

Cost-effective variable selection in habitat surveys

Douglas C. Braun* and John D. Reynolds

Earth to Ocean Research Group, Department of Biological Sciences, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada

Summary

1. Researchers usually expect to understand the ecological systems better when they examine more variables. However, we cannot measure everything because time and money are limited, so we need to make difficult choices. Decisions are complicated by the fact that variables are often either uninformative or highly correlated, leading to diminishing returns on information with new variables. Correlated variables and diminishing returns on information per variable can be explicitly incorporated with costs of data collection to design cost-effective survey programmes.

2. We develop a step-by-step quantitative protocol to evaluate the cost-effectiveness of survey designs under different cost scenarios to help scientists and managers design cost-effective surveys. We illustrate this protocol using a case study that relates physical stream habitat variables to variation in sockeye salmon spawning populations.

3. We present our protocol by comparing linear regression models containing different combinations of variables representing different survey designs. The steps of the protocol are to (i) eliminate redundant variables, (ii) calculate costs scenarios, (iii) calculate survey performance metrics and (iv) identify and compare a subset of survey designs that maximize effectiveness at a given cost. Survey designs are compared by their ranked performance using R^2 , AICc, average cost-effectiveness ratio and incremental cost-effectiveness ratio.

4. Our case study shows diminishing returns on the information provided by the addition of more variables as survey costs increase. The protocol supports the design of cost-effective monitoring programmes and leads to a general discussion relating changing environmental conditions to survey costs, including the need for clear and measurable objectives, which allow scientific information to be translated into management options.

Key-words: conservation, cost-effectiveness, diminishing returns, efficiency, monitoring, multicollinearity, survey design

Introduction

A common perception among ecologists is that the more variables collected, the more we learn about the ecology of a system. However, the costs of data collection are often prohibitive, and increasing sampling intensity (either the number of samples or the number of variables) leads to diminishing returns on information. For example, additional variables may be uninformative or strongly correlated with variables already considered; these variables provide little new information and can come with significant costs. Although a large body of literature discusses cost-effective sample effort with regard to replication (Conquest 1983; Skalski 1985; Watson 2010), spatial scales (Marignani, Vico & Maccherini 2007; Eigenbrod, Hecnar & Fahrig 2011) and temporal scales (Schreuder,

Hansen & Kohl 1999; Mackenzie & Royle 2005), few studies have quantitatively explored the cost-effectiveness associated with the number of variables measured. This issue is complicated by the fact that the most expensive variables are not necessarily the most informative.

Analyses that quantify the trade-off between the non-monetary benefit of some practice and the associated costs are becoming more prevalent in ecological research (Wintle, Runge & Bekessy 2010). Classical economic theories such as marginal gains theory and return on investment can be used to compare the cost-effectiveness of a range of survey designs (Grantham *et al.* 2008; Underwood *et al.* 2008). For example, Grantham *et al.* (2008) showed diminishing returns in the value of biodiversity survey data. They found that a reduced survey design (i.e. reduced number of survey plots) provided a similar amount of information to the largest survey but cost 25 times less. Similar results were found by Gardner *et al.* (2008) in identifying taxa that indicate ecological integrity and

*Correspondence author. E-mail: dbraun@sfu.ca
Correspondence site: <http://www.respond2articles.com/MEE/>

biodiversity, whereby the least expensive taxa to survey provided the most information on ecological integrity and biodiversity. These results suggest that less expensive survey designs can provide similar levels of accuracy when compared to more expensive surveys.

While the ideas of diminishing returns on information and the application of different cost-effectiveness metrics have been well demonstrated, these studies only examined one monitoring option (i.e. variable or taxa) at a time, whereas potential combinations of variables or taxa have not been considered. A priority for this area of research is to understand how different options interact (McCarthy *et al.* 2010). Many studies have covered the different components of survey designs (Haila & Margules 1996; Caughlan & Oakley 2001; Green *et al.* 2005) and some the whole process (Vos, Meelis & Ter Keurs 2000), but to our knowledge, no study has considered both the financial and statistical costs of different monitoring options. This is important, because of the surge in monitoring precipitated by world-wide concerns about habitat alteration and global climate change (Caughlan & Oakley 2001). Therefore, there is a need for a general step-by-step protocol that can be applied to a range of scenarios to assess the cost-effectiveness of alternative survey designs.

The scope of the problem is illustrated by Fig. 1, which shows that the relationship between the number of variables measured and the information gleaned can be expected to be positive and asymptotic. Figure 1a demonstrates this relationship in a regression framework where R^2 values indicate the fit between some explanatory variable collected during a habitat survey and a response variable of interest (e.g. population metrics). R^2 values for the models increase with the number of variables in each model and so do the cumulative costs (Fig. 1b). The nonlinear relationship between the number of variables and R^2 is attributed to the decrease in unique information as new variables are sampled, whereas the nonlinear relationship between variables and cost is attributed to the increased sampling efficiency (e.g. a visit to a site becomes more cost-effective when more variables are sampled per visit). However, there is inherently no expected relationship between the cost of measuring a variable and the benefit from the amount of information it contains; more expensive surveys will not always provide more information. Combinations of variables that maximize effectiveness for a given cost need to be identified. Then, the cost-effective combinations can be compared by their overall effectiveness (e.g. R^2), and the trade-off between cost and effectiveness (i.e. cost-effectiveness). Additional performance metrics that trade off the amount of information and the statistical cost of including too many variables, such as Akaike Information Criterion (AIC), can be useful in comparing how parsimonious combinations of variables are. Selecting the most appropriate criterion will depend on the objectives of the monitoring programme and the cost-effectiveness analysis.

