Ecological Applications, 25(2), 2015, pp. 441–450

Evidence-based tool surpasses expert opinion in predicting probability of eradication of aquatic nonindigenous species

David Drolet,[1,4] Andrea Locke,[2] Mark A. Lewis,[3] and Jeff Davidson[1]

[1]*Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, 550 University Avenue, Charlottetown, Prince Edward Island C1A 4P3 Canada*

[2]*Science Branch, Department of Fisheries and Oceans Canada, Gulf Fisheries Centre, P.O. Box 5030, Moncton, New Brunswick E1C 9B6 Canada*

[3]*Departments of Mathematical Sciences and Biological Sciences, University of Alberta, 455B CAB, Edmonton, Alberta T6G 2G1 Canada*

Abstract. The main objective of evidence-based management is to promote use of scientific data in the decision-making process of managers, with data either complementing or replacing expert knowledge. It is expected that this will increase the efficiency of environmental interventions. However, the relative accuracy and precision of evidence-based tools and expert knowledge has seldom been evaluated. It is therefore essential to verify whether such tools provide better decision support before advocating their use. We conducted an elicitation survey in which experts were asked to (1) evaluate the influence of various factors on the success of eradication programs for aquatic nonindigenous species and (2) provide probabilities of success for real case studies for which we knew the outcome. The responses of experts were compared with the results and predictions of a newly developed evidence-based tool: a statistical model calibrated with a meta-analysis of case studies designed to evaluate probability of eradication. Experts and the model generally identified the same factors as influencing the probability of success. However, the model provided much more accurate estimates for the probability of eradication than expert opinion, strongly suggesting that an evidence-based approach is superior to expert knowledge in this case. Uncertainty surrounding the predictions of the evidence-based tool was similar to among-expert variability. Finally, a model based on $\geq$30 case studies returned more accurate predictions than expert opinion. We conclude that decision-making processes based on expert judgment would greatly benefit from incorporating evidence-based tools.

Key words: conservation; elicitation survey; eradication; evidence-based management; expert knowledge; nonindigenous species.

Introduction

Managers' experience and expert opinion have played, and will continue to play, a fundamental role in environmental management decision making (Fazey et al. 2006). In situations where empirical information is scarce, expert knowledge provides an inexpensive and quick alternative to data acquisition. For example, experts can help assess the severity of anthropogenic impacts (Martin et al. 2005, O'Neill et al. 2008, Whitfield et al. 2008, Donlan et al. 2010), develop species distribution models (Pearce et al. 2001, Yamada et al. 2003, Johnson and Gillingham 2004, MacMillan and Marshall 2006, Murray et al. 2009) and make plans for land use (Clevenger et al. 2002, Geneletti 2005). However, in the absence of independent empirical tests of expert models, it is difficult to assess their accuracy and the validity of decisions they inform.

Over a decade ago, a call was made for the implementation of evidence-based environmental management (Pullin and Knight 2001, Sutherland et al. 2004), coming from the realization that managers often use limited scientific evidence in their decision-making processes, even when such evidence is available (Pullin et al. 2004, Pullin and Knight 2005, Cook et al. 2010). A great deal has been done to increase the use of evidential approaches: several systematic reviews, critically evaluating the effectiveness of common management techniques, have been published (e.g., Stewart et al. 2005, 2009, Stewart and Pullin 2008, Smith et al. 2010, 2011), meta-analyses identified factors influencing success of actions (e.g., Fischer and Lindenmayer 2000, Brooks et al. 2006, Padgee et al. 2006, Smith et al. 2010, Pluess et al. 2012*a*, *b*), and models have been developed to predict the qualitative response of a system to different interventions (Newton et al. 2007, Raymond et al. 2011, Holzkämper et al. 2012). It is expected that complementing, or replacing, experiential with evidential approaches will improve the overall efficiency of

Manuscript received 27 January 2014; revised 23 June 2014; accepted 24 July 2014. Corresponding Editor: E. Nelson.

[4] Present address: Fisheries and Oceans Canada, Institut Maurice-Lamontagne, 850 route de la Mer, Mont-Joli, Quebec G5H 3Z4 Canada.
E-mail: david.drolet@dfo-mpo.gc.ca

conservation and environmental management interventions (Sutherland et al. 2004). However, improved efficiency will only occur if data-based tools are more accurate than expert-based approaches.

In a few cases, the accuracy of expert opinion was evaluated against empirical evidence. Whitfield et al. (2008) found that simple metrics, such as distance at which birds react to human presence, could be well approximated by expert evaluations. This result was obtained as part of the broader objective of evaluating the quality of evidence used to set buffer zones surrounding nests. McCarthy et al. (2004) found that experts consistently overestimated the probability of decline of populations when compared to predictions from population dynamics models. However, the models used were for fictitious populations, so the predictions could not be compared with population trends in nature. To our knowledge, only two studies were specifically designed to compare expert- and data-based approaches, both in the field of habitat modeling. Clevenger et al. (2002) found that expert opinion models were less accurate than empirical models in predicting road-crossing areas by wildlife. Finally, Pearce et al. (2001) found that incorporating expert opinion in empirical models generally did not improve distribution models for various Australian species. In addition, models based solely on expert opinion performed worse than various data-derived models.

Proliferation of nonindigenous biota is a major conservation issue that can also have significant economic impacts (Mack et al. 2000). The decision to intervene and try to eradicate or control a particular species should be driven by the ratio of costs to benefits from both an ecological and economic perspective (e.g., McEnnulty et al. 2001). Tools to evaluate (1) risk associated with a species (e.g., Miller et al. 2007, Leung et al. 2012) and (2) costs of interventions (e.g., Martins et al. 2006, Crombie et al. 2008) are being developed. The efficacy of control techniques has been investigated through systematic reviews for particular species (e.g., Tyler et al. 2006, Roberts and Pullin 2008) and meta-analysis (Pluess et al. 2012*a*, *b*, Tobin et al. 2014). While these tools can be used to evaluate chances of success of eradication programs, the probability of success of different interventions is still usually evaluated through expert opinion.

