contrasts a prior valuation $\bar{V}(a_1)$ involving fixed action (as in scenario 4) with a posterior valuation $\max_a \bar{V}_z(a)$ (as in scenario 1). This situation arises naturally with the potential replacement of established management actions by new actions that are optimal with respect to improved understanding of the resource system.

HYBRID(a₂) contrasts a prior valuation $\max_a \bar{V}(a)$ with a posterior valuation $\bar{V}_z(a_2)$. Though less useful, this situation might apply when long-standing management that was previously regarded as optimum is adjusted at the direction of an oversight authority, which also mandates follow-up monitoring.

It is straightforward to show that $VSI_{HYBRID(a_2)} < CVSI_z < VSI_{HYBRID(a_1)}$. The inequality $CVSI_z < VSI_{HYBRID(a_1)}$ indicates that the replacement of optimal with fixed (suboptimal) prior actions produces greater information value than $CVSI_z$. On the other hand, the replacement of optimal with fixed posterior management reduces the information value ($VSI_{HYBRID(a_2)} < CVSI_z$). In particular, management corresponding to *HYBRID(a₂)* is inferior to *HYBRID(a₁)*, in that $VSI_{HYBRID(a_2)} < VSI_{HYBRID(a_1)}$.

Bal et al. (2018) documented another hybrid scenario in which they used valuation to determine an appropriate model form along with an optimal sampling strategy, in order to compare the effectiveness of surveillance monitoring versus targeted monitoring in managing multiple threats to multiple species, under economic constraints. Their application included uncertainty among species models, and focused on decisions about monitoring that increase value of monitoring information in reducing uncertainty.

7.2 | Spatial applications

In our development thus far we have considered uncertain states with a prior probability distribution $p(x)$, and data z related to the state with conditional distribution $f(z|x)$. The state and data distributions can be combined via Bayes' theorem to produce a posterior distribution $p(x|z)$ showing the influence of the data on the state distribution. The value of sample information in the scenarios described earlier is calculated as a comparison of value averaged over the posterior and prior distributions.

An extension of *VSI* in a spatial context involves additional structure imposed on the states, data and their distributions. Here we consider a spatial region that is divided into areas s_i , $i = 1, \dots, n$, such that the state $x_i = x(s_i)$ for area s_i is uncertain with distribution $p(x_i)$.

With n area-specific states, the system state of the region is represented collectively by a vector $\mathbf{x}' = (x_1, \dots, x_n)$. Spatial structure is characterized by a joint probability distribution $p(\mathbf{x})$ with a multivariate mean $\underline{\mu}' = (\mu_1, \dots, \mu_n)$, and a covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \dots & \sigma_n^2 \end{bmatrix}$$

of variances σ_i^2 for state x_i and covariances σ_{ij} between states x_i and x_j that may vary with the distance between the areas. Such a framework is especially relevant to remote sensing of landscape features, where multispectral data are recorded over a pixelated area and pixel data exhibit patterns of correlation depending on their proximate locations in the area.

A specific example might consist of a forested region that is gridded into areas s_i containing an uncertain amount of woody biomass x_i . A survey of the region produces data z_i for some (perhaps all) of the areas, which can be used to update the distribution for the region to $p(\mathbf{x}|\mathbf{z})$. A value function $V(a, \mathbf{x}) = \sum_i V(a, x_i)$ aggregates the net profit from timber harvest over the areas, with harvest actions that include clear cutting (a_1), moderate thinning (a_2), or heavy thinning (a_3).

Computing the value of sample information for this situation follows the same general pattern as shown in Figure 1, except that averaging to produce $\bar{V}(a)$ and $\bar{V}_z(a)$ is complicated by the multivariate nature of the states and data over the region (Bhattacharjya et al., 2013; Bickel, 2008). Assume for example a prior distribution for states over the region that is multivariate normal, $\mathbf{x} \sim N(\underline{\mu}, \Sigma)$, with data that are linked to the states by $z_i = x_i + \varepsilon_i$, $\varepsilon \sim N(0, \gamma^2 I)$. A design \underline{F} for data collection produces a marginal distribution $z \sim N(\underline{F}\underline{\mu}, \underline{F}\Sigma\underline{F}' + \gamma^2 I)$ for the data, along with the multivariate normal posterior distribution.

$$\mathbf{x}|\mathbf{z} \sim N\left(\underline{\mu} + \Sigma\underline{F}'(\underline{F}\Sigma\underline{F}' + \gamma^2 I)^{-1}(\mathbf{z} - \underline{F}\underline{\mu}), \Sigma - \Sigma\underline{F}'(\underline{F}\Sigma\underline{F}' + \gamma^2 I)^{-1}\underline{F}\Sigma\right).$$

for the states (Graybill, 1976). The prior and posterior state distributions are used in $\bar{V}(a)$ and $\bar{V}_z(a)$ to produce $CVSI_z$, and averaging the $CVSI_z$ values over the marginal distribution of z yields $EVSI_e$ for the spatial problem.