The objectives of this study are to outline a general protocol for comparing the cost-effectiveness of different habitat survey designs and to illustrate its application. We first present the protocol using four steps, beginning by assessing collinearity among variables and ending by comparing the cost-effective-

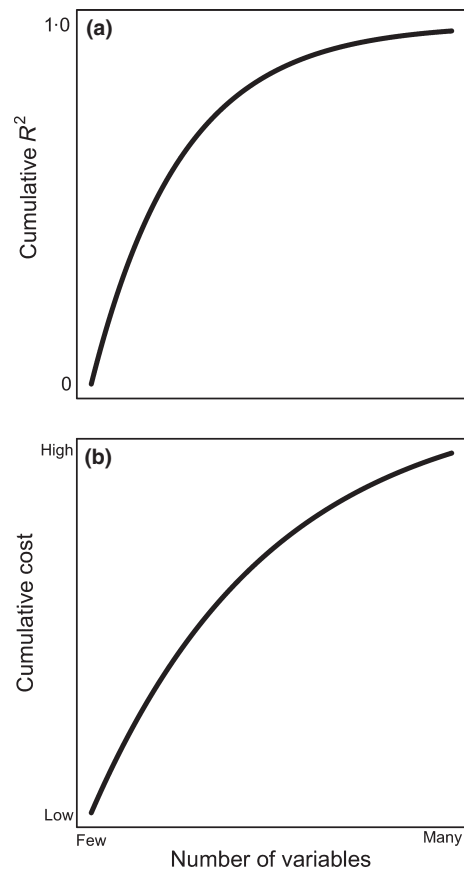


Fig. 1. Hypothetical relationships between: (a) R^2 values and the number of variables in linear regression models, and (b) number of variables sampled and the cumulative cost of sampling.

ness of survey designs (Fig. 2). We then illustrate this protocol using a case study where we evaluate the cost-effectiveness of physical stream habitat variables as predictors of spawning sockeye salmon (*Oncorhynchus nerka*) population densities. The information gained (i.e. effectiveness) from each survey design is measured by relating the habitat variables measured to spawning sockeye population densities using regression models. Survey costs are represented as the money required to measure each variable in each survey design. Statistical costs are incorporated using AIC metrics, which evaluate the trade-off between the number of variables and model fit. Cost-effectiveness metrics are calculated using survey costs and R^2 values. For each step in this method, we discuss the general considerations of their application to designing cost-effective surveys for a diverse range of habitat monitoring scenarios.

Conceptual framework

STEP 1 – COLLINEARITY

Collinearity occurs when variables are highly correlated. Models that contain variables that are highly correlated can have reduced performance of both the overall model and the individual predictors, which can lead to type II errors (Zuur Ieno & Elphick 2009). Statistical redundancy (i.e. when vari-

Step 1 – Collinearity

Use variance inflation factor (cutoff of <3) and cost to determine removal of collinear variables



Step 2 – Cost

Calculate survey costs (time, dollars) for each design
Incorporate sampling efficiencies
Consider statistical costs of additional variables using AIC



Step 3 – Survey performance metrics

Build models from survey designs and calculate effectiveness (R^2), cost-effectiveness metrics (ACER and ICER) and statistical efficiency (AIC)



Step 4 – Compare survey designs

Identify cost-effective set of models using R^2 and cost
Compare survey designs using performance metrics

Fig. 2. The conceptual framework for evaluating cost-effectiveness of survey designs.

ables share much of the same information) decreases sampling efficiency by reducing the amount of information per unit effort despite the cost of measuring more variables. A few simple tools can be used to help determine the degree of collinearity among variables. Most commonly used are correlation matrices, which can display either correlation coefficients between variables or simple scatter-plots. Another alternative is to use the variance inflation factor (VIF) (Zuur, Ieno & Elphick 2009), which is a measure of how correlated a single variable is with all other variables within a set.

There are numerous methods to reduce collinearity, each with pros and cons, see Graham (2003) for review. Two common approaches include (i) eliminate one or more of the collinear variables (Zuur, Ieno & Elphick 2009) and (ii) multivariate approaches, such as principal components analysis, which combines correlated variables into uncorrelated components. The most appropriate method will depend on the objective of the analysis. For this protocol, the objective is to reduce the statistical and sampling redundancy in variables. This can be achieved by eliminating variables. While some explanatory power will be lost, eliminating variables not only reduces statistical redundancy, but also sampling redundancy because the variables do not have to be surveyed again. Approaches where variables remain in the analysis, such as principal components analysis, deal with the statistical redundancy but not sampling

redundancy because the variables still need to be collected to create the principal components. Ridge regression and structural equation modelling can also be used (Graham 2003).

The decision of which variable(s) to remove, based on statistical correlations, can be aided by considering survey cost (Anderson 2008). For example, if two variables have similar correlations among variables but one is less expensive than the other, it would make sense to remove the more expensive variable.

STEP 2 – COST

Survey costs

Survey costs can be defined as either fixed or unfixed. Fixed costs do not vary with sampling intensity, such as planning and designing survey protocols, whereas unfixed costs increase with sampling intensity, which includes time to measure habitat characteristics (similar to variable costs in economics). It is important to explore different fixed cost scenarios because field site location can influence cost-effectiveness. For example, the rank order of cost-effectiveness for surveys at remote sites with larger fixed costs will differ from surveys at more accessible sites with smaller fixed costs, given the same set of survey designs. For metrics that are derived from one variable or sample from the field, the field cost is shared by both metrics, which increases the cost-effectiveness (Fig. 1b). For example, water samples are commonly collected in the field and later subsampled for both nitrate and phosphorus.