To increase the use of evidence in the planning of intervention targeting aquatic nonindigenous species (ANIS), we recently developed the software MIPE (Model Informing Probability of Eradication of aquatic nonindigenous species; Drolet et al. 2014). The application uses a review of 143 case studies to fit a generalized linear model (with binomial response variable and logit link function) relating the outcome (eradication or non-eradication; depending on whether individuals were seen in post-treatment surveys) to characteristics of the target system and the ANIS to be eradicated (taxonomy of the ANIS [plant/algae, invertebrate, or vertebrate], habitat [marine intertidal, marine subtidal, river/stream, or lake/pond], area of infestation, log-transformed spatial extent of target population [m^2], and population status ["introduced" if not breeding, "established" if breeding, or "invasive" if causing problems), and the intervention (method [mechanical, chemical, biological, or combination], containment [whether or not measures were taken to prevent spread to or from the target area], and duration of program including post-treatment survey). Based on independent evaluations of the predictive power of the model, it seems to provide reliable probabilities of eradication, e.g., management actions resulting in eradication are ranked higher than management actions resulting in non-eradication 92% of the time. However, whether managers would benefit from using the model depends on its relative performance when compared to the currently used expert evaluations. MIPE provides a unique opportunity to compare the value of evidence-based tools with that of expert opinion in the field of control of ANIS, and environmental management in general.

Here we report on the relative value of MIPE and expert judgment in evaluating feasibility of eradication of nonindigenous populations in aquatic systems. We conducted an expert elicitation survey and compared the results with the information derived from MIPE. Specifically, we compared (1) the relative importance of the different independent variables considered, (2) the direction of effects for the levels of the factors, (3) the accuracy of probability of success for real case studies, and associated uncertainty, derived from the survey and the model, and (4) the number of data points needed by the statistical model to equal or surpass the accuracy of expert predictions. This study provides a general evaluation of the usefulness of evidence-based approaches. In addition, it provides insights on the type of information (e.g., independent vs. dependent and qualitative vs. quantitative variables) where expert knowledge is most valuable.

Materials and Methods

Survey development

We developed the expert survey using the online-based tool SurveyMonkey. The survey had an introductory page in which the objectives were explained to potential respondents and we collected information about the professional experience of respondents (see the Appendix for a sample survey). We asked about type of current position held by the respondent, the number of years of experience working with ANIS, number of years working in control of ANIS, and the field of expertise. Following the introduction page, the survey had two different sections.

In the first section, we collected information about the independent variables perceived by respondents as being important in influencing the probability of eradication. For each categorical factor considered (taxonomy, habitat, population status, method, and containment), experts were asked to rank the categories in increasing order of feasibility of eradication. For example, the question about taxonomy asked experts to rank plants/algae, invertebrates, and vertebrates, from easiest to hardest to eradicate. For containment, which only has two categories (yes or no) experts were asked if they perceived containment measures to increase, decrease, or as having no effect on the probability of eradication. For continuous variables (area and duration) experts were asked to provide the direction of the effect. For example, we asked whether a larger population is easier, harder, or equally hard to eradicate than a smaller population of the same species. Finally, we asked respondents to rank all seven factors for their perceived importance in influencing probability of eradication. Experts could express perceived ties by using the same rank more than once.

In the second section, we asked experts to provide perceived probability of eradication, both qualitatively and quantitatively, when details for real case studies were provided. For each case study in the data set ($n = 143$), we wrote a paragraph in which we described (1) broad taxonomy of the target species (the species names were not given to reduce the chances of the case studies being recognized by the experts), (2) the size of individuals at the adult stage, (3) the history of invasion of the species in other areas, (4) the type and size of the habitat in which the eradication program was conducted, (5) the suspected vector of introduction, (6) the area of infestation, (7) the means of dispersal of the target species, (8) the population status at the start of the program, (9) the estimated time between introduction (or first detection) and the start of program, (10) details about the methods used (including containment) to remove individuals, and (11) the duration of the program including post-treatment monitoring. When information was missing (e.g., time of introduction), this was stated in the description. Means of dispersal were not described unless the life history of the species could result in mobility above or below what might be considered typical of the taxonomic group. For example, means of dispersal for fish would be described for migratory species. Experts were asked to qualify probability of eradication for each case study as being very unlikely, unlikely, likely, or very likely. Experts were also asked to provide a quantitative probability of eradication between 0 and 1. A total of eight case studies were provided to each expert in a stratified-random manner, i.e., one question randomly selected from each of eight groups of case studies (17–18 studies per group). If experts recognized a case study and knew the outcome, they were asked to skip that question.

Elicitation of experts

By e-mail, we invited a list of potential respondents to take the survey. We directly contacted 46 experts including researchers from the Canadian Aquatic Invasive Species Network, other Canadian invasive species specialists, and relevant authors of the case studies we included in our data set, resulting in a list of authors from around the world. In addition, the coordinators of the Northeastern, Great Lakes, and Western Aquatic Nuisance Species Panels agreed to forward our e-mail invitation to their members, who represent several major networks of ANIS researchers and managers in North America. In all invitations, potential respondents were encouraged to distribute the survey to qualified colleagues.

Statistical analyses

Our statistical analyses addressed several questions. (1) What independent variables (and levels within factors) did the experts perceive as being important in influencing the probability of eradication when asked directly (section 1)? (2) What independent variables (and levels within factors) did the experts use in judging feasibility of eradication for real case studies (section 2)? (3) Were the important independent variables (and levels within factors), identified above, consistent among methods of elicitation and the same factors used by the empirical model MIPE? (4) How did the predictions based on expert opinion (section 2) compare to the real outcomes of case studies, and to MIPE's predictions for those case studies? And (5) how many case studies need to be included in the MIPE data set to obtain an accuracy equal or greater than that of experts?