This formulation includes a number of simplifying assumptions about the value function, state and data distributions, and management alternatives. Even with the simplifications, computing $EVSI$ can be challenging depending on the structure of the covariance matrix and the design used for data collection. There are only a limited number of ecological and environmental applications of the value of sample information in a spatial context. Examples that incorporate spatial covariance structure include weed management (Wiles, 2004), salmon farming (Forsberg & Guttormsen, 2006) and forest information (Kangas, 2010). Eidsvik et al. (2015) provide a detailed technical review of spatial VSI , aimed primarily at the geosciences. Again, we note the pertinence to VSI of spatial patterns of covariation in multispectral data in remotely sensed images of landscapes.

7.3 | Dynamic systems

Value of sample information can be extended to include fluctuating environmental conditions and actions over time, in which value is based on system responses to a management strategy rather than to single actions (Williams, 2011a). With uncertainty about system structure, the optimum average value is equivalent to active adaptive optimization (Williams & Johnson, 2018), so the comparison of prior and posterior values is actually a comparison of optimal adaptive values under $\{q(k)\}$ and $\{q(k|z)\}$. For example, Smith et al. (2013) developed an adaptive management programme for regulating the horseshoe crab *Limulus polyphemus* fishery, which provides the basics needed for analysing the value of information; Johnson et al. (2014) considered the value of information in a dynamic framework for managing northern bobwhites *Colinus virginianus*; and Moore and Runge (2012) used a dynamic model to identify the best strategy for management of an invasive willow species in Australia. Of course, any application of VSI requires an experiment to get the distribution of data needed for $\{q(k|z)\}$ (Williams, 2015).

With uncertainty about system state under Markovian state transitions (Puterman, 1994), optimal average value is equivalent to analysis of partially observable Markov decision processes (POMDP) (Kaelbling et al., 1998; Williams, 2009). The comparison of prior and posterior values is then a comparison of POMDP solutions under $\{p(\mathbf{x})\}$ and $\{p(\mathbf{x}|\mathbf{z})\}$. Examples in fish and wildlife biology that allow for value of information under partial observability include fishermen's decision-making (Lane, 1989), seabird habitat management (Tomberlin, 2010) and resolution of structural uncertainty (Williams, 2011a). Again, the computation of VSI requires an experiment to get the distribution of data needed for $\{q(\mathbf{x}|\mathbf{z})\}$ (Williams, 2015).

A somewhat simpler framework for iterative decision-making involves sequential experimentation and a stopping rule for its termination (James & Gorelick, 1994). In this situation an initial prior probability distribution and set of potential experiments are used to generate $EVSI_e$. A positive expected $EVSI_e$ suggests that the gain in knowledge from experimentation will produce a positive net benefit, so a decision is made to conduct the optimal experiment. The resulting experimental data z are used to update uncertainty about system structure or status, which then becomes a new prior probability distribution for use in calculating a new $EVSI_e$. This process is repeated as long as $EVSI_e$ is positive. When a negative $EVSI_e$ is produced (e.g. the data z are statistical outliers that obscure the actual system state and lower the average posterior value), further experimentation is stopped.

7.4 | Cost analysis

In many instances it is useful to reframe VSI in terms of minimizing costs rather than maximizing benefits. For example, Bennett et al. (2018) and Bal et al. (2018) considered sampling costs versus the

information value of different monitoring schemes under a limited budget.

In Section 4 we described a generic value function in terms of costs and benefits,

$$V_k(a, x, e, z) = B_k(a, x) - C(e, z),$$

which includes costs $C(e, z)$ for experiment e as well as net management benefits

$$B_k(a, x) = R_k(a, x) - C_k(a, x).$$

By omitting returns $R_k(a, x)$, the value function records only the costs of management and experimentation,

$$V_k(a, x, e, z) = -[C_k(a, x) + C(e, z)],$$

and the decision focus is on actions and experiments that can reduce costs.

