Proceedings of the 9th International Conference on Marine Bioinvasions (19–21 January 2016, Sydney, Australia)

Research Article

Optimizing screening protocols for non-indigenous species: are currently used tools over-parameterized?

David Drolet[1],*, Claudio DiBacco[2], Andrea Locke[3], Cynthia H. McKenzie[4], Christopher W. McKindsey[1] and Thomas W. Therriault[5]

[1]*Fisheries and Oceans Canada, Institut Maurice-Lamontagne, Mont-Joli, Québec, G5H 3Z4 Canada*

[2]*Fisheries and Oceans Canada, Bedford Institute of Oceanography, Dartmouth, Nova Scotia, B2Y 4A2 Canada*

[3]*Fisheries and Oceans Canada, Gulf Fisheries Centre, Moncton, New Brunswick, E1C 9B6 Canada*

[4]*Fisheries and Oceans Canada, Northwest Atlantic Fisheries Centre, St. John 's, Newfoundland and Labrador, A1C 5X1 Canada*

[5]*Fisheries and Oceans Canada, Pacific Biological Station, Nanaimo, British Columbia, V9T 6N7 Canada*

*Corresponding author

E-mail: David.Drolet@dfo-mpo.gc.ca

Editor's note:

This study was first presented at the 9th International Conference on Marine Bioinvasions held in Sydney, Australia, January 19–21, 2016 (http://www.marinebioinvasions.info/previous-conferences). Since their inception in 1999, ICMB series have provided a venue for the exchange of information on various aspects of biological invasions in marine ecosystems, including ecological research, education, management and policies tackling marine bioinvasions.

Abstract

While screening-level risk assessment (SLRA) tools for non-indigenous species generally provide managers with reliable information for decision making (e.g., a proposed species introduction should be allowed or rejected), the results are affected by several sources of uncertainty. In particular, model uncertainty, related to the influence of factors/questions included in SLRA tools has rarely been addressed. Here we undertook an investigation of model uncertainty using a detailed evaluation of the contribution of questions included in the Canadian Marine Invasive Screening Tool (CMIST) and determined if the tool can be made more accurate through a series of optimization procedures. Accuracy was defined as the fit between assessment scores and the results of an expert opinion survey of risk posed by 48 marine invertebrate species known to have been introduced into Canadian coastal waters. We first measured the contribution of each question to accuracy and removed the ones that did not improve the fit. We then derived optimal weights that adjust the contribution of each question to maximize accuracy. Eight of 17 questions were found not to improve accuracy, or even decreased it; removing these questions, followed by addition of weights made the tool gradually more accurate when all species were included. However, an independent cross-validation test showed these weights to be too variable to consistently improve fit; this was probably related to the relatively small number of species included in the tests. Tools that have previously been tested using a large number of species should be used to determine if addition of optimal weights can improve independent predictions. The evidence that risk assessment tools are over-parameterized is building and we suggest that currently used tools would benefit from a detailed evaluation of the value of questions they include. Careful selection of questions and weights, based on accuracy improvement and other elements (e.g., organizational mandate) could greatly benefit SLRA tools, by providing more accurate estimations of risk and accelerating assessments.

Key words: risk assessment, invasions, marine invertebrates, Canadian Marine Invasive Screening Tool, Weed Risk Assessment tool

Introduction

Effective management of non-indigenous species requires an evaluation of the risk posed by a species to the ecological integrity of a target area. The most commonly used tools are semi-quantitative scoring systems that operate at the screening-level (reviewed in Kumschick and Richardson 2013) and generally provide reliable risk assessments with minimal data

and time investment (e.g., Gordon et al. 2008). Most of these tools have been adapted from the Australian Weed Risk Assessment tool (WRA; Pheloung et al. 1999) and ask several questions about factors or traits thought to be related to the invasive potential of a species. Answers to questions are typically converted to an ordinal scale and combined mathematically to provide a final risk score. Many such tools have been developed and tested for different taxa and geographical areas (e.g., Daehler et al. 2004; Copp et al. 2009; Tricarico et al. 2010; Gordon and Gantz 2011). While they provide good first-pass evaluations for many species, tools could be improved as some species are invariably classified incorrectly and others require further evaluation before a recommendation can be made (Kumschick and Richardson 2013).