Statistical costs

Complex models (large number of variables) pay a statistical cost for over-parameterization (Anderson 2008). Statistical cost can be considered by taking an information theoretic approach. For example, AIC identifies the most parsimonious model, trading off between model fit and complexity (Johnson & Omland 2004). This model selection criterion penalizes models with more parameters.

STEP 3 – SURVEY PERFORMANCE METRICS

Effectiveness

We define effectiveness as how well a survey informs the question(s) being asked. One way to quantify the effectiveness is to build statistical models from survey data and measure their performance using R^2 . For example, multiple regression models represent the relationship between different survey designs and some response variable of interest (i.e. population density, species diversity). How well the model fits the data indicates how well the survey data explain variation in the response variable, which can be quantified by the coefficient of determination, R^2 (Table 1). Therefore, we take the model R^2 as survey effectiveness. The general framework we present can be used with any type of linear model ranging from simple (e.g. linear regression) to complex (e.g. multilevel mixed-effects models). For our

Table 1. Survey evaluation metrics

Survey evaluation metric	Calculation	Elements	Definition	When to use	References
R^2	$R^2 = 1 - \frac{SS_{\text{residual}}}{SS_{\text{total}}}$	Fit	Model goodness-of-fit or the amount of information in each survey design	To accompany cost-effectiveness metrics (only considers effectiveness)	Johnson & Omland 2004
AICc	$AICc = -2\ln(L) + \frac{2p(p+1)}{n-p-1}$	Fit, model complexity, and sample size	Trade-off between model fit and the number of variables – model parsimony	To accompany cost-effectiveness metrics (costs only considered as number of variables)	Johnson & Omland 2004
Average cost-effectiveness ratio – ACER	$ACER = \frac{TC}{E}$	Fit and cost	Total cost per unit effectiveness	When comparing multiple options at once – e.g. evaluating options for a new monitoring programme	Laska, Meisner, and Siegel 1997
Incremental cost-effectiveness ratio – ICER	$ICER_i = \frac{VC_i - VC_c}{E_i - E_c}$	Fit and cost	Change in cost per change in effectiveness of a new practice relative to current practice	When explicitly comparing alternative options to a current programme – e.g. evaluating potential changes to an existing monitoring programme	Briggs & Fenn, 1997

SS, linear model sum of squares.

example, we use multiple regression because it is the most appropriate given our study design and is familiar to ecologists.

Cost-effectiveness metrics

Cost-effectiveness analyses (CEA) evaluate the trade-off between the financial costs and the non-monetary benefits of some practice. CEAs are used extensively in health economics to aid in making decisions about whether or not a treatment is worth implementing (Donaldson, Currie & Mitton 2002). For example, the average cost-effectiveness ratio (ACER) is a metric often used to trade off the cost of a treatment or screening with some health benefit or increased detection probability (Hoch & Dewa 2008). This cost-effectiveness approach can be applied to ecological monitoring programmes by using survey costs along with some measure of monitoring effectiveness and will identify which survey option is the most cost-effective overall (Table 1). ACER is calculated as:

$$ACER = \frac{TC}{E} \quad \text{eqn 1}$$

where TC is the total survey cost (fixed + unfixed costs), and E is the effectiveness (R^2) of the survey. ACER is useful when the objective of the CEA is to develop new monitoring programmes because it provides absolute cost-effectiveness values (Table 1).

To compare the cost-effectiveness of two different survey options, we use the incremental cost-effectiveness ratio (ICER) (Hoch & Dewa 2008) (Table 1), calculated as:

$$ICER_i = \frac{VC_i - VC_c}{E_i - E_c} \quad \text{eqn 2}$$

where VC is the unfixed survey cost (fixed cost not included), E is the effectiveness (i.e. R^2) of the survey for the current survey design c , and an alternative survey design i . ICER is useful when the objective of the CEA is to re-evaluate existing programmes because it compares alternatives that are either more or less intensive than the current design (Table 1). While ICER identifies *which alternative potential change* is more cost-effective than another one, it does not indicate whether *the change itself* is cost-effective compared to the current design. We agree with the recommendation that the two cost-effectiveness metrics should be used together to make decisions (Laska, Meisner & Siegel 1997).

Statistical efficiency

We define statistical efficiency as the trade-off between model fit and complexity and approximated by Akaike Information Criteria (AIC), which incorporates a penalty for model complexity determined by the number of parameters included in the model (Table 1). In an AIC framework, simple models are deemed more efficient than complex models for a given level of fit. This approach is similar to cost-effectiveness metrics (trade-off between financial cost and model fit) except that here the cost is a statistical one. For each model, AIC is calculated as:

$$AIC = 2p - 2\ln(L)$$

where p is the number of parameters and L is the likelihood function. An additional penalty for small sample sizes is usually added:

$$AICc = AIC + \frac{2p(p+1)}{n-p-1}$$

where n is the number of observations. Anderson (2008) recommends using AICc over AIC, regardless of sample size, because the two metrics converge when sample size is large relative to the number of variables. ΔAIC is the difference in AIC values between the model with the lowest AIC value and model i . Larger values indicate less efficient models relative to each other. AIC is useful when the objective of the CEA is to identify parsimonious models that also provide reliable parameter estimates, compared to R^2 which only evaluates the fit of a model and does not penalize a model for complexity (Johnson & Omland 2004) (Table 1).