To illustrate, consider the expected value of sample information in scenario 2, in which CVSIs are averaged over the distribution of data values z . As shown in Figure 2, the optimal actions a^* and a_z^* in $CVSI_z$ are obtained by maximizing $\bar{V}(a)$ and $\bar{V}_z(a)$ over the potential actions to get the net benefits $\bar{V}^* = \bar{V}(a^*)$ and $\bar{V}_z^* = \bar{V}_z(a^*)$. If the value function accounts only for the costs $C_k(a, x)$ and $C(e, z)$, optimal decision-making seeks to *minimize* these costs:

$$\begin{aligned}\bar{V}^* &= \min_a \bar{V}(a) \\ &= \min_a \sum_x p(x) C(a, x),\end{aligned}$$

and

$$\begin{aligned}\bar{V}_z^* &= \min_a \bar{V}_z(a) \\ &= \min_a \sum_x p(x|z) [C(a, x) + C(e, z)].\end{aligned}$$

All other calculations for $EVSI_z$ are the same as described in Section 5.2 and shown in Figure 1.

Thus, a focus on costs follows the same logic as with net benefits, except that costs are minimized, whereas net benefits are maximized. Because minimizing a cost produces the same result as maximizing its negative, for example,

$$\begin{aligned}\bar{V}^* &= \min_a \sum_x p(x) C(a, x) \\ &= -\max_a \sum_x p(x) [-C(a, x)],\end{aligned}$$

both approaches can be described with maximization as in Section 5.1.

There are at least two ways in the literature that costs are handled with VSI . One way is to include the costs (as above), with net benefits and experimental costs actually included in the value function itself. In this case, VSI measures net returns (or losses) that accrue

to new information, indicating the relative gains for data collection (Raiffa & Schlaifer, 1961). The other, and more traditional, way is to use a value function that includes only returns, so that VSI highlights the marginal gain in returns from new information, without any consideration of the costs of acquiring and using the information. In a typical traditional application, the VSI value would then be compared to the associated costs, and data collection could be justified by a VSI value greater than the costs of its acquisition. Alternatively, the VSI value itself could be identified as an upper limit of the cost one should be willing to incur for purchase of the information (Maxwell et al., 2015).

It is straightforward to show that these two approaches do not necessarily yield the same results. Consider a situation in which returns for an action a_1 are large but only marginally exceed its costs; and returns for action a_2 are smaller than for a_1 , but nevertheless substantially greater than its costs. Maximum valuation based on net returns $R(a) - C(a)$ identifies optimal action a_2 for a value of $R(a_2) - C(a_2)$. However, maximum valuation based on returns alone identifies optimal action a_1 for a value of $R(a_1)$, producing a net value of $R(a_1) - C(a_1)$. This difference in overall valuation reflects the difference in value functions, and suggests caution in deciding which approach to take.

7.5 | Analysis issues

There are several analysis issues concerning patterns in VSI values. For example, consider variation in the prior probabilities that serve as the starting point for the conditional value of information in scenario 1, and assume state uncertainty and Bayesian posterior probabilities $p(x|z)$. As uncertainty declines (i.e. $p(x)$ approaches 1 for some state x'), so does the corresponding posterior uncertainty $p(x|z)$, irrespective of the data z . The result is that $\bar{V}(a) = \sum_x p(x) V(a, x)$ and $\bar{V}_z(a) = \sum_x p(x|z) V(a, x)$ both converge to $V(a, x')$, and $CVSI_z$ converges to

$$\max_a V(a, x') - \max_a V(a, x') = 0.$$

In contrast, the magnitude of $CVSI$ grows as uncertainty increases from 0, up to a maximum value in the interior of the probability space for \underline{p} . This mirrors a similar pattern for expected values of perfect and partial perfect information (Williams & Johnson, 2015). As with $EVPI$, the use of maximization in determining $EVSI$ induces a partition of the probability space, wherein the same action is used for each \underline{p} in a given partition (Williams et al., 2011). Among other things, this implies that a positive value of $EVSI$ does not by itself suggest a change in action.