The results returned by a risk assessment tool are influenced by several sources of uncertainty. The quality of information used to answer each question will vary among species. This information, and the wording of questions, can be interpreted differently by different assessors (judgment subjectivity and linguistic uncertainty, *sensu* Regan et al. [2002]). The same sources of uncertainty apply to the data used to calibrate and test tools (typically expert opinion about the realized impact or risk of species already present in an area). These processes lead to scores of varying levels of uncertainty among species, areas, assessors, and experts. While progress is being made in addressing these sources of uncertainty (judgment subjectivity and linguistic uncertainty, see Copp et al. 2009; Holt et al. 2012; Drolet et al. 2015), model uncertainty (*sensu* Regan et al. [2002]), related to which variables are needed to best represent a biological phenomenon, has rarely been addressed. For example, the WRA and its derivatives ask 49 questions, most of which deal with biological attributes/characteristics thought to influence invasiveness. However, empirical evidence linking many of these factors to invasiveness is weak (e.g., Bomford and Glover 2004). This has led many to suggest that risk assessment tools for invasive species are over-parameterized; tools may be equally or even more accurate if they are made more parsimonious through a careful selection of questions (Caley and Kuhnert 2006; Gordon et al. 2008; Koop et al. 2012; Bradie et al. 2015). In addition, if questions have variable discriminatory power, one would expect scores to be more accurate if their contribution to the final risk score was weighted according their discriminatory power. However, in most current tools, questions are either weighted equally or based on their perceived importance rather than on empirically-derived influence (but see Koop et al. 2012).

Recently, Drolet et al. (2015) developed and tested the Canadian Marine Invasive Screening Tool (CMIST), a tool designed to evaluate risk of accidental introductions of non-indigenous marine invertebrate species. The CMIST scores provided good approximation of expert opinion on biological risk associated with species known to have been introduced to Canadian coastal waters. Here we attempt to optimize CMIST by undertaking a detailed evaluation of the relative importance of each assessment question and adjusting their contribution to the final risk score. We first evaluate the influence of each question on the fit between expert opinion and CMIST scores, and the consequences of removing the questions that do not contribute to the fit. We then derive optimal weights for each question, based on their empirical relationship to expert opinion scores, and evaluate how this weighting influences score accuracy and precision.

Methods

Data acquisition

The CMIST was developed with potential over-parameterization in mind (i.e., keeping it as parsimonious as possible), and thus includes only 17 questions directly related to the invasion process: eight pertaining to likelihood of invasion and nine to potential impacts (Table 1; see Supplementary material Table S1 for full question formulation). A semi-quantitative score is assigned to each question based on assessors' answers (Low = 1, Moderate = 2, or High = 3) and level of uncertainty (see Drolet et al. 2015). Scores are averaged for the "likelihood" and "impact" questions and these two scores are multiplied to obtain a final score, potentially ranging from 1 to 9. The CMIST was tested by comparing the average score returned by two assessors and expert opinion scores on risk posed by species known to have been introduced to three Canadian marine ecoregions (DFO 2009; one on the west coast and two on the east coast). The expert opinion scores were based on level of risk and uncertainty (both being scored as Low, Moderate, or High; see Drolet et al. 2015) and potentially ranged from one to three. The assessment scores were well correlated to expert opinion for the two east coast ecoregions ($R^2 = 0.48$ and 0.74), but the fit was poorer on the west coast ($R^2 = 0.23$), probably due to lower information availability (to assessors and experts) resulting in wider confidence limits in this ecoregion (Drolet et al. 2015).

In this study, we analyzed species from all ecoregions together. Initially, we planned to use species-ecoregion combinations as independent data, but many of the species included were introduced to

Table 1. Contribution of CMIST questions to the fit between assessment scores and expert opinion for all species and geographical and taxonomic subsets. Values are differences in percentage of variation explained when a question is included and when it is not. Positive values (shaded) represent questions that contributed positively to accuracy; negative values represent questions that decreased accuracy. Full question formulation can be found in Supplementary material Table S1.

Question	Both coasts	East coast	West coast	Mollusca	Crustacea	Tunicata
Likelihood of invasion						
1) Established	2.54	-2.28	4.37	4.05	1.92	-0.01
2) Arrival	-1.81	0.61	-0.86	-2.80	-0.16	-2.39
3) Habitat	-1.57	-2.07	-1.03	2.33	-2.62	-0.59
4) Climate	-0.91	0.91	-2.06	-3.33	-1.09	0.36
5) Reproduction	2.82	1.20	2.37	1.86	2.14	0.14
6) Control agents	0.56	4.46	1.41	3.47	-0.61	-0.40
7) Natural dispersal	-0.80	0.51	-4.63	-5.05	-0.92	0.39
8) Anthropogenic dispersal	0.81	-0.76	2.05	2.20	0.23	0.26
Impact of invasion						
9) Impact on populations	0.52	-0.69	0.41	0.51	0.50	-0.06
10) Impact on communities	1.22	0.43	1.00	3.26	0.39	-0.15
11) Impact on habitat	1.45	-0.94	0.60	0.56	-0.31	0.39
12) Impact on ecosystems	1.93	-2.22	1.87	0.84	1.80	0.00
13) Diseases and parasites	-0.57	2.24	-1.06	-2.04	0.95	-0.63
14) Genetic effects	-4.00	-0.74	-3.06	-4.85	-2.34	-0.37
15) Impact on at-risk species	-0.30	1.94	-0.03	-0.33	0.03	0.16
16) Impact on aquaculture	-0.26	2.20	-0.19	-0.51	-2.60	0.19
17) Invasive elsewhere	2.48	1.59	2.84	6.79	0.67	-0.41

more than one ecoregion. The scores for species from the 2 coasts were different enough to be considered as unique replicates; however, they were very similar between the two east coast ecoregions and were therefore pooled.