The importance of the different independent variables in influencing feasibility of eradication obtained by directly questioning the experts (section 1) was investigated by sorting variables in ascending order of the sum of ranks assigned by experts (factors with small sum of ranks represent factors seen as being more important). We then used sign rank tests (Zar 2010) to determine the perceived influence of levels within factor. A sign was determined for each pairwise comparison by each expert (−, =, or +; e.g., for taxonomy, we compared the assigned ranks for plants/algae vs. invertebrates, plants/algae vs. vertebrates, and invertebrates vs. vertebrates) and we evaluated significance with a Bonferroni correction.

To determine which variables were influencing the judgment of experts when predicting eradication probability, we fitted a general linear main-effect model with seven independent variables (taxonomy, habitat, area, population status, method, containment, and duration). The dependent variable was the quantitative probability of eradication returned by experts for real case studies in section 2 ($n = 226$). The relative importance of each variable was evaluated with corrected Akaike information criterion (AIC_c) determined for every single-effect model. AIC_c values were sorted from smallest to largest; variables with lower AIC_c values were the most influential. To examine within-variable rankings, we used the same model and sorted the least square means of the level of each categorical variable; significance was evaluated with LSD post-hoc tests, with Bonferroni correction, when the main effect was significant. For continuous variables, the sign and significance of the regression coefficient was used.

To determine the relative importance of each independent variable empirically (i.e., MIPE), we used a generalized linear model with the seven independent variables and real outcome of case studies as the dependent variable ($n = 143$). We sorted AIC_c values determined for every single-effect model, as above. We ranked the level of each categorical factor with the estimated marginal means, followed with pair-wise contrasts with a Bonferroni correction to determine significance. For continuous variables, the sign and significance of the coefficient was used.

We qualitatively compared the importance of independent variables, and the levels within a variable, among the two means of expert elicitation (sections 1 and 2) and the empirical model MIPE. Given the wide range of data types and analyses used, no formal statistical analysis was conducted.

We analyzed the accuracy of probabilities of eradication returned by experts and compared their accuracy with that of MIPE. We first evaluated if expert judgments were consistent when elicited qualitatively and quantitatively using boxplots of quantitative probability of eradication for the different categories (very unlikely, unlikely, etc.). To determine the general accuracy of qualitative expert judgments, we built frequency distributions of expert-assigned categories, separately for case studies with real outcomes of eradication and non-eradication. A chi-square test was conducted to determine if these distributions differed. Given that MIPE returns quantitative probabilities of eradication, assigning cut-off values for the different categories would be an arbitrary process. In addition, individual experts may have used different cut-off values between "unlikely" and "very unlikely" and between "likely" and "very likely" because of differences in linguistic interpretation. Therefore we did not compare expert judgments and MIPE using all four categories but compared the correct classification rates using only two. Correct eradication classification rates were defined as the proportion of case studies categorized as likely or very likely (for experts) and with a predicted probability of eradication greater than 0.5 (for MIPE) that actually resulted in eradication. Conversely, correct non-eradication classification rates were defined as proportion of case studies categorized as unlikely or very unlikely (for experts) and with a predicted probability of eradication lower than 0.5 (for MIPE) that actually resulted in non-eradication. Correct classification rates of experts and MIPE were compared with Z tests (Zar 2010). The quantitative responses of experts were compared to the predictions of the model in two different ways. First, for the experts' and model's predictions, we calculated the deviation between the prediction and the outcome for each case study. For example, a predicted probability of 0.3 for a case study for which the real outcome was eradication would have a deviation of -0.7 (i.e., $0.3 - 1$). Values close to zero correspond to accurate predictions, negative values correspond to pessimistic predictions, positive values correspond to optimistic predictions, with maximum pessimism and optimism corresponding to -1 and 1, respectively (i.e., a predicted probability of 0 for a real outcome of eradication, and a predicted probability of 1 for a real outcome of non-eradication). Frequency distribution histograms of deviations were built for experts and MIPE, and were compared visually. Second, we compared the receiver operating characteristic (ROC) curves built from the responses of experts and the model. For each curve, we calculated the area under the curve (AUC) corresponding to the probability that a case study that actually resulted in eradication would obtain a higher probability of eradication than a case study that actually resulted in non-eradication (Hanley and McNeil 1982). We also determined the optimal probability threshold and evaluated the true and false positive rates at that threshold. The true positive rate corresponds to the proportion of case studies, with an estimated probability of success higher than the threshold, which actually resulted in eradication. Conversely, the false positive rate corresponds to the proportion of case studies, with an estimated probability of success lower than the threshold, which resulted in eradication. For all comparisons between experts' and model's accuracy, predictions from multiple experts on the same case study were considered as independent data; for these case studies, model's predictions were included the same number of times in the distributions (e.g., if we had three expert predictions for a case study, the prediction of MIPE was included three times when building the data set). Also, MIPE predictions were made using a model calibrated with a data set that excluded the particular case study for which the prediction was made.

We did not ask experts to provide a measure of certainty surrounding their predictions, but we used among-expert variability to address uncertainty. We used the case studies for which we obtained predictions from three experts or more in section 2 ($n = 31$). From these, we calculated a mean prediction and the range (highest prediction minus lowest prediction), corresponding to among-expert uncertainty. We compared the width of experts' prediction range with the 95% confidence limits (based on the precision of parameter estimates of the generalized linear model) surrounding the MIPE predictions graphically.

Finally, to investigate how many case studies need to be included in the MIPE data set to obtain accuracy equal or greater than that of experts, we conducted a bootstrapped cross-validation procedure. We randomly selected case studies out of the 143 possibilities to calibrate models, and used these models to obtain an independent probability of eradication for the remaining case studies. We then built ROC curves and kept track of the AUC. We varied the number of training case

TABLE 1. Summary of the relative importance of independent variables in influencing the outcome of eradication attempts of nonindigenous species in aquatic environments evaluated with different methods.