In general, system status can be represented with multiple state attributes $\underline{x} = (x_1, \dots, x_i)$, and management with multivariate combinations $\underline{a} = (a_1, \dots, a_m)$ of actions. In this situation values are averaged over multivariate probabilities $p(\underline{x})$ and $p(\underline{x}|\underline{z})$ to produce $\bar{V}(\underline{a})$ and $\bar{V}_z(\underline{a})$, and the maximizations

$$\bar{V}^* = \max_{\underline{a}} \bar{V}(\underline{a}),$$

and

$$\bar{V}_z^* = \max_a \bar{V}_z(a)$$

over a offer the possibility of considering trade-offs in value among the action elements. Framing the value of sample information in this way allows for the use of multivariate optimization in identifying optimal actions and experiments.

Finally, there is a question about the functional form of the value function that will ensure non-negative values of sample information. One way to ensure non-negativity is simply to impose that condition on results. Another way is to identify positive gradients in the value function such that reducing uncertainty with new data actually leads to increased value. Yet another way is to change the comparison $\bar{V}_z(a_z^*) - \bar{V}(a^*)$ based on optimal prior and posterior actions a^* and a_z^* , to a comparison $\bar{V}_z(a_z^*) - \bar{V}_z(a^*)$ that instead uses the posterior distribution $\{p(x|z)\}$ in both the prior and posterior valuations (McDonald & Smith, 1997). It is straightforward to show that the latter metric is necessarily non-negative. Using the posterior distribution for both values in the comparison isolates the influence of the actions, in much the same way that the analysis in scenario 5 isolates the influence of new data.

8 | EVPI AS A FORM OF EVSI

In most discourses about value of information the introduction of *EVSI* occurs after a discussion of *EVPI*, perhaps because *EVPI* is analytically simpler. However, there are insights to be gained by beginning with *EVSI* and developing *EVPI* from it. Consider, for example, the use of data z from a sampling effort to produce an estimator $\hat{\theta}$ of parameter θ for an ecological system. Here we use the generic term θ to represent any system attribute, whether it is a system state, structural feature, or parameter, to express the idea that there is a system feature with uncertainty that can be reduced with additional data. Assuming $E(\hat{\theta}) = \theta$ and $\hat{\theta}$ is sufficient for θ , we have

$$f_e(z|\theta) = g(\hat{\theta}|\theta) h(z),$$

so that

$$\begin{aligned} f_e(z) &= \sum_{\theta} f_e(z|\theta) p(\theta) \\ &= \sum_{\theta} g(\hat{\theta}|\theta) h(z) p(\theta), \end{aligned}$$

and

$$\begin{aligned} p(\theta|z) &= \frac{f_e(z|\theta) p(\theta)}{f_e(z)} \\ &= \frac{g(\hat{\theta}|\theta) p(\theta)}{\sum_{\theta} g(\hat{\theta}|\theta) p(\theta)}, \end{aligned}$$

which converges to 1 with increasing sample size. In *EVSI* the average of posterior maxima,

$$\sum_z f_e(z) \max_a \sum_{\theta} p(\theta|z) V(a, \theta),$$

is therefore approximated for large samples by

$$\begin{aligned} \sum_z \left[\sum_{\theta} f_e(z|\theta) p(\theta) \right] \max_a V(a, \theta) &= \sum_{\theta} p(\theta) \left[\sum_z f_e(z|\theta) \right] \max_a V(a, \theta) \\ &= \sum_{\theta} p(\theta) \max_a V(a, \theta), \end{aligned}$$

so that

$$EVSI_e = \left[\sum_z f_e(z) \max_a \sum_{\theta} p(\theta|z) V(a, \theta) \right] - \left[\max_a \sum_{\theta} p(\theta) V(a, \theta) \right],$$

is seen to converge to

$$EVPI_e = \left[\sum_{\theta} p(\theta) \max_a V(a, \theta) \right] - \left[\max_a \sum_{\theta} p(\theta|z) V(a, \theta) \right],$$

as sample size increases. This means that *EVSI*, which may be negative, can be expected to increase in value with more intensive sampling. That is, as 'perfect' information is approached with greater sampling effort, *EVSI* approaches the upper bound of *EVPI*.

9 | EXAMPLES IN THE LITERATURE

In the ecological literature, studies of the value of sample information (e.g. Canessa et al., 2015; Costello et al., 2010; Runge et al., 2011; Sahlin et al., 2011; Williams, 2015), can be described in terms of the essential features in Section 2 and the computations in Sections 5 and 6, as shown by the examples in Table 1. The studies in Table 1 include both state and structural uncertainties, as well as discrete and continuous variation in actions and uncertainties. Value functions in the examples range from biological indicators such as population size and survival rates, to economic losses from invasions of non-native species, to measures of sampling efficiency and effectiveness. Experimental updating of uncertainty is represented by formal experimentation with Bayesian updating, as well as a simple replacement of one distribution with another from an external source. Yet despite their variety, all the examples can be formulated in terms of the five essential *VSI* elements identified in Section 2.