Evaluation of question importance

To determine how each question contributed to the accuracy of CMIST (i.e., the ability of the tool to correctly estimate the expert evaluation of risk for a given species-coast combination), we first calculated the proportion of variation in expert opinion scores explained by CMIST scores (R^2 of a linear regression) when all questions were included. We then removed one question at a time in score calculation and again calculated the R^2. The contribution of a question was defined as the difference in percentage of variation explained by the regression when it is included and when it is not. Therefore, if a question contributes positively to the match between CMIST and expert opinion scores, the value is positive, i.e., the R^2 is larger (greater accuracy) when the question is included than when it is not. Conversely, a negative value means that the match between CMIST and expert opinion scores is worse when the question is included. This analysis was first done using species from both coasts (n = 48) in a single analysis. A similar analysis was done on subsets of species to evaluate the consistency of the questions' contribution between coasts: the east coast only (n = 18), and the west coast only (n = 30) and among the major taxonomic groups

included in the study: molluscs (n = 20), crustaceans (n = 10), and tunicates (n = 13).

Optimization of CMIST

We attempted to optimize CMIST in three different ways. First, we simply removed all the questions that did not contribute to accuracy when all species were included (identified above) and recalculated the CMIST scores. Second, we assigned weights to the questions that were retained. In this system the score for a species was calculated as:

$$\frac{\sum_{a=1}^{x} S_a w_a}{x} \times \frac{\sum_{b=1}^{y} S_b w_b}{y}$$

where a refers to a question related to likelihood of invasion, x is the number of questions related to likelihood of invasion that were retained, b refers to a question related to impact of invasion, y is the number of questions related to impact of invasion that were retained, S_a and S_b are the scores assigned to likelihood and impact of invasion questions, respectively, and w_a and w_b are the weights for likelihood and impact questions, respectively. We found the combination of weights (w_a and w_b) that provided the greatest match between CMIST and expert opinion scores, i.e., the combination of weights that maximized the R^2 between the two set of scores. The optimal weights were determined using the solver add-on in Microsoft Excel, with starting values of 1 for the weight of each question. In this analysis, the weights were constrained to positive values since the

questions whose answers were potentially negatively correlated with expert opinion scores were already removed. The third optimization technique used the same equation and methodologies to calculate the scores, but this time all questions were included and the optimal weights were allowed to include negative values. This permits the contribution by the questions for which answers are negatively correlated with expert opinion scores (i.e., some of the ones removed in the previous scenario).

Evaluation of optimization procedures

To determine if the optimization procedures improved accuracy, we first compared the relationships between CMIST and expert opinion scores using linear regression analysis. This was done for 1) all questions included, 2) questions removed, 3) questions removed with weights (positive), and 4) all questions with weights (positive or negative). To evaluate which set of scores best predicted the expert opinion data, we calculated the corrected Akaike Information Criterion (AICc) for each linear regression and used these values to determine the likelihood that each technique is the most accurate among the set of models tested.

The three optimization procedures (2–4 as described above) were used to calculate independent scores for each species using leave-one-out cross validation; the optimization was done using all but one species and these independent optimized models were used to calculate scores for the excluded species. We kept track of the proportion of times each question was retained ("question removed") and the weight for each question ("questions removed with weights" and "all questions with weights"). We then used AICc values for these independent models (and the original CMIST scores) to determine which provided the best fit to the expert opinion scores.

The CMIST allows for calculations of confidence limits surrounding assessment scores as explained in Drolet et al. (2015). In brief, it uses probabilities that an assessor might have provided a different answer to a question (under all possible combinations of score and uncertainty) to generate a range of potential scores for a species and derive associated 95% confidence limits. This procedure was used to evaluate how the optimization procedures influence score precision. We used a two-way main effect ANOVA with the fixed factor Method (4 levels: all questions, questions removed, questions removed with weights, and all questions with weights) and the random blocking factor Species (48 levels). Since the different optimization methods return scores on different scales, the dependent variable was the width of confidence limits for

a species divided by the mean score for a particular method. Thus, the Method effect tests for differences in proportional uncertainty: uncertainty being a proportion of the average scores for a method. Significant fixed effects were further evaluated with Tukey post-hoc tests (Day and Quinn 1989).