Other metrics

Although R^2 and AIC are commonly used to evaluate models, other metrics can be used to measure the effectiveness, such as intercept and slope of observed vs. fitted regressions (Pineiro *et al.* 2008). It may also be of greater interest to evaluate predictive power (root mean predicted squared error), as opposed to explanatory power (R^2); this will depend on the objectives of the survey. In addition, statistical efficiency can be measured using Bayesian information criterion and likelihood ratio testing (Johnson & Omland 2004). The performance of survey designs with regard to effectiveness, cost-effectiveness and statistical efficiency will depend on the cost and effectiveness metrics used. There is no one way to measure the effectiveness. Cost or effectiveness metrics can be chosen to reflect the priorities of surveyors, with the understanding that this choice can affect the decisions about which surveys designs are most cost-effective.

STEP 4 – COMPARE SURVEY DESIGNS

The final step is to identify a cost-effective subset of survey designs that maximizes explanatory power (i.e. statistical effectiveness) for a given cost. In our example, this cost-effective set is a subset of all possible models with the highest R^2 for a given cost (closed circles in Fig. 3). Once the cost-effective subset of models is identified, we can compare models using the evaluation metrics outlined in Step 3 and Table 1. There can be more than one cost-effective survey design in the subset, according to the number of cost scenarios. Cost-effective surveys are identified using R^2 values as effectiveness and cost is the price. The number of cost-effective models and shape of the relationship between R^2 and the number of variables or cumulative cost will depend on how correlated variables are with each other and the amount of variance they explain in the dependent variable.

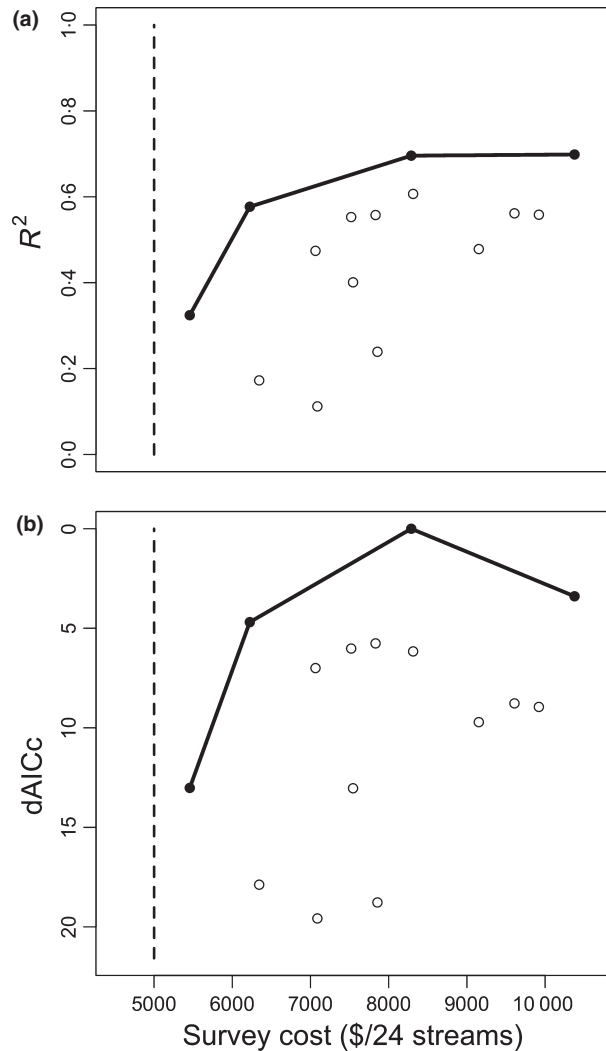


Fig. 3. (a) R^2 vs. cost and (b) $\Delta AICc$ vs. cost for all 15 linear regression models relating stream habitat variables to sockeye salmon spawning density. Closed circles are the cost-effective set of models and dashed line indicates the fixed cost. Note: cost-effective set of models identified by R^2 values and cost may not include the most parsimonious model according to AICc values.

Once a cost-effective subset of models is identified, ranking and comparing different the performance metrics against one another can aid in prioritizing survey designs that management objectives. Models within the cost-effective subset can vary in their cost (dollars), effectiveness (R^2), cost-effectiveness (ACER and ICER) and statistical efficiency (ΔAIC).

CASE STUDY: SOCKEYE SALMON BREEDING HABITAT STATUS

We illustrate our protocol with data that were collected to assess stream habitat quality for spawning sockeye salmon (Braun & Reynolds 2011). For simplicity, we selected a subset of data from the original study. The subset consisted of six commonly measured abiotic variables from 24 streams measured from June to August, 2007, in the Stuart watershed of the Fraser River, British Columbia, Canada. The variables

were measured within a single study section in each stream, whose length was determined as 30 times the width of the stream at the high water mark. Our response variable was the density of adult spawning fish in each section, as determined by surveys by the Canadian Department of Fisheries and Oceans. The overarching goal of our study was to identify habitat variables that could be used to inform management on the status of stream habitat for sockeye salmon. Therefore, we identified habitat variables that are important to spawning sockeye density, because they may mediate predation risk experienced by spawning adults (Braun & Reynolds 2011). A variable's effectiveness was determined as the proportion of variation it explained in salmon population densities.

Step 1 – Collinearity

For this example, we assess the collinearity for six variables using the criteria from Step 1 (Table 2) based on the procedure outlined by Zuur, Ieno & Elphick (2009). We use VIF scores as our primary criteria for eliminating variables accompanied by financial cost (Zuur, Ieno & Elphick 2009). If VIF scores for two variables were close (i.e. within 2), we use cost to determine which variable to eliminate. This is an arbitrary but clear criterion for bringing in cost to eliminate variables, and it fits with the overarching objective of selecting cost-effective variables. After removing a variable, we iteratively assess the remaining variables until all variables have VIF scores of < 3 (Zuur, Ieno & Elphick 2009). That procedure reduced our candidate set of variables from 6 to 4 (Table 2), which were used to construct the regression models.