Variable	MIPE	Direct	Indirect
Taxonomy	7 (34.84)	7 (147)	7 (32.09)
Habitat	3 (21.30)	2 (98)	5 (30.34)
Area	1 (0.00)	1 (89)	1 (0.00)
Status	5 (32.09)	3 (112)	4 (26.42)
Method	2 (14.66)	5 (125)	2 (7.13)
Containment	6 (33.11)	4 (122)	3 (25.98)
Duration	4 (31.28)	6 (145)	6 (30.79)

Notes: Values are ranks from each method. MIPE (Model Informing Probability of Eradication) is the evidence-based tool. Direct is the ranking of experts when asked to provide a relative importance. Indirect presents the ranks obtained indirectly from fitting a statistical model to the quantitative probabilities of eradication returned by experts. Relative measures of importance are given in parentheses; for MIPE and Indirect this is ΔAIC_c (the change in Akaike's information criterion corrected for small sample sizes) and for Direct, it is the sum of the ranks.

studies from 20 to 100 by increments of 5. Each was repeated 1000 times and we calculated the mean AUC and associated 95% confidence intervals by determining the 2.5th and 97.5th percentiles. These were then compared graphically to the AUC for the quantitative predictions of experts. At low numbers of training case studies, we often obtained insufficient replication to include some of the categorical factors. When we had less than two case studies for any level of a factor, this factor was removed entirely from the statistical model.

RESULTS

A total of 38 experts completed the survey. Of these, four were academic researchers, 11 were environmental managers, 14 were government researchers, four were government biologists, and four worked in industry (e.g., consultants). Experts had between 1 and 35 years of experience working in the general field of ANIS (mean of 15.4) and between 0 and 30 years of experience working in control of ANIS (mean of 10.25).

The importance of factors influencing the probability of eradication identified by MIPE showed that area, method, and habitat were the most influential, with the other factors having small influence (Table 1). When asked directly, experts generally ranked the factors similar to MIPE, with the exception of method, which appeared in fifth position. When assessed indirectly, through fitting a statistical model to the quantitative response of experts, ranks were generally similar to those obtained by the other methods. An exception was in the evaluation of containment that was perceived as more important (Table 1).

Overall, the rankings of categories or levels within factors by experts (from the two methods of elicitation) tended to agree with those obtained by MIPE (Table 2). One notable exception is for habitat; when experts were asked directly, freshwater ecosystems were seen as more likely to result in eradication, whereas the other methods revealed that the marine intertidal environment is one of the habitats where eradication is the most likely (Table 2).

When provided with details of case studies (section 2), the qualitative and quantitative predictions of experts were generally consistent (Fig. 1) with the exception of a few obvious outliers. We decided to keep these outliers, as all analyses were also done after removing them and the results were similar. When asked to provide a qualitative probability of eradication, experts' frequency distributions of categories for case studies that actually resulted in eradication and non-eradication differed (Fig. 1; $\chi^2 = 26.09$, df = 3, $P < 0.001$). Among case studies classified as likely or very likely by experts, 62% actually resulted in eradication; this correct classification rate was less than that of

TABLE 2. Summary of ranks of levels within variables in order of feasibility of eradication or effectiveness of intervention.

Variable and source	Ranks
Taxonomy	
Direct	vertebrate[a] > plant[ab] > invertebrate[b]
Indirect	vertebrate[a] > invertebrate[ab] > plant[b]
MIPE	vertebrate[a] > plant[a] > invertebrate[a]
Habitat	
Direct	lake[a] > river[b] > intertidal[c] > subtidal[d]
Indirect	intertidal[a] > subtidal[a] > river[a] > lake[a]
MIPE	intertidal[ab] = lake[a] > subtidal[ab] > river[b]
Area	
Direct	small[a] > large[b]
Indirect	small > large *
MIPE	small > large *
Status	
Direct	introduced[a] > established[b] > invasive[b]
Indirect	invasive[a] > introduced[a] > established[a]
MIPE	introduced[a] > established[ab] > invasive[b]
Method	
Direct	combination[a] > chemical[b] > biological[c] > mechanical[c]
Indirect	chemical[a] > combination[a] > biological[ab] > mechanical[b]
MIPE	combination[a] > biological[a] > chemical[a] > mechanical[b]
Containment	
Direct	yes[a] > no[b]
Indirect	no[a] > yes[a]
MIPE	no[a] > yes[a]
Duration	
Direct	longer[a] > shorter[b]
Indirect	longer > shorter[ns]
MIPE	longer > shorter[ns]

Notes: Direct is the ranking of experts when asked to provide a relative importance. Indirect presents the ranks obtained indirectly from fitting a statistical model to the quantitative probabilities of eradication returned by experts. MIPE (Model Informing Probability of Eradication) is the evidence-based tool. Categories not sharing a common letter are significantly different after Bonferroni correction.

* Show significant continuous effects ($P < 0.05$); ns, nonsignificant continuous effects.

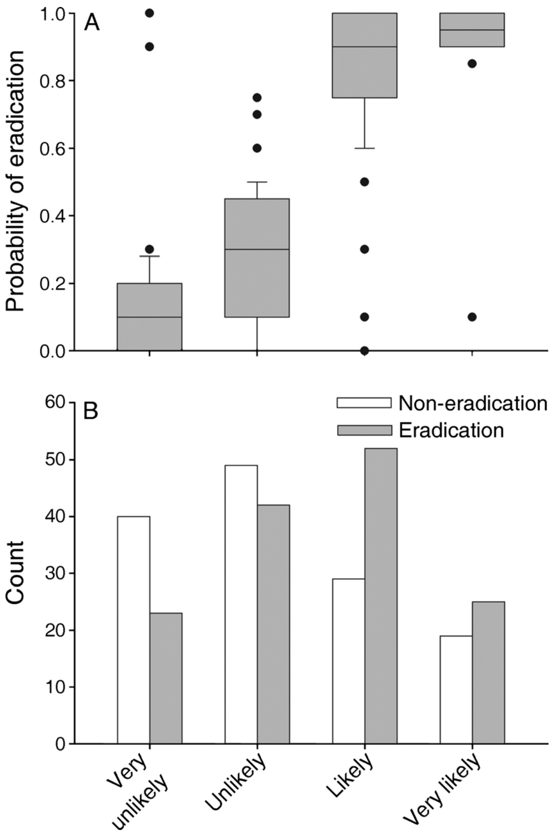

FIG. 1. (A) The relationship between quantitative and qualitative probability of eradication assigned by experts to real case studies. Lines show medians, boxes show 25th and 75th percentiles, whiskers show the limits of 90% confidence limits, and dots are outliers. (B) Frequency distribution of expert-assigned categories for case studies that actually resulted in eradication and non-eradication.