Results

Evaluation of question importance

Approximately half the questions (9 of 17) contributed positively to the fit between CMIST and expert opinion scores when all species were included (i.e., positive values for contribution to accuracy in Table 1). However, the accuracy of CMIST was greater if the remaining eight questions were ignored (i.e., negative values for contribution to accuracy in Table 1). With few exceptions, the questions related to likelihood of invasion contributed consistently to accuracy across geographical areas and taxa considered (Table 1). For example, the questions pertaining to establishment, reproduction, control agents, and anthropogenic dispersal contributed positively to accuracy in at least four of the six analyses. In contrast, questions about arrival, habitat, climate, and natural dispersal only contributed positively in one or two of the six analyses. The questions pertaining to impacts of invasion influenced accuracy in a much less consistent manner, with particularly important differences between coasts. However, the question pertaining to effects on community and ecosystems, and history of invasion elsewhere improved accuracy for five of six analyses, whereas the question about genetic effects was excluded for all analyses (Table 1).

Optimization of CMIST

After removing the questions that decreased accuracy (when all species were considered), the optimal combination of weights ignored three further questions (weight of 0; Table 2): Control agents, Impacts on populations, and Impacts on communities. Three questions appeared to be particularly important (high weights; Table 2): requirements for reproduction, impact on ecosystems, and history of invasion elsewhere. When all questions were considered and the optimal weights derived, the weights for likelihood of invasion questions tended to correspond to the contribution to accuracy (i.e., similar signs between values in column 1 Table 1 and column 3 in Table 2), but important differences were observed for the impact of invasion questions. Again, the questions pertaining to requirements for reproduction and history of invasion elsewhere were of particular importance (high

Table 2. Results of optimization procedures for CMIST. The "questions retained" column shows the proportion of times a question was retained during the cross-validation procedure. The other two columns shows the optimal weights derived when all species were included and the range of values, in parenthesis, obtained in the cross-validation. Full question formulation can be found in Table S1.

Question	Questions retained	Weight for questions retained	Weight for all questions
Likelihood of invasion			
1) Established	1	0.97 (0.43, 1.64)	3.22 (1.92, 7.88)
2) Arrival	0	na	-2.45 (-4.38, 2.12)
3) Habitat	0	na	-1.65 (-7.05, -0.04)
4) Climate	0	na	-3.68 (-5.63, -1.26)
5) Reproduction	1	1.86 (0.95, 2.33)	4.66 (3.19, 7.86)
6) Control agents	1	0.00 (0.00, 0.00)	0.21 (-6.98, 2.97)
7) Natural dispersal	0	na	2.36 (1.72, 3.37)
8) Anthropogenic dispersal	1	1.08 (0.32, 1.86)	3.85 (1.12, 8.98)
Impact of invasion			
9) Impact on populations	1	0.00 (0.00, 0.08)	0.79 (-2.50, 5.66)
10) Impact on communities	1	0.00 (0.00, 0.46)	-4.33 (-7.89, 1.19)
11) Impact on habitat	1	0.45 (0.00, 1.35)	4.01 (0.45, 6.00)
12) Impact on ecosystems	1	1.93 (1.10, 2.56)	1.47 (-9.56, 6.94)
13) Diseases and parasites	0	na	2.30 (0.39, 5.45)
14) Genetic effects	0	na	-4.08 (-10.94, -1.61)
15) Impact on at-risk species	0.02	na	6.02 (4.38, 8.18)
16) Impact on aquaculture	0.10	na	0.10 (-0.85, 1.88)
17) Invasive elsewhere	1	2.00 (1.71, 2.48)	5.32 (3.87, 12.15)

weights). However, questions about effects on habitat and species at risk were important in this analysis but not in the other ones.

Evaluation of optimization procedures

There was a gradual increase in accuracy when going from 1) all questions, to 2) questions removed, to 3) questions removed with weights, to 4) all questions with weights (Figure 1) when all species were included (i.e., non-independent test). This is evidenced by a more than two-fold increase in R^2 values between "all questions" and "all questions with weights" (Figure 1). The model with all questions and weights was by far the most accurate, with a more than 0.99 probability of providing the best fit to the expert opinion data among the four models considered (Table 3). The independent test of the models (cross-validation procedure) provided very different results. The questions that were retained were consistent among the different independent tests (Table 2); only two questions were not either always retained or always rejected. Similarly, the weights for the model with questions removed were somewhat consistent (small range of values among the independent evaluations; Table 2). However, the weights when all questions were retained varied substantially. This resulted in a gradual decrease in R^2 with increasing model complexity (Figure 2). Support for the model with questions removed was greatest (Table 3), and the model with questions removed and weights had greater support than the model with all questions.

Finally, the model with all questions and weights had very weak support (Table 3).

Proportional uncertainty varied among species (MS = 0.02, $F_{47, 141}$ = 4.27, p < 0.001), simply meaning that the scores for some species were more uncertain than for others, irrespective of the method used. The effect of method was also significant (MS = 0.02, $F_{3, 141}$ = 304.05, p < 0.001); proportional uncertainty was low and similar for "all questions", "questions removed" and "questions removed with weights", but was much greater for "all questions with weights" (Figure 3).