Step 2 – Cost

For our primary analysis, we consider a moderate fixed cost of \$5000, which is an estimate of the total fixed costs for vehicles, boats, fuel, and field logistics for the duration of our surveys. A second analysis compares the ACERs using three fixed costs, low (\$0) moderate (\$5000) and high (\$10 000).

Unfixed costs are incorporated as time and money spent measuring each habitat variable. Time was recorded during field surveys, and a mean time per stream is calculated. The mean time is then converted into dollars according to an

Table 2. The variance inflation factors (VIF) used to drop candidate variables. The variable with the highest VIF score is dropped until all remaining variables have a score of < 3 .

Variables	Variance inflation factor		
	Round 1	Round 2	Round 3
Maximum water depth	6.5	4.7	1.1
Mean water depth	4.02	3.9	Dropped
Cross-section area	4.5	Dropped	–
% Pools	1.6	1.7	1.6
Woody debris	1.3	1.3	1.2
Undercut banks	1.4	1.4	1.4

In round 2 and 3, cross-section area and mean water depth were dropped instead of maximum water depth because of cost.

hourly wage based on the level of expertise required to measure the metrics. We account for sampling efficiencies by splitting the cost of measuring a variable into its potential shared and unique costs. For example, variables taken at transects share the cost of setting up those transects. We calculate the total cost of measuring each model (i.e. survey scenario) by summing the total labour costs with equipment costs, less any shared costs. Statistical costs of different survey designs were considered using AICc (small sample sizes).

Step 3 – Survey evaluation metrics

We built regression models that describe sockeye salmon densities using all possible combinations of abiotic variables from the candidate set for a total of 15 models. From these models, we calculated four survey evaluation metrics (Table 1). We extracted each model's R^2 as our metric of effectiveness. ACER was calculated as in eqn 1 where TC is the total survey cost (fixed + unfixed costs) and E is the R^2 from the model representing the survey design. ICER was calculated as in eqn 2 where VC is the variable survey cost, E is the R^2 from the model representing the current survey design c , and an alternative survey design i . Statistical efficiency was measured as ΔAICc . Model assumptions were evaluated using standard regression diagnostics, and we used plots to visually assess the normality of residuals and heteroscedasticity. Salmon densities were log transformed to better meet the assumptions of linear regression models.

Step 4 – Compare survey designs

Cost-effective subset of models. First, we identified a cost-effective subset of four survey designs by iteratively assessing the R^2 values for each model from the least to most expensive using the moderate fixed cost scenario of \$5000. If a model had an R^2 value that was higher than all models that had equal or higher costs, it was deemed cost-effective; it gives 'the best bang for your buck' (Fig. 3). Consider a budget of \$7500 to survey 24 streams. The most effective survey option would be to measure undercut banks and maximum water depth, which costs \$6224 and gives an R^2 of 0.58 (Table 3). More expensive survey options but still within the \$7500 budget would provide less information. In fact, surveys that measure woody debris (\$7088) or % pools (\$7064) produce lower R^2 values of 0.11 and 0.47, respectively. We show this graphically in Fig. 3 (closed circles are four cost-effective survey designs out of a possible 15 combinations of survey designs). Although the highest R^2 value is achieved with the most expensive survey design, which includes all four variables, Fig. 3 also shows the diminishing returns on information with increasing cost.

Comparison of evaluation metrics. Once the cost-effective subset of models was identified, we ranked the performance of survey designs under each of the four criteria (R^2 , ACER, ICER, and ΔAICc) (Table 3). All criteria ranked survey designs differently, that is to say there was no clear 'best' survey design. None of the survey designs ranked first more than once.

Table 3. The cost-effective set of survey designs ranked by their performance in each of the four performance criteria. Survey cost is the total estimated cost of surveying 24 streams (fixed cost + variable cost) using a moderate fixed cost of \$5000. Effectiveness is the R^2 value from each linear regression model (i.e. survey design). Cost-effectiveness metrics are ACER and ICER. Statistical efficiency is represented by Δ AICc. The ranked orders for each criterion are in brackets

Survey design/model	Survey cost (\$)	R^2	Δ AICc	ACER	ICER
Undercut banks + Maximum water depth + % Pools + Woody debris	10 376	0.698 (1)	3.4 (2)	14 865 (3)	1 044 000 (4)
Undercut banks + Maximum water depth + % Pools	8288	0.696 (2)	0.0 (1)	11 908 (2)	173 455 (3)
Undercut banks + Maximum water depth	6224	0.577 (3)	4.7 (3)	10 787 (1)	3036 (2)
Undercut banks	5456	0.324 (4)	13.0 (4)	16 840 (4)	1407 (1)

ACER, average cost-effectiveness ratio; ICER, incremental cost-effectiveness ratio.

However, the survey that included undercut banks was the fourth ranked design for three out of the four criteria. ACER and ICER did not agree on any of their rankings of the survey designs. The lack of agreement among performance metrics should be of little concern as long as decision-makers are clear about which evaluation metric(s) best suit their CEA objectives and their monitoring needs or constraints. Ranking survey designs is useful because it simplifies the selection process but it also masks nonlinearities in the performance metrics and therefore should always be accompanied by the actual values. For example, the first and second survey designs ranked according to effectiveness have R^2 values of 0.698 and 0.696; they have equal effectiveness but are ranked differently. Ranks accompanied by the actual metric values will help elucidate these nonlinearities.