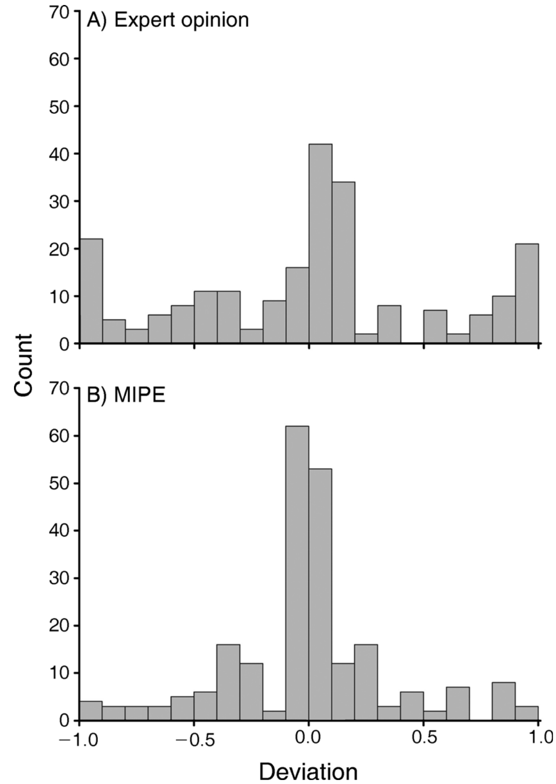

FIG. 2. Frequency distribution of the deviation between probability of eradication provided by (A) experts for real case studies and the actual outcome (0, non-eradication; 1, eradication) and (B) predictions of the Model Informing Probability of Eradication of aquatic nonindigenous species (MIPE).

MIPE (80%; $Z = 3.34$, $P = 0.001$). Among case studies classified as unlikely or very unlikely by experts, 58% actually resulted in non-eradication; this correct classification rate was less than that of MIPE (81%; $Z = 4.15$, $P < 0.001$).

The frequency distribution of deviations (prediction minus outcome) of experts asked to provide a quantitative probability of eradication was symmetrical and centered on zero (Fig. 2). Thus experts did not systematically under- or overestimate probability of eradication, and the average prediction was accurate. However, there were peaks at the two extremes of the histograms, meaning that experts often completely misjudged case studies (i.e., low probability for eradication or high probability for non-eradication). The frequency distribution of deviations for MIPE was also centered on zero, but there were no peaks at the two extremes. ROC curves further confirmed that MIPE provided more accurate estimations of probability of eradication than experts (Fig. 3). The ROC curve for MIPE lay well above that of the experts. The AUC was 0.90 for MIPE and 0.60 for experts. The optimal probability threshold for the MIPE ROC curve was 0.39, with true and false positive rates of 0.91 and 0.24, respectively. In other words, case studies with a calculated probability of success above 0.39 resulted in

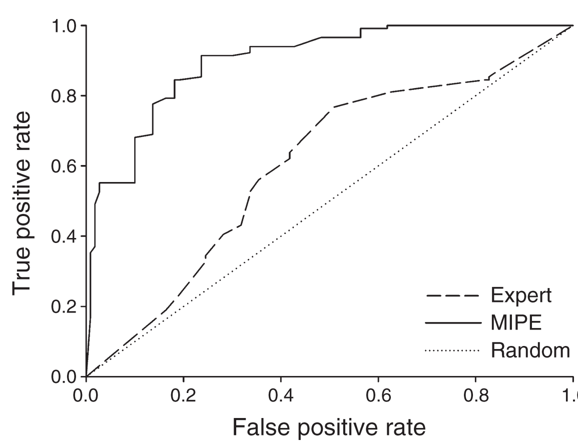

FIG. 3. Receiver operating characteristics curves for the predicted probability of eradication of MIPE (solid lines) and experts (dashed lines) when details about the case studies were provided. The dotted line represents expected curve if predictions were made at random.

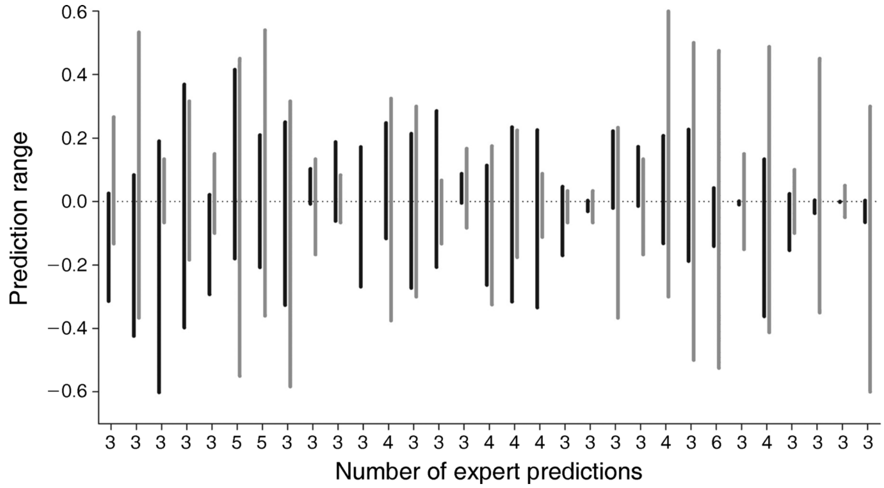

FIG. 4. Comparison of precision of predicted probability of eradication for MIPE (black lines) and experts (gray lines). The lines show the width of the 95% confidence intervals for MIPE and range of predictions made by multiple experts. Intervals and ranges were adjusted to bring the means to zero. The numbers on the x-axis represent the number of expert predictions returned for a particular case study.

eradication 91% of times, whereas only 24% of case studies with a prediction below 0.39 succeeded. In contrast, the ROC curve fitted to quantitative expert answers had an optimal threshold of 0.25, with a 0.77 and 0.51 true and false positive rate, respectively.