In the Appendix we use an example of frog population translocation from Canessa et al. (2015) to demonstrate the framing of a problem with discrete uncertainty, and an example involving sustainability of a hunted population to demonstrate the framing of a problem with continuous uncertainty.

10 | DISCUSSION

Most analyses of the value of information concern the valuation of perfect information (or partial perfect information). However, due

to the ubiquity of uncertainty about ecological structure, processes and states, updating uncertainty with new but less-than-perfect information—that is, sample information—is especially useful in ecology.

In this paper, we have described several variants of *VSI* and a common framework for them, which includes system features, a range of actions, value functions, experimental data and uncertainties. Recognition of these variants expands the usual framing of the value of sample information. In particular, scenarios 4 and 5 and their hybrids include pre-selected actions that occur frequently in ecological management, even if such situations typically are not described in terms of *VSI*.

The value of sample information holds promise in developing areas of ecology such as ecological forecasting and the use of remote sensing in conservation. In forecasting, models can produce the new data needed for Bayesian updating. Thus, understanding expressed by a prior probability distribution can be model-generated (Dakins et al., 1994, 1996) and updated by means of model-based predictions (see e.g. Costello et al., 1998). Value is essentially based on knowledge of the system with and without forecasts, and a comparison of the two values represents the value of the forecasting information.

In the use of remote sensing for conservation, satellite data provide the new information for computation and analysis of *VSI*. For example, a posterior analysis that compares system status or structure with versus without satellite coverage (e.g. Bernknopf et al. 2019; Macauley, 2006) expresses the value of the satellite coverage that accounts for spatial patterns of covariation in multispectral data. In remote sensing as well as forecasting, the value of new information can be determined prospectively through a comparative valuation before the information is collected, thus permitting prioritization of experimental as well as management decisions.

The variants of the value of sample information outlined in scenarios 1–5 may play different roles in actual decision-making. Expected values of sample information (*EVSI* and maximum *EVSI*) are forward-looking analyses that show the potential usefulness of new information before it is obtained. They therefore can help in making decisions about experimentation, as shown in the example about testing a proposed population translocation site for disease (Canessa et al., 2015; see Appendix). On the other hand, the conditional value of sample information (*CVSI*) measures the value of information that is in hand, and hence can be used for looking back at the comparative value of known data. Retrospective analyses with preselected actions (scenarios 4 and 5) are disconnected from active management, in the sense that decisions are made irrespective of the data. However, information value can be useful for experimental design, and there is at least implicit potential for it to guide future management. For example, remote sensing data that allow a new understanding of the ecosystem services provided by a nature preserve (and thus a perceived increase in the preserve's value) might lead to future management or policy changes, if ecosystem service assessments such as those described by the National Ecosystem Services Partnership (2014) are incorporated in decision-making.

As the complexity and expense of resource management grows, deciding whether further investigations are needed will become increasingly important. Wider use of the value of sample information can demonstrate the usefulness of collecting more data, and highlight research and monitoring programmes that are effective in enhancing ecological understanding and outcomes.

ACKNOWLEDGEMENTS

We thank the USGS Science and Decisions Center for support for B.K.W. during preparation of this paper. We are grateful to Fred Johnson and two anonymous reviewers for comments that improved the manuscript. This material is based on work supported by the U.S. Geological Survey under Cooperative Agreement No. G19AC00146. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Geological Survey. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. government.

AUTHORS' CONTRIBUTIONS

B.K.W. and E.D.B. conceived the concept; B.K.W. led the writing of the manuscript. Both authors contributed critically to the drafts and gave final approval for publication.

DATA AVAILABILITY STATEMENT

We used no data, simulated data, or code in preparing this manuscript.