Comparison of fixed costs scenarios. We explored how different fixed costs influence the ACER by performing a second analysis, which compared two other fixed cost scenarios of low (\$0) and high (\$10 000) with the moderate fixed cost (\$5000) scenario (Table 4). The rankings for all other metrics we used are independent of fixed cost as well as selection of the cost-effective set of survey designs. As fixed costs increased relative to unfixed costs, more-intense surveys ranked higher than less-intense surveys. The top-ranked models for ACER differed for the three fixed cost scenarios, whereby at low fixed cost, the top-ranked model was the least expensive in the cost-effective set (maximizing R^2 for a given cost); at the moderate cost, the top-ranked model was the second cheapest; and at the high-cost scenario, the third most expensive survey was ranked the highest (Table 4). Note that it would take a fixed cost of c. \$600 000 for the most intense survey to be ranked the highest. This is

because the difference in R^2 values between the two most intense surveys is extremely small (R^2 for undercut banks + maximum water depth + % pools = 0.696, for undercut banks + maximum water depth + % pools + woody debris = 0.698). This suggests the importance of accounting for different fixed costs, especially if surveys are conducted across sites that vary in remoteness or if different infrastructure is required (e.g. accommodation and transportation).

Robustness of cost-effective subset to variation in sample size. Changes in the number of sites surveyed could influence the survey designs selected for the cost-effective set, which would influence the rankings for surveys under each of our criteria. We address the first part of this issue by comparing the cost-effective sets derived from subsampled data sets to the set selected using the full data set ($N = 24$). We subsampled the full data set at sample sizes ranging from 12 to 21 sites, 1000 times for each sample size. For each new data set, we reran the 15 models and then used the new costs (fewer samples led to reduced costs) and model R^2 to select a cost-effective set (as in step 4). This produced 1000 cost-effective sets for each sample sizes. This allowed us to determine the influence of sample size on (i) the pattern of diminishing returns on information with cost and (ii) our confidence that the cost-effective subset is the true set. The true set is the cost-effective subset if the entire population of streams were sampled. We also counted the number of new survey designs that entered into the cost-effective sets for each sample size.

The pattern of diminishing returns on information with cost was maintained at all sample sizes (Fig. 4). Figure 5 shows how often the cost-effective subsets from the reduced data set were identical to that observed with the full data set. As sample size

Table 4. The cost-effective set of survey designs ranked by their average cost-effectiveness ratio (ACER) performance for three fixed cost scenarios: low, moderate and high

Survey design/model	Variable costs (\$)	R^2	ACER		
			Low (\$0)	Moderate (\$5000)	High (\$10 000)
Undercut banks + Maximum water depth + % Pools + Woody debris	5376	0.698	7702 (4)	14 865 (3)	22 029 (3)
Undercut banks + Maximum water depth + % Pools	3288	0.696	4724 (3)	11 908 (2)	19 092 (1)
Undercut banks + Maximum water depth	1224	0.577	2121 (2)	10 787 (1)	19 452 (2)
Undercut banks	456	0.324	1407 (1)	16 840 (4)	32 272 (4)

increases, the confidence in the cost-effective subset increases linearly; in other words, there is greater confidence in cost-effective sets derived from larger sample sizes. The cost-effective subsets for sample sizes > 18 always included the cost-effective subset from the full data set. This suggests that when different cost-effective sets were identified, it was because new survey designs were added, rather than because the original ones were being removed. Furthermore, the number of new survey designs that entered the cost-effective set decreased with sample size. This suggests that as more sites are surveyed, fewer survey designs would be included in the cost-effective set. This can inform decisions about how many locations should be sampled.

Discussion

To our knowledge, this is the first step-by-step protocol for identifying cost-effective variables to be used in habitat surveys by trading off explanatory power and cost (financial and statistical). Results from our case study demonstrate that collecting more variables is not always better. Different fixed cost scenarios influenced the ACER whereby increases in fixed costs led to more-intense survey designs being more cost-effective.

The logistic and statistical issues we tackle are common to all habitat and biodiversity surveys. The underlying concepts to each of our steps are independent of scale, habitat, cost of survey, response variable metrics and statistical methods used to evaluate models. These steps can be applied to new data collected as part of a pilot project, with the goal of designing a new monitoring programme, or they can be used to examine the performance of an existing programme. Evaluating the cost-effectiveness of existing monitoring programmes can help

streamline the use of limited resources. It can also be used to adapt existing programmes by eliminating variables that are no longer or never were informative.

Many monitoring programmes are long-term. This poses a challenge to cost-effective designs because variables that are redundant or uninformative today could become important tomorrow (Wintle, Runge & Bekessy 2010). Twenty years ago, a few coral reef ecologists considered measuring pH and a few amphibian ecologists measured UV-B. Furthermore, advances in technology may lead to greater sampling efficiency, improving cost-effectiveness. Therefore, any programme can be expected to require periodic re-evaluation according to the criteria of interest, such as cost-effectiveness. A rule of thumb that may be helpful when gazing into the future could be 'if it's cheap, measure it'.

Our framework assumes that survey objectives are measurable, which is essential, but may not always be true (Haila & Margules 1996; Vos, Meelis & Ter Keurs 2000; Green *et al.* 2005). For example, we may want to assess 'biodiversity', but that will be impossible unless someone translates this objective into specific indices that can be measured, such as species richness or evenness. Use of a framework such as this, which evaluates cost-effectiveness of surveys, forces practitioners to tackle this problem head-on.

Future work on this topic might include incorporating observation error as a performance criterion. The trade-off between variable number and sample replication could also be considered. This also would inform the discussion about when to select efficient designs over effective designs. More efficient designs allow for money to be spent on sampling more replicates, which will influence the power of the study and level of inference that can be drawn (Molloy *et al.* 2010).