The qualitative comparison of precision of experts' and MIPE predictions showed no particular patterns (Fig. 4). Sometimes the confidence of the model was higher, and sometimes it was lower. However, there was a weak link between the width of MIPE's confidence intervals and the range of expert predictions for the same case study (correlation coefficient = 0.20, df = 29, P = 0.28; data not presented).

Increasing the number of training case studies gradually increased the AUC of MIPE and decreased its confidence intervals (Fig. 5). After ~30 case studies used to calibrate MIPE, the confidence limits no longer included the AUC of experts' predictions (Fig. 5).

DISCUSSION

Expert knowledge is an important, and often the only, source of information used in the decision-making process of environmental managers (Pullin et al. 2004, Fazey et al. 2006, Cook et al. 2010). The empirical data required to inform decisions often do not exist or are not readily accessible. Making the evidence available and the development of evidence-based tools is expected to improve the overall efficiency of environmental management actions (Pullin and Knight 2001, Sutherland et al. 2004). However, improved efficiency will only happen if evidence-based tools are more accurate than expert opinion. This assumption has seldom been tested thoroughly because expert information is normally not elicited when empirical data already exist. Here we gathered expert opinion after the development of a model designed to evaluate the probability of success of a conservation intervention. This was done to directly evaluate the relative accuracy of both, and to evaluate the type of information where expert opinion is most valuable. Recent work formalized expert elicitation, in particular in the context of establishing priors in a Bayesian modeling framework (Choy et al. 2009, Kuhnert et al. 2010), or obtaining among-expert consensus (MacMillan and Marshall 2006). We did not intend to develop such elaborate models, but rather focused on direct expert elicitation. We believe this is more representative of what a manager would seek from one (or a few)

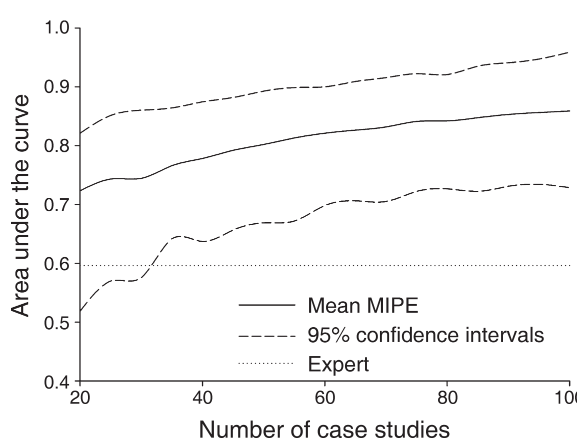

FIG. 5. Results of a bootstrapped cross-validation procedure to determine the number of case studies required for MIPE to be equally accurate to expert judgment. The graph shows the area under receiver operating characteristic (ROC) curves fitted to MIPE's predictions when varying the number of training and testing case studies by increments of 5. Confidence limits are 2.5th and 97.5th percentiles based on 1000 iterations and dotted line shows the area under the curve fitted to quantitative expert predictions.

expert(s) consulted rapidly prior to taking action in a crisis situation.

Overall, experts and MIPE agreed on the importance of the independent variables, and the among-level direction of effects within a variable; experts had a good understanding of factors influencing probability of eradication. For this particular problem and potentially others, reliable qualitative information on the influence of independent variables can be obtained from expert consultations. This confirms the value of asking experts how they think a group of independent variables influences a metric, and the relative importance of the variables. This approach was used in the fields of species distribution mapping (e.g., Pearce et al. 2001) and assessment of threats to endangered species (O'Neill et al. 2008, Donlan et al. 2010). Also, the information on independent variables was similar whether it was derived by asking experts directly or obtained indirectly through fitting a model to the quantitative responses of experts. In other words, experts were able to coherently integrate their qualitative understanding of the influence of independent variables when making a quantitative prediction. This suggests one can identify predictors by asking experts to provide an expected metric under different combinations of independent variables. This approach was used previously in the field of species distribution modeling; experts were asked to provide an estimated density or probability of occurrence when provided with a description of local conditions; the information was then used to identify habitat features most influencing distribution (Murray et al. 2009, O'Leary et al. 2009). Generally, our results support the use of expert opinion to derive information on predictor variables.

Experts generally provided estimated chances of eradication that went in the right direction (when compared to the real outcome of case studies), whether this probability was elicited qualitatively or quantitatively. Decisions concerning the planning of an eradication attempt, and possibly other types of interventions, would therefore be well informed by expert consultation. Interestingly, there was no systematic bias in experts' predictions; there was no trend toward either overoptimism or overpessimism as was observed in other studies evaluating the value of expert opinion (McCarthy et al. 2004, Murray et al. 2009). However, there were tremendous differences in experts' accuracy when compared to that of MIPE; any way we looked at it, MIPE was much more accurate. We would therefore advocate for its use as a complement to expert judgment in the planning of management interventions toward nonindigenous aquatic species. Expert judgment will always be necessary even when predictive statistical models are available. This is particularly important for a situation that is not well represented in the data set. For example, out of the 65 case studies targeting vertebrates in the MIPE data set, only one was targeting an amphibian (the rest were fish). Therefore managers should evaluate how well the model applies to their situation (in the case of amphibians, it does not); if applicability is questionable, expert judgment should be sought and relied on to put the model's predictions in perspective.