ORCID

Byron K. Williams  <https://orcid.org/0000-0001-7644-1396>

Eleanor D. Brown  <https://orcid.org/0000-0001-7798-830X>

REFERENCES

- Ades, A. E., Lu, G., & Claxton, K. (2004). Expected value of sample information calculations in medical decision making. *Medical Decision Making*, 24, 207–237.
- Back, P.-E., Rosen, L., & Norberg, T. (2007). Value of information analysis in remedial investigations. *Ambio*, 36, 486–493. [https://doi.org/10.1579/0044-7447\(2007\)36\[486:VOIAIR\]2.0.CO;2](https://doi.org/10.1579/0044-7447(2007)36[486:VOIAIR]2.0.CO;2)
- Bal, P., Tulloch, A. I., Chadès, I., Carwardine, J., McDonald-Madden, E., & Rhodes, J. R. (2018). Quantifying the value of monitoring species in multi-species, multi-threat systems. *Methods in Ecology and Evolution*, 9, 1706–1717. <https://doi.org/10.1111/2041-210X.13037>
- Bennett, J. R., Maxwell, S. L., Martin, A. E., Chadès, I., Fahrig, L., & Gilbert, B. (2018). When to monitor and when to act: Value of information theory for multiple management units and limited budgets. *Journal of Applied Ecology*, 55, 2102–2113. <https://doi.org/10.1111/1365-2664.13132>
- Bernknopf, R., Kuwayama, Y., Gibson, R., Blakely, J., Macbee, B., Clifford, T. J., Quayle, B., Epting, J., Hardy, T., & Goodrich, D. (2019). *The cost-effectiveness of satellite earth observations to inform a post-wildfire response*. Working Paper 19–16. Resources for the Future.
- Bhattacharjya, D., Eidsvik, J., & Mukerji, T. (2013). The value of information in portfolio problems with dependent projects. *Decision Analysis*, 10, 341–351. <https://doi.org/10.1287/deca.2013.0277>
- Bickel, J. E. (2008). The relationship between perfect and imperfect information in a two-action risk-sensitive problem. *Decision Analysis*, 5, 116–128. <https://doi.org/10.1287/deca.1080.0118>