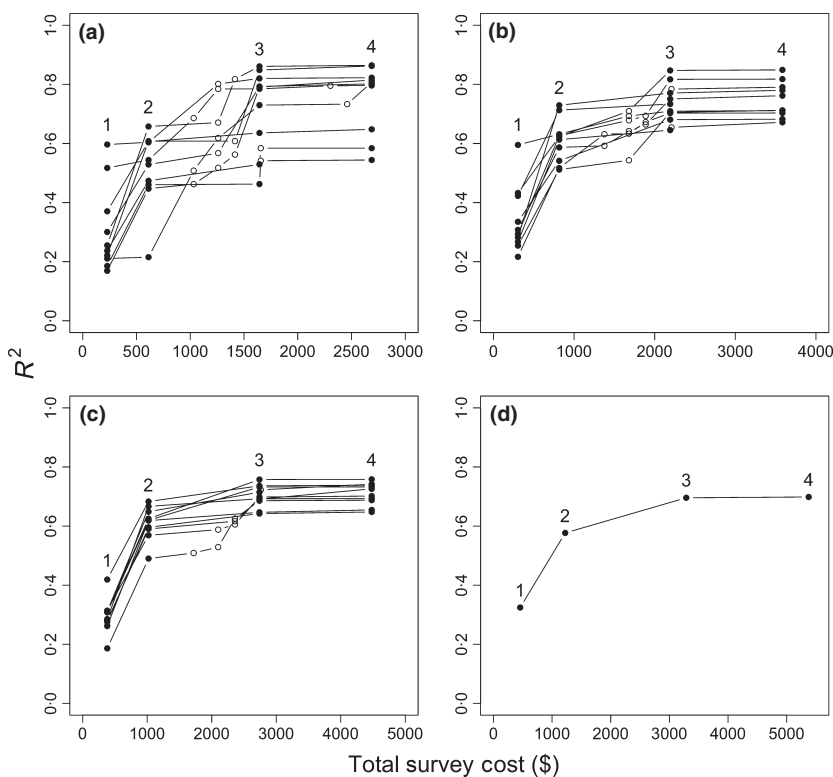


Fig. 4. The cost-effective sets for the first 10 iterations (of 1000) from subsampled datasets at sample sizes of: (a) 12, (b) 16 and (c) 20 and (d) the full dataset ($n = 24$). Closed circles are models found in the cost-effective set identified using the full dataset and open circles are new models identified by the full dataset. Labels: 1 = undercut banks, 2 = undercut banks + maximum water depth, 3 = undercut banks + maximum water depth + % pools, and 4 = undercut banks + maximum water depth + % pools + woody debris. No fixed costs were included.

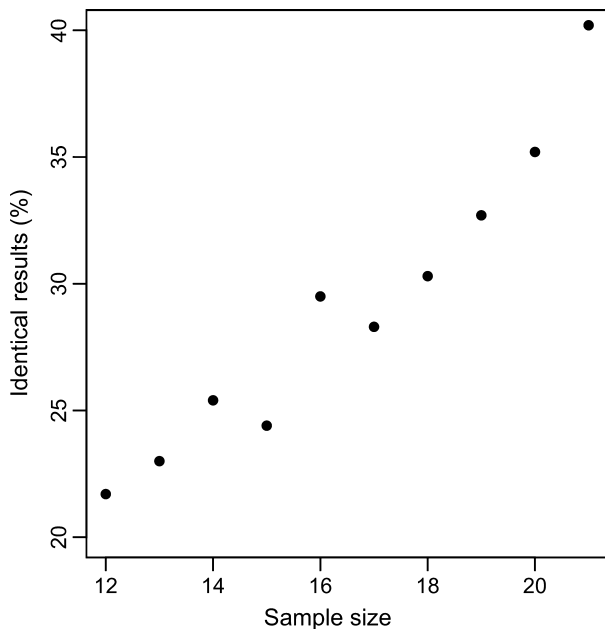


Fig. 5. Variation in the percent of iterations where subsampled data produced identical cost-effective subsets to the full dataset. One thousand iterations were run per sample size (ranging from 12 to 21). For each iteration, all 15 models were re-run using a unique dataset based on random draws (without replacement) from the full dataset. We did not use sample sizes > 21 because they do not produce 1000 unique datasets.

In conclusion, we hope that these protocol considerations will encourage researchers and managers to think about the idea that collecting more variables is not always better. More specifically, we hope this protocol will help identify cost-effective survey designs, for a given set of objectives and budget, through direct comparisons of cost and effectiveness among variables. Understanding how to evaluate cost-effectiveness of survey designs can improve decision-making and support the design of sustainable monitoring programmes, thereby making better use of limited resources for conservation and management efforts.

Acknowledgements

We thank the Fraser Salmon Watershed Programme, the Natural Sciences and Engineering Research Council of Canada, the Watershed Watch Salmon Society, the Northern Scientific Training Programme, and the Canadian Department of Fisheries and Oceans (DFO) for financial support. We appreciate help from DFO staff, including David Patterson, Herb Herunter, Erland MacIsaac, Tracy Cone, Dennis Klassen, Kerry Parish, and Keri Benner for logistical support and valuable advice on the design of the case study that was used here. We also appreciate the field support from Jan Verspoor, Mike Sawyer, Rudi Verspoor, Krista Braun and Craig Losos. We also thank Andrew Cooper, Emily Darling, Brett Favaro, Joel Harding, Scott Hinch, Morgan Hocking, Jenny Linton, Phil Molloy, Craig Orr, Bernie Roitberg and Jan Verspoor for comments on the study.

References

Anderson, D.R. (2008) *Model Based Inference in the Life Sciences: A Primer on Evidence*. Springer, New York, NY.