The availability of case studies in the literature will determine to what degree statistical models will be helpful in complementing expert opinion. For many potential fields of application, there may not be enough published information to calibrate reliable models. It is therefore important to investigate how many case studies are required for models to be of value. In our case, higher accuracy (when compared to expert judgment) was achieved when more than 30 case studies were used to calibrate the model. This should by no means be seen as a rule of thumb; the very low number of case studies required for this result is probably related to the overwhelming influence of area of infestation in predicting outcome. Similar work should be undertaken with other fields of application to relate minimum number of studies required for models to surpass expert judgment with metrics of model accuracy. This would allow determining a priori if enough information is available for an evidence-based approach to be useful.

In retrospect, it would have been desirable to ask experts to provide a measure of uncertainty surrounding their quantitative predictions as recommended by Martin et al. (2012). Had we asked for a range of probabilities the experts felt 95% confident contains their true estimate, we could have compared their precision and that of MIPE directly. Failing to do so, we had to rely on the among-expert variability to address uncertainty. Overall, experts tended to disagree more for case studies with wider MIPE confidence intervals. This suggests that uncertainty of expert judgments and the statistical model were in accord with each other. However, we could not address important questions such as expert overconfidence (Martin et al. 2012).

In conclusion, we provided the first evaluation of the relative value of quantitative evidence-based tools and expert opinion in predicting probability of success of environmental management actions. Although experts understood the importance of various factors and were able to integrate this information to provide generally reliable chances of success, the accuracy of their quantitative predictions was markedly lower than that of the evidence-based tool. MIPE can therefore be recommended for use in the planning of eradication attempts of ANIS as a complement to expert judgments. This study highlights the value of evidence-based management; quantitative decision support tools could help experts and managers in their decision-making process.

Acknowledgments

Funding for this project was provided by the Canadian Aquatic Invasive Species Network (CAISN II) of the Natural Science and Engineering Research Council (NSERC) of

Canada. M. A. Lewis gratefully acknowledges a Canada Research Chair, NSERC Discovery and Accelerator grants and a Killam Research Fellowship. We are grateful to all experts who participated in the survey and to the coordinators of the Northeastern, Great Lakes, and Western Aquatic Nuisance Species Panels for help in disseminating the survey. We thank three anonymous reviewers for comments and suggestions that greatly improved the manuscript.

Literature Cited

Brooks, J. S., M. A. Franzen, C. M. Holmes, M. N. Grote, and M. B. Mulder. 2006. Testing hypotheses for the success of different conservation strategies. Conservation Biology 20:1528–1538.

Choy, S. L., R. O'Leary, and K. Mengersen. 2009. Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. Ecology 90:265–277.

Clevenger, A. P., J. Wierzchowski, B. Chruszcz, and K. Gunson. 2002. GIS-generated, expert-based models for identifying wildlife habitat linkages and planning mitigation passages. Conservation Biology 16:503–514.

Cook, C. N., M. Hockings, and R. W. Carter. 2010. Conservation in the dark? The information used to support management decisions. Frontiers in Ecology and the Environment 8:181–186.

Crombie, J., E. Knight, and S. Barry. 2008. Marine pest incursions—a tool to predict the cost of eradication based on expert assessments. Report of the Australian Government, Bureau of Rural Sciences, Canberra, Australia.

Donlan, C. J., D. K. Wingfield, L. B. Crowder, and C. Wilcox. 2010. Using expert opinion surveys to rank threats to endangered species: a case study with sea turtles. Conservation Biology 24:1586–1596.

Drolet, D., A. Locke, M. A. Lewis, and J. Davidson. 2014. User-friendly and evidence-based tool to evaluate probability of eradication of aquatic non-indigenous species. Journal of Applied Ecology. http://dx.doi.org/10.1111/1365-2664.12263

Fazey, I., J. A. Fazey, J. G. Salisbury, D. B. Lindenmayer, and S. Dovers. 2006. The nature and role of experiential knowledge for environmental conservation. Environmental Conservation 33:1–10.

Fischer, J., and D. B. Lindenmayer. 2000. An assessment of the published results of animal relocations. Biological Conservation 96:1–11.

Geneletti, D. 2005. Formalising expert opinion through multi-attribute value functions: an application in landscape ecology. Journal of Environmental Management 76:255–262.

Hanley, J. A., and B. J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36.

Holzkämper, A., V. Kumar, B. W. J. Surridge, A. Paetzold, and D. N. Lerner. 2012. Bringing diverse sources together—a meta-model for supporting integrated catchment management. Journal of Environmental Management 96:116–127.

Johnson, C. J., and M. P. Gillingham. 2004. Mapping uncertainty: sensitivity of wildlife habitat ratings to expert opinion. Journal of Applied Ecology 41:1032–1041.

Kuhnert, P. M., T. G. Martin, and S. P. Griffiths. 2010. A guide to eliciting and using expert knowledge in Bayesian ecological models. Ecology Letters 13:900–914.

Leung, B., et al. 2012. TEASIng apart alien species risk assessments: a framework for best practices. Ecology Letters 15:1475–1493.

Mack, R. N., D. Simberloff, W. M. Lonsdale, H. Evans, M. Clout, and F. A. Bazzaz. 2000. Biotic invasions: causes, epidemiology, global consequences and control. Ecological Applications 10:689–710.

MacMillan, D. C., and K. Marshall. 2006. The Delphi process—an expert-based approach to ecological modelling in data-poor environments. Animal Conservation 9:11–19.

Martin, T. G., M. A. Burgman, F. Fidler, P. M. Kuhnert, S. Low-Choy, M. McBride, and K. Mengersen. 2012. Eliciting expert knowledge in conservation science. Conservation Biology 26:29–38.

Martin, T. G., P. M. Kuhnert, K. Mengersen, and H. P. Possingham. 2005. The power of expert opinion in ecological models using Bayesian methods: impact of grazing on birds. Ecological Applications 15:266–280.

Martins, T. L. F., M. d. L. Brooke, G. M. Hilton, S. Farnsworth, J. Gould, and D. J. Pain. 2006. Costing eradications of alien mammals from islands. Animal Conservation 9:439–444.