- Bolam, F. C., Grainger, M. J., Mengersen, K. L., Stewart, G. B., Sutherland, W. J., Runge, M. C., & McGowan, P. J. (2019). Using the value of information to improve conservation decision making. *Biological Reviews*, 94, 629–647. <https://doi.org/10.1111/brv.12471>
- Canessa, S., Guillera-Aroita, G., Lahoz-Monfort, J. J., Southwell, D. M., Armstrong, D. P., Chadès, I., Lacy, R. C., & Converse, S. J. (2015). When do we need more data? A primer on calculating the value of information for applied ecologists. *Methods in Ecology and Evolution*, 6, 1219–1228. <https://doi.org/10.1111/2041-210X.12423>
- Clemen, R. T., & Reilly, T. (2014). *Making hard decisions with decision tools* (3rd ed.). South-Western CENGAGE Learning.
- Costello, C. J., Adams, R. M., & Polasky, S. (1998). The value of El Niño forecasts in the management of salmon: A stochastic dynamic assessment. *American Journal of Agricultural Economics*, 80, 765–777. <https://doi.org/10.2307/1244062>
- Costello, C., Rassweiler, A., Siegel, D., De Leo, G., Micheli, F., & Rosenberg, A. (2010). The value of spatial information in MPA network design. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 18294–18299. <https://doi.org/10.1073/pnas.0908057107>
- Dakins, M. E., Toll, J. E., & Small, M. J. (1994). Risk-based environmental remediation: Decision framework and role of uncertainty. *Environmental Toxicology and Chemistry*, 13, 1907–1915.
- Dakins, M. E., Toll, J. E., Small, M. J., & Brand, K. P. (1996). Risk-based environmental remediation: Bayesian Monte Carlo analysis and the expected value of sample information. *Risk Analysis*, 16, 67–79.
- Davis, K. J., Chadès, I., Rhodes, J. R., & Bode, M. (2019). General rules for environmental management to prioritise social ecological systems research based on a value of information approach. *Journal of Applied Ecology*, 56, 2079–2090.
- Eidsvik, J., Mukerji, T., & Bhattacharjya, D. (2015). *Value of information in the earth sciences*. Cambridge University Press.
- Forsberg, O. I., & Guttormsen, A. G. (2006). The value of information and salmon farming: Harvesting the right fish at the right time. *Aquaculture Economics and Management*, 10, 183–200.
- Grafton, R. Q., Hilborn, R., Squires, D., Tait, M., & Williams, M. (2010). *Handbook of marine fisheries conservation and management*. Oxford University Press.
- Grantham, H. S., Wilson, K. A., Moilanen, A., Rebelo, T., & Possingham, H. P. (2009). Delaying conservation actions for improved knowledge: How long should we wait? *Ecology Letters*, 12, 293–301. <https://doi.org/10.1111/j.1461-0248.2009.01287.x>
- Graybill, F. A. (1976). *Theory and application of the linear model*. Duxbury Press.
- Hobbs, N. T., & Hooten, M. B. (2015). *Bayesian models: A statistical primer for ecologists*. Princeton University Press.
- Howard, R., & Abbas, A. E. (2015). *Foundations of decision analysis*. Pearson Education.
- James, B. R., & Gorelick, S. M. (1994). When enough is enough: The worth of monitoring data in aquifer remediation design. *Water Resources Research*, 30, 3499–3513. <https://doi.org/10.1029/94WR01972>
- Johnson, F. A., Hagan, G., Palmer, W. E., & Kemmerer, M. (2014). Uncertainty, robustness, and the value of information in managing a population of northern bobwhites. *Journal of Wildlife Management*, 78, 531–539. <https://doi.org/10.1002/jwmg.682>
- Johnson, F. A., Smith, B. J., Bonneau, M., Martin, J., Romagosa, C., Mazzotti, F., Waddle, H., Reed, R. N., Eckles, J. K., & Vitt, L. J. (2017). Expert elicitation, uncertainty, and the value of information in controlling invasive species. *Ecological Economics*, 137, 83–90. <https://doi.org/10.1016/j.ecolecon.2017.03.004>
- Kaelbling, L. P., Littman, M. I., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101, 99–134. [https://doi.org/10.1016/S0004-3702\(98\)00023-X](https://doi.org/10.1016/S0004-3702(98)00023-X)
- Kangas, A. S. (2010). The value of forest information. *European Journal of Forest Research*, 129, 863–874. <https://doi.org/10.1007/s10342-009-0281-7>
- Keisler, J. M., Collier, Z. A., Chu, E., Sinatra, N., & Linkov, I. (2014). Value of information analysis: The state of application. *Environment Systems and Decisions*, 34, 3–23. <https://doi.org/10.1007/s10669-013-9439-4>
- Lane, D. (1989). A partially observable model of decision making by fishermen. *Operations*. <https://doi.org/10.1287/opre.37.2.240>
- Lee, P. M. (1989). *Bayesian statistics: An introduction*. Edward Arnold.
- Macauley, M. K. (2006). The value of information: Measuring the contribution of space-derived earth science data to resource management. *Space Policy*, 22, 274–282. <https://doi.org/10.1016/j.spacepol.2006.08.003>
- Maxwell, S. L., Rhodes, J. R., Runge, M. C., Possingham, H. P., Ng, C. F., & McDonald-Madden, E. (2015). How much is new information worth? Evaluating the financial benefit of resolving management uncertainty. *Journal of Applied Ecology*, 52, 12–20. <https://doi.org/10.1111/1365-2664.12373>
- McDonald, A. D., & Smith, A. D. M. (1997). A tutorial on evaluating expected returns from research for fishery management using Bayes' Theorem. *Natural Resource Modeling*, 10, 185–216. <https://doi.org/10.1111/j.1939-7445.1997.tb00106.x>
- Mood, A., Graybill, F., & Boes, D. (1974). *Introduction to the theory of statistics* (3rd ed.). McGraw-Hill.
- Moore, J. L., & Runge, M. C. (2012). Combining structured decision making and value-of-information analyses to identify robust management strategies. *Conservation Biology*, 26, 810–820. <https://doi.org/10.1111/j.1523-1739.2012.01907.x>
- Moore, J. L., Runge, M. C., Webber, B. L., & Wilson, J. R. (2011). Contain or eradicate? Optimizing the management goal for Australian acacia invasions in the face of uncertainty. *Diversity and Distributions*, 17, 1047–1059. <https://doi.org/10.1111/j.1472-4642.2011.00809.x>
- Morgan, M. G., & Henrion, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press.
- National Ecosystem Services Partnership. (2014). *Federal resource management and ecosystem services guidebook*. Duke University, National Ecosystem Services Partnership. Retrieved from www.nespguidebook.com
- Pearlman, F., Lawrence, C. B., Pindilli, E. J., Geppi, D., Shapiro, C. D., Grasso, M., Pearlman, J., Adkins, J., Sawyer, G., & Tass, A. (2019). *Demonstrating the value of Earth observations—Methods, practical applications, and solutions—Group on Earth Observations side event proceedings*. U.S. Geological Survey Open-File Report 2019–1033. Retrieved from <https://pubs.usgs.gov/of/2019/1033/ofr20191033.pdf>
- Puterman, M. L. (1994). *Markov decision processes: Discrete stochastic dynamic programming*. Wiley.
- Raiffa, H., & Schlaifer, R. O. (1961). *Applied statistical decision theory*. Harvard University Graduate School of Business Administration.
- Runge, M. C., Converse, S. J., & Lyons, J. E. (2011). Which uncertainty? Using expert elicitation and expected value of information to design an adaptive program. *Biological Conservation*, 144, 1214–1223. <https://doi.org/10.1016/j.biocon.2010.12.020>
- Ryder, T. J. (2018). *State wildlife management and conservation*. Johns Hopkins University Press.
- Sahlin, U., Ryden, T., Nyberg, C. D., & Smith, H. G. (2011). A benefit analysis of screening for invasive species - base-rate uncertainty and the value of information. *Methods in Ecology and Evolution*, 2, 500–508. <https://doi.org/10.1111/j.2041-210X.2011.00097.x>
- Silvy, N. J. (2012). *The wildlife techniques manual: Management* (7th ed.). Johns Hopkins University Press.
- Smith, D. R., McGowan, C. P., Daily, J. P., Nichols, J. D., Sweka, J. A., & Lyons, J. E. (2013). Evaluating a multispecies adaptive management framework: Must uncertainty impede effective decision-making? *Journal of Applied Ecology*, 50, 1431–1440. <https://doi.org/10.1111/1365-2664.12145>
- Taylor, A. C., Evans, J. S., & McKone, T. E. (1993). Value of animal test information in environmental control decisions. *Risk Analysis*, 13, 403–412.