Braun, D.C. & Reynolds, J.D. (2011) Relationships between habitat characteristics and breeding population densities in sockeye salmon. *Canadian Journal of Fisheries & Aquatic Sciences*, **68**, 758–767.

Briggs, A. & Fenn, P. (1997) Trying to do better than average: a commentary on 'statistical inference for cost-effectiveness ratios'. *Health Economics*, **6**, 491–495.

Caughlan, L. & Oakley, K. (2001) Cost considerations for long-term ecological monitoring. *Ecological Indicators*, **1**, 123–134.

Conquest, L. (1983) Assessing the statistical effectiveness of ecological experiments: utility of the coefficient of variation. *The International Journal of Environmental Studies*, **20**, 209–221.

Donaldson, C., Currie, G. & Mitton, C. (2002) Cost effectiveness analysis in health care: contraindications. *British Medical Journal*, **325**, 891–894.

Eigenbrod, F., Hecnar, S.J. & Fahrig, L. (2011) Sub-optimal study design has major impacts on landscape-scale inference. *Biological Conservation*, **144**, 298–305.

Gardner, T.A., Barlow, J., Araujo, I.S., Avila-Pires, T.C., Bonaldo, A.B., Costa, J.E., Esposito, M.C., Ferreira, L.V., Hawes, J., Hernandez, M.I.M., Hoogmoed, M.S., Leite, R.N., Lo-Man-Hung, N.F., Malcolm, J.R., Martins, M.B., Mestre, L.A.M., Miranda-Santos, R., Overal, W.L., Parry, L., Peters, S.L., Ribeiro-Junior, M.A., da Silva, M.N.F., Motta, C.d.S. & Peres, C.A. (2008) The cost-effectiveness of biodiversity surveys in tropical forests. *Ecology Letters*, **11**, 139–150.

Graham, M.H. (2003) Confronting multicollinearity in ecological multiple regression. *Ecology*, **84**, 2809–2815.

Grantham, H.S., Moilanen, A., Wilson, K.A., Pressey, R.L., Rebelo, T.G. & Possingham, H.P. (2008) Diminishing return on investment for biodiversity data in conservation planning. *Conservation Letters*, **1**, 190–198.

Green, R.E., Balmford, A., Crane, P.R., Mace, G.M., Reynolds, J.D. & Turner, R.K. (2005) A framework for improved monitoring of biodiversity: responses to the world summit on sustainable development. *Conservation Biology*, **19**, 56–65.

Haila, Y. & Margules, C. (1996) Survey research in conservation biology. *Ecography*, **19**, 323–331.

Hoch, J.S. & Dewa, C.S. (2008) A clinician's guide to correct cost-effectiveness analysis: think incremental not average. *Canadian Journal of Psychiatry*, **53**, 267–274.

Johnson, J. & Omland, K. (2004) Model selection in ecology and evolution. *Trends in Ecology and Evolution*, **19**, 101–108.

Laska, E.M., Meisner, M. & Siegel, C. (1997) The usefulness of average cost-effectiveness ratio. *Health Economics*, **6**, 497–504.

Mackenzie, D.I. & Royle, J.A. (2005) Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*, **42**, 1105–1114.

Marignani, M., Vico, E.D. & Maccherini, S. (2007) Spatial scale and sampling size affect the concordance between remotely sensed information and plant community discrimination in restoration monitoring. *Biodiversity and Conservation*, **16**, 3851–3861.

McCarthy, M.A., Thompson, C.L., Hauser, C., Burgman, M.A., Possingham, H.P., Moir, M.L., Tiensin, T. & Gilbert, M. (2010) Resource allocation for efficient environmental management. *Ecology Letters*, **13**, 1280–1289.

Molloy, P.P., Anticamara, J.A., Rist, J.L. & Vincent, A.C.J. (2010) Frugal conservation: what does it take to detect changes in fish populations? *Biological Conservation*, **143**, 2532–2542.

Pineiro, G., Perelman, S., Guerschman, J.P. & Paruelo, J.M. (2008) How to evaluate models: observed vs. predicted or predicted vs. observed? *Ecological Modelling*, **216**, 316–322.

Schreuder, H.T., Hansen, M. & Kohl, M. (1999) Relative costs and benefits of a continuous and periodic forest inventory in Minnesota. *Environmental Monitoring and Assessment*, **59**, 135–144.

Skalski, J.R. (1985) Construction of cost-functions for tag-recapture research. *Wildlife Society Bulletin*, **13**, 273–283.

Underwood, E.C., Shaw, R.M., Wilson, K.A., Kareiva, P., Klausmeyer, K.R., McBride, M.F., Bode, M., Morrison, S.A., Hoekstra, J.M. & Possingham, H.P. (2008) Protecting biodiversity when money matters: maximizing return on investment. *PLoS ONE*, **3**, e1515. doi: 10.1371/journal.pone.0001515.

Vos, P., Meelis, E. & Ter Keurs, W.J. (2000) A framework for the design of ecological monitoring programs as a tool for environmental and nature management. *Environmental Monitoring and Assessment*, **61**, 317–344.

Watson, D. (2010) Optimizing inventories of diverse sites: insights from Barro Colorado Island birds. *Methods in Ecology and Evolution*, **1**, 280–291.

Wintle, B., Runge, M. & Bekessy, S. (2010) Allocating monitoring effort in the face of unknown unknowns. *Ecology Letters*, **13**, 1325–1337.

Zuur, A.F., Ieno, E.N. & Elphick, C.S. (2009) A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, **1**, 3–14.

Received 4 March 2011; accepted 30 July 2011
Handling Editor: David Warton