Discussion

Screening-level risk assessment (SLRA) tools can provide quick and relatively reliable information to inform policy and management decisions concerning non-indigenous species. Here we evaluated how model uncertainty, related to unknown influences of factors included in a SLRA tool, affected accuracy of CMIST, the first risk assessment tool developed explicitly for marine non-indigenous species that are usually introduced accidentally. Though the tool includes fewer questions than most of its counterparts (e.g., 17 for CMIST vs 49 for WRA-derivatives) and the questions were carefully selected to capture factors affecting both likelihood of invasion and impact of invasion, many of the questions did not improve or even decreased model accuracy. This finding is similar to that of Koop et al. (2012) who found that the answers to approximately half the
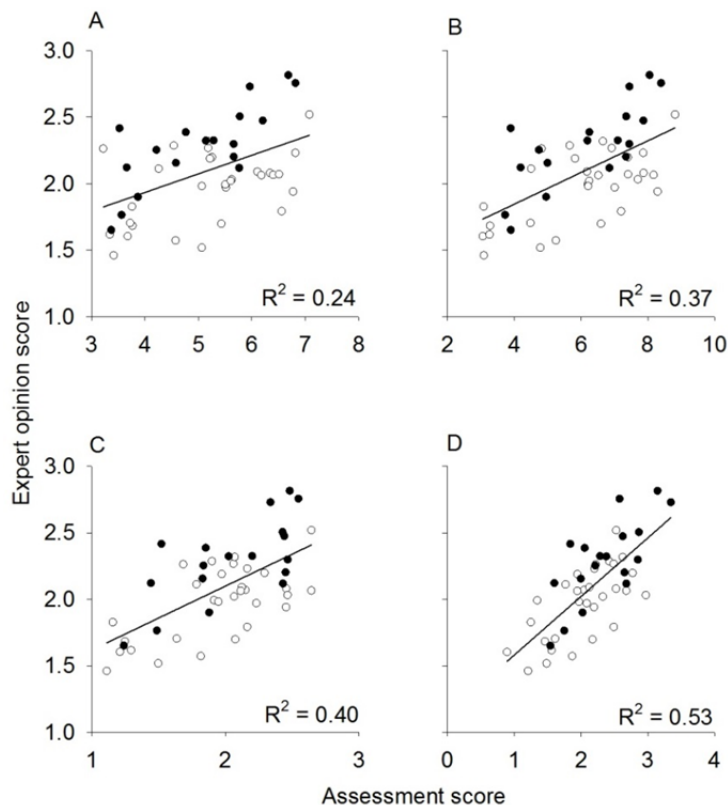
Figure 1. Relationship between expert opinion scores and assessment scores for CMIST with A) All questions, B) Questions removed, C) Questions removed with weights, and D) All questions with weights. Closed circles show species from the Canadian west coast and open circles are east coast species. Line shows best fit regression with associated R^2 values.

Table 3. Evaluation of fit of different models to the expert opinion scores for non-indigenous marine invertebrate species in Canadian coastal waters. Results of the non-independent (All species) and independent (Cross-validation) tests are presented. Likelihood is the probability that a model is the best predictor among the models evaluated.

	AICc	dAICc	AICc weight	Likelihood
All species				
All questions	-117.89	23.00	1.01E-5	<0.001
Questions removed	-127.09	13.80	1.00E-3	0.001
Questions removed with weights	-129.62	11.27	3.56E-3	0.007
All questions with weights	-140.89	0.00	1.00	0.99
Cross-validation				
All questions	-117.89	6.68	0.04	0.03
Questions removed	-124.57	0.00	1.00	0.80
Questions removed with weights	-121.43	3.13	0.21	0.17
All questions with weights	-113.52	11.05	0.004	0.003

questions in the WRA were not statistically related to pest status of plants in the United States. This does not necessarily mean that these questions pertain to factors unrelated to risk, but simply implies a lack of relationship between the answers and the data used to test tools. For example, Weber et al. (2009) found that many questions did not contribute to variation in scores because most species received the same answer. This may explain why current population status (establishment) did not contribute to accuracy for the east coast species; all species evaluated on this coast

have established populations, whereas some species from the west coast have been observed sporadically with no known established population. Similarly, all tunicate species included have a history of invasion elsewhere, explaining why this question was not influential for this taxon. Nonetheless, the results of this study suggest that some of the questions included in CMIST did not contribute to model accuracy.