- Tomberlin, D. (2010). Endangered seabird habitat management as a partially observable Markov decision process. *Marine Resources Economics*, 25, 93–104.
- Wiles, L. J. (2004). Economics of weed management: Principles and practices. *Weed Technology*, 18, 1403–1407.
- Willan, A. R., & Eckermann, S. (2010). Optimal clinical trial design using value of information methods with imperfect implementation. *Health Economics*, 19, 549–561.
- Willan, A. R., & Pinto, E. M. (2005). The value of information and optimal clinical trial design. *Statistics in Medicine*, 24, 1791–1806. <https://doi.org/10.1002/sim.2069>
- Williams, B. K. (2009). Markov decision processes in natural resources management: Observability and uncertainty. *Ecological Modelling*, 220, 830–840. <https://doi.org/10.1016/j.ecolmodel.2008.12.023>
- Williams, B. K. (2011a). Resolving structural uncertainty in natural resources management using POMDP approaches. *Ecological Modelling*, 222, 1092–1102. <https://doi.org/10.1016/j.ecolmodel.2010.12.015>
- Williams, B. K. (2011b). Adaptive management of natural resources: Framework and issues. *Journal of Environmental Management*, 2011, 1346–1353. <https://doi.org/10.1016/j.jenvman.2010.10.041>
- Williams, B. K. (2015). Integrating external and internal learning in resource management. *Journal of Wildlife Management*, 79, 148–155. <https://doi.org/10.1002/jwmg.814>
- Williams, B. K., Eaton, M., & Breininger, D. R. (2011). Adaptive resource management and the value of information. *Ecological Modelling*, 222, 3429–3436. <https://doi.org/10.1016/j.ecolmodel.2011.07.003>
- Williams, B. K., & Johnson, F. A. (2015a). Value of information in natural resource management: Technical developments and application to pink-footed geese. *Ecology and Evolution*, 5, 466–474. <https://doi.org/10.1002/ece3.1363>
- Williams, B. K., & Johnson, F. A. (2015b). Value of information and natural resources decision making. *Wildlife Society Bulletin*, 39, 488–496. <https://doi.org/10.1002/wsb.575>
- Williams, B. K., & Johnson, F. A. (2018). Value of sample information in dynamic, structurally uncertain resource systems. *PLoS ONE*, 13(6), e0199326. <https://doi.org/10.1371/journal.pone.0199326>
- Yokota, F., & Thompson, K. A. (2004). Value of information literature analysis: A review of applications in health risk management. *Medical Decision Making*, 24, 287–298. <https://doi.org/10.1177/0272989X04263157>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Williams BK, Brown ED. Scenarios for valuing sample information in natural resources. *Methods Ecol Evol*. 2020;11:1534–1549. <https://doi.org/10.1111/2041-210X.13487>