McCarthy, M. A., et al. 2004. Comparing predictions of extinction risk using models and subjective judgement. Acta Oecologica 26:67–74.

McEnnulty, F. R., N. J. Bax, B. Schaffelke, and M. L. Campbell. 2001. A review of rapid response options for the control of ABWMAC listed introduced marine pest species and related taxa in Australian waters. Centre for Research on Introduced Marine Pests. Technical Report no. 23. CSIRO Marine Research, Hobart, Tasmania, Australia.

Miller, A. W., G. M. Ruiz, M. S. Minton, and R. F. Ambrose. 2007. Differentiating successful and failed molluscan invaders in estuarine ecosystems. Marine Ecology Progress Series 332:41–51.

Murray, J. V., A. W. Goldizen, R. A. O'Leary, C. A. McAlpine, H. P. Possingham, and S. L. Choy. 2009. How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock-wallabies *Petrogale penicillata*. Journal of Applied Ecology 46:842–851.

Newton, A. C., G. B. Stewart, A. Diaz, D. Golicher, and A. S. Pullin. 2007. Bayesian belief networks as a tool for evidence-based conservation management. Journal for Nature Conservation 15:144–160.

O'Leary, R. A., S. L. Chow, J. V. Murray, M. Kynn, R. Denham, T. G. Martin, and K. Mengersen. 2009. Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby *Petrogale penicillata*. Environmetrics 20:379–398.

O'Neill, S. J., T. J. Osborn, M. Hulme, I. Lorenzoni, and A. R. Watkinson. 2008. Using expert knowledge to assess uncertainties in future polar bear populations under climate change. Journal Applied Ecology 45:1649–1659.

Padgee, A., Y. Kim, and P. J. Daugherty. 2006. What makes community forest management successful: a meta-study from community forests throughout the world. Society and Natural Resources: An International Journal 19:33–52.

Pearce, J. L., K. Cherry, M. Drielsma, S. Ferrier, and G. Whish. 2001. Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. Journal of Applied Ecology 38:412–424.

Pluess, T., R. Cannon, V. Jarošík, J. Pergl, P. Pyšek, and S. Bacher. 2012*a*. When are eradication campaigns successful? A test of common assumptions. Biological Invasions 14:1365–1378.

Pluess, T., V. Jarošík, R. Cannon, J. Pergl, A. Breukers, and S. Bacher. 2012*b*. Which factors affect the success or failure of campaigns against alien species. PLoS ONE 7:e48157.

Pullin, A. S., and T. M. Knight. 2001. Effectiveness in conservation practice: pointers from medicine and public health. Conservation Biology 15:50–54.

Pullin, A. S., and T. M. Knight. 2005. Assessing conservation management's evidence base: a survey of management-plan compilers in the United Kingdom and Australia. Conservation Biology 19:1989–1996.

Pullin, A. S., T. M. Knight, D. A. Stone, and K. Charman. 2004. Do conservation managers use scientific evidence to support their decision-making? Biological Conservation 119:245–252.

Raymond, B., J. McInnes, J. M. Dambacher, S. Way, and D. M. Bergstrom. 2011. Qualitative modelling of invasive species eradication on subantarctic Maquarie Island. Journal of Applied Ecology 48:181–191.

Roberts, P. D., and A. S. Pullin. 2008. The effectiveness of management interventions for the control of *Spartina* species: a systematic review and meta-analysis. Aquatic Conservation: Marine and Freshwater Ecosystems 18:592–618.

Smith, R. K., A. S. Pullin, G. B. Stewart, and W. J. Sutherland. 2010. Effectiveness of predator removal for enhancing bird populations. Conservation Biology 24:820–829.

Smith, R. K., A. S. Pullin, G. B. Stewart, and W. J. Sutherland. 2011. Is nest predator exclusion an effective strategy for enhancing bird populations? Biological Conservation 144:1–10.

Stewart, G. B., H. R. Bayliss, D. A. Showler, W. J. Sutherland, and A. S. Pullin. 2009. Effectiveness of engineered in-stream structure mitigation measures to increase salmonid abundance: a systematic review. Ecological Applications 19:931–941.

Stewart, G. B., C. F. Coles, and A. S. Pullin. 2005. Applying evidence-based practice in conservation management: lessons from the first systematic review and dissemination projects. Biological Conservation 126:270–278.

Stewart, G. B., and A. S. Pullin. 2008. The relative importance of grazing stock type and grazing intensity for conservation of mesotrophic "old meadow" pasture. Journal for Nature Conservation 16:175–185.

Sutherland, W. J., A. S. Pullin, P. M. Dolman, and T. M. Knight. 2004. The need for evidence-based conservation. Trends in Ecology and Evolution 19:305–308.

Tobin, P. C., J. M. Kean, D. M. Suckling, D. G. McCullough, D. A. Herms, and L. D. Stringer. 2014. Determinants of successful arthropod eradication programs. Biological Invasions 16:401–414.

Tyler, C., A. S. Pullin, and G. B. Stewart. 2006. Effectiveness of management interventions to control invasion by *Rhododendron ponticum*. Environmental Management 37:513–522.

Whitfield, D. P., M. Ruddock, and R. Bullman. 2008. Expert opinion as a tool for quantifying bird tolerance to human disturbance. Biological Conservation 141:2708–2717.

Yamada, K., J. Elith, M. McCarthy, and A. Zerger. 2003. Eliciting and integrating expert knowledge for wildlife habitat modelling. Ecological Modelling 165:251–264.

Zar, J. H. 2010. Biostatistical analysis. Fifth edition. Prentice Hall, Upper Saddle River, New Jersey, USA.

Supplemental Material

Ecological Archives

The Appendix is available online: http://dx.doi.org/10.1890/14-0180.1.sm

Data Availability

Data associated with this article have been archived at Dryad: http://dx.doi.org/10.5061/dryad.vv2c1