Current risk assessment tools either assign an equal weight to each question or weights that are based on perceived importance (most WRA-derivatives) and,

in one case, on empirical importance; Koop et al. (2012) assigned higher weights to questions whose answers were highly statistically related to pest status. However, for a given data set, each factor has a unique influence; the R^2 maximization procedure presented in this study allows empirical determination of the combination of weights that maximizes accuracy. Deriving weights after the questions that did not contribute to accuracy were removed resulted in exclusion of three more questions (i.e., optimal weight of 0). This is probably linked to correlation in answers to these questions (collinearity). For example, species with an overall high ecological impact will have an impact on populations, communities, and ecosystems. This would explain why effect on ecosystems alone is sufficient to obtain maximum accuracy. Overall, when all species were included, model accuracy increased with increasing level of refinement. Maximum R^2 was obtained when all questions were included and weighting allowed negative values to permit contribution of questions whose answers are negatively linked to expert opinion (but see below for independent predictions).

While it is possible to optimize accuracy for any particular dataset using the methods we presented, it is unclear if the results are consistent enough to obtain more accurate independent predictions. In general, the contribution of questions to accuracy was similar among ecoregions and taxonomic groups. In addition, the cross validation procedure retained very similar models when only removing questions that did not increase accuracy. Therefore, it appears that the influence of questions is consistent enough among species to safely remove them from the tool, which results in a notable gain in accuracy, as noted elsewhere. Gordon et al. (2008) found that the classification of plants as weeds was more accurate when only one question (about history of invasion elsewhere) was used compared to when all questions (49 in total) are considered. Caley and Kuhnert (2006) derived a classification tree (based on only four of the WRA questions) with similar accuracy to the original WRA, while Weber et al. (2009) used various approaches to show that much reduced collections of questions (4 or 5) yielded similar outcomes to the complete WRA for 1844 species. Koop et al. (2012) developed a more accurate tool after removing several WRA questions. Thus, it appears that many tools are over-parameterized and many questions simply add noise and mask the discriminatory power of a few important ones.

However, further model refinement (by adding weights) did not improve independent predictions. It also increased uncertainty in the case where all questions were included and negative weighting was
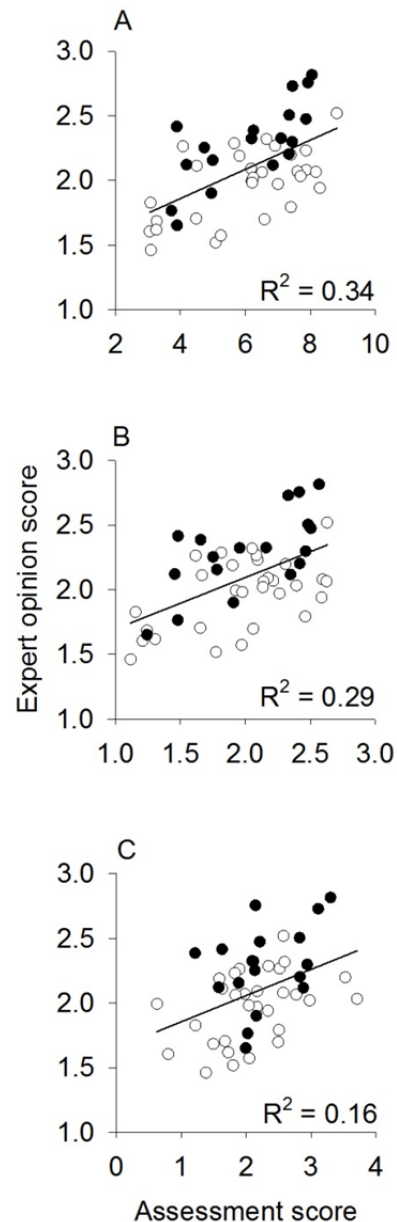

Figure 2. Relationship between expert opinion scores and independent assessment scores for CMIST with A) Questions removed, B) Questions removed with weights, and C) All questions with weights. Closed circles show species from the Canadian west coast and open circles are east coast species. Line shows best fit regression with associated R^2 values.

allowed. This might be related to the relatively small number of species included in the study; each species has a high relative influence (leverage) on the optimal weights, thus the predictions based on the rest of the species are less accurate. This is evidenced by the high variability in cross-validated weights, in particular when all questions are included and weights are

allowed to be negative. This should be tested with larger data sets (e.g., several WRA tests include hundreds of species) to see if it would result in stable weights, and ultimately how this could influence accuracy.

Ecological risk assessments are rarely evaluated *a posteriori* (Gibbs 2011). In the realm of non-indigenous species, this would involve evaluating a large number of species, waiting for one to arrive in the assessment area, and when detected, comparing the outcome with assessment results. This is highly impractical; the only other option to evaluate screening procedures is to assess species known to have been introduced into an area after the fact. This bypasses the arrival process since all species have arrived at some point, even potentially more than once. It is therefore not surprising that the question related to frequency of arrivals turned out not to contribute to accuracy here. However, there is considerable evidence that propagule pressure is an important determinant of probability of establishment (Forsyth and Duncan 2001; Lockwood et al. 2005; Colautti et al. 2006; Simberloff 2009; Britton and Gozlan 2013). Future optimized risk assessment tools should still include questions related to propagule pressure, even though the current test did not find it significant. This is particularly true for risk assessment tools used in the screening-level context, the purpose of which is to identify species not present in an area that pose the greatest risk to it.

In conclusion, we found that the original version of CMIST was over-parameterized and several questions could be ignored to improve accuracy. However, what questions to retain should be a trade-off between accuracy and other considerations. For example, resource management or conservation agencies might want to keep questions about effects on aquaculture and at-risk species (even though they did not improve accuracy in the current dataset) to adapt scores in a way to be consistent with their mandate. The results of our analyses are dependent on the species we used; it is possible that the outcome of future incursions by non-indigenous species will be influenced by factors other than those of past introductions, thus an eliminated question might represent an important predictor in the future. Also, the techniques presented result in tools optimized to expert opinions of risk; a more objective metric might be desired, but quantification of impacts of non-indigenous species is a complex task (Barney et al. 2013; Kumschick et al. 2015; Ojaveer et al. 2015). Based on these results and those of others (Caley and Kuhnert 2006; Gordon et al. 2008; Koop et al. 2012), we would recommend an in-depth examination of the importance of questions included
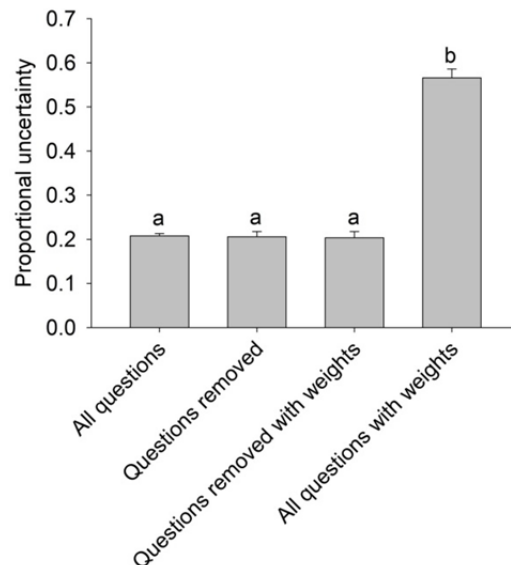

Figure 3. Influence of optimization methods of CMIST on the adjusted 95% confidence limits around species assessment scores. Error bars show standard errors and columns not sharing a common letter are significantly different (Tukey post hoc tests).

in the commonly used SLRA tools. Eliminating questions consistently found to be unimportant would make these tools more accurate and faster and easier to use. While independent scores were not improved by addition of weights here, this technique seems promising as it allows the contribution of each question to be adjusted to its real influence. If weights derived from large data sets are consistent enough, the technique has the potential to greatly improve risk assessment tools for non-indigenous species.

Acknowledgements

References

Barney JN, Tekiela DR, Dollete ESJ, Tomasek BJ (2013) What is the "real" impact of invasive plant species? *Frontiers in Ecology and the Environment* 11: 322–329, https://doi.org/10.1890/120120

Bomford M, Glover J (2004) Risk assessment model for the import and keeping of exotic freshwater and estuarine finfish. Report of the Australian Bureau of Rural Sciences, 125 pp

Bradie J, Pietrobon A, Leung B (2015) Beyond species-specific assessments: an analysis and validation of environmental distance metrics for non-indigenous species risk assessment. *Biological Invasions* 17: 3455–3465, https://doi.org/10.1007/s10530-015-0970-8

Britton JR, Gozlan RE (2013) How many founders for a biological invasion? Predicting introduction outcomes from propagule pressure. *Ecology* 94: 2558–2566, https://doi.org/10.1890/13-0527.1

Caley P, Kuhnert PM (2006) Application and evaluations of classification trees for screening unwanted plants. *Austral Ecology* 31: 647–655, https://doi.org/10.1111/j.1442-9993.2006.01617.x

Colautti RI, Grigorovich IA, MacIsaac HJ (2006) Propagule pressure: a null model for biological invasions. *Biological Invasions* 8: 1023–1037, https://doi.org/10.1007/s10530-005-3735-y

Copp GH, Vilizzi L, Mumford J, Fenwick GV, Godard MJ, Gozlan RE (2009) Calibration of FISK, an invasiveness screening tool for nonnative freshwater fishes. *Risk Analysis* 29: 457–467, https://doi.org/10.1111/j.1539-6924.2008.01159.x

Daehler CC, Denslow J, Ansari S, Kuo H-C (2004) A risk-assessment system for screening out invasive pest plants from Hawaii and other Pacific islands. *Conservation Biology* 18: 360–368, https://doi.org/10.1111/j.1523-1739.2004.00066.x

Day RW, Quinn GP (1989) Comparisons of treatments after an analysis of variance in ecology. *Ecological Monographs* 59: 433–463, https://doi.org/10.2307/1943075

DFO (2009) Development of a framework and principles for the biogeographic classification of Canadian marine areas. DFO Canadian Science Advisory Secretariat Science Advisory Report 2009/056: 17 pp, http://www.dfo-mpo.gc.ca/CSAS/Csas/Publications/SAR-AS/2009/2009_056_e.pdf

Drolet D, DiBacco C, Locke A, McKenzie CH, McKindsey CW, Moore AM, Webb JL, Therriault TW (2015) Evaluation of a new screening-level risk assessment tool applied to non-indigenous marine invertebrates in Canadian coastal waters. *Biological Invasions* 18: 279–294, https://doi.org/10.1007/s10530-015-1008-y

Forsyth DM, Duncan RP (2001) Propagule size and the relative success of exotic ungulate and bird introductions to New Zealand. *The American Naturalist* 157: 583–595, https://doi.org/10.1086/320626

Gibbs M (2011) Ecological risk assessment, prediction, and assessing risk predictions. *Risk Analysis* 31: 1784–1788, https://doi.org/10.1111/j.1539-6924.2011.01605.x

Gordon DR, Onderdonk DA, Fox AM, Stocker RK (2008) Consistent accuracy of the Australian weed risk assessment system across varied geographies. *Diversity and Distributions* 14: 234–242, https://doi.org/10.1111/j.1472-4642.2007.00460.x

Gordon DR, Gantz CA (2011) Risk assessment for invasiveness differs for aquatic and terrestrial plant species. *Biological Invasions* 13: 1829–1842, https://doi.org/10.1007/s10530-011-0002-2

Holt J, Leach AW, Knight JD, Griessinger D, MacLeod A, van der Gaag DJ, Schrader G, Mumford JD (2012) Tools for visualizing and integrating pest risk assessment ratings and uncertainties. *EPPO Bulletin* 42: 35–41, https://doi.org/10.1111/j.1365-2338.2012.02548.x

Koop AL, Fowler L, Newton LP, Caton BP (2012) Development and validation of a weed screening tool for the United States. *Biological Invasions* 14: 273–294, https://doi.org/10.1007/s10530-011-0061-4

Kumschick S, Gaertner M, Vilà M, Essl F, Jeschke JM, Pyšek P, Ricciardi A, Bacher S, Blackburn TM, Dick JTA, Evans T, Hulme PE, Kühn I, Mrugała A, Pergl J, Rabitsch W, Richardson DM, Sendek A, Winter M (2015) Ecological impacts of alien species: Quantification, scope, caveats, and recommendations. *BioScience* 65: 55–63, https://doi.org/10.1093/biosci/biu193

Kumschick S, Richardson DM (2013) Species-based risk assessments for biological invasions: advances and challenges. *Diversity and Distributions* 19: 1095–1105, https://doi.org/10.1111/ddi.12110

Lockwood JL, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution* 20: 223–228, https://doi.org/10.1016/j.tree.2005.02.004

Ojaveer H, Galil BS, Campbell ML, Carlton JT, Canning-Clode J, Cook EJ, Davidson AD, Hewitt CL, Jelmert A, Marchini A, McKenzie CH, Minchin D, Occhipinti-Ambrogi A, Olenin S, Ruiz G (2015) Classification of non-indigenous species based on their impacts: considerations for application in marine management. *PLoS Biology* 13: e1002130, https://doi.org/10.1371/journal.pbio.1002130

Pheloung PC, Williams PA, Halloy SR (1999) A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. *Journal of Environmental Management* 57: 239–251, https://doi.org/10.1006/jema.1999.0297

Regan HM, Colyvan M, Burgman MA (2002) A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* 12: 618–628, https://doi.org/10.1890/1051-0761(2002)012[0618:ATATOU]2.0.CO;2

Simberloff D (2009) The role of propagule pressure in biological invasions. *Annual Review of Ecology, Evolution, and Systematics* 40: 81–102, https://doi.org/10.1146/annurev.ecolsys.110308.120304

Tricarico E, Vilizzi L, Gherardi F, Copp GH (2010) Calibration of FI-ISK, an invasiveness screening tool for nonnative freshwater invertebrates. *Risk Analysis* 30: 285–292, https://doi.org/10.1111/j.1539-6924.2009.01255.x

Weber J, Panetta FD, Virtue J, Pheloung P (2009) An analysis of assessment outcomes from eight years' operation of the Australian border weed risk assessment system. *Journal of Environmental Management* 90: 798–807, https://doi.org/10.1016/j.jenvman.2008.01.012

Supplementary material

The following supplementary material is available for this article:

Table S1. Formulation of CMIST questions.

This material is available as part of online article from:
http://www.reabic.net/journals/mbi/2017/Supplements/MBI_2017_Drolet_etal_Supplement.xls