Ecological Informatics 48 (2018) 147–157

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

An integrative methodological framework for setting environmental criteria: Evaluation of stakeholder perceptions

Dong-Kyun Kim[a], Maryam Ramin[a], Vincent Cheng[a], Aisha Javed[a], Samarth Kaluskar[a], Noreen Kelly[a], Demetrios Kobiliris[a], Alex Neumann[a], Felicity Ni[a], Tianna Peller[a], Gurbir Perhar[a], Yuko Shimoda[a], Ariola Visha[a], Christopher Wellen[a], Cindy Yang[a], Shan Mugalingam[b], George B. Arhonditsis[a,*]

[a] Ecological Modelling Laboratory, Department of Physical & Environmental Sciences, University of Toronto, Toronto, Ontario M1C1A4, Canada
[b] Lower Trent Conservation Authority, Trenton, Ontario, Canada, K8V 5P4

ARTICLE INFO

Keywords:
Expert elicitation
Decision making
Bayesian inference
Water quality
Bay of Quinte

ABSTRACT

In water quality management, delisting decisions for impaired waterbodies bear resemblance to the statistical null hypothesis testing in which the prevailing conditions in the system at hand are compared against a *reference* (or non-impaired) state. This binary comparison presupposes the existence of a robust delineation of what constitutes a non-impaired state along with the establishment of threshold values for key environmental variables that act as surrogates for the degree of system impairment. Drawing the dichotomy between impaired and non-impaired conditions can be a challenging exercise, as it can be influenced by different trade-offs between environmental priorities and socioeconomic values. In this study, we present an integrative methodological framework that first uses mathematical modelling to reproduce the fundamental relationships between external stressors and ecosystem response and then statistically links the projected patterns with the likelihood to achieve acceptable water quality conditions. Our case study is the Bay of Quinte, Ontario, Canada; an embayment at the northeastern end of Lake Ontario with a long history of eutrophication. Our survey was conducted during the winter and spring of 2014 and included a total of 51 individuals who were actively involved with the Bay of Quinte Remedial Action Plan. Our analysis found that there is a perceptual difference between public and experts in that the latter group tends to more favorably characterize the present conditions, but is also more conservative about the delisting prospects of the system. Statistical analysis showed an average level of confidence lower than 50% about the delisting likelihood when experts were asked to assess the current water quality criteria. We conclude by arguing that the occurrence frequency of extreme conditions (i.e., exceedance of maximum allowable nutrient levels and/or toxic algal blooms) should be an integral component of the delisting process.

1. Introduction

In the Great Lakes region, many management actions have focused on improving the prevailing ecological conditions over the past four decades (EC and USEPA, 2013). The International Joint Commission (IJC) designated Areas of Concern (AOCs), where significant environmental degradation occurred from anthropogenic activities at the local level (IJC, 1985, 2003). Management processes revolved around the restoration of impaired beneficial uses, referred to as Beneficial Use Impairments or BUIs. These BUIs largely suggest poor water and sediment quality, loss of habitat, and/or impairments that may have adverse effects on human health (George and Boyd, 2007). Delisting of an AOC can take place only once all "*specific, measurable, achievable and scientifically defensible*" actions have been undertaken and environmental conditions are comparable to those at similar non-AOC sites (EC and OMOE, 2007; George and Boyd, 2007).[1] The pillars of the process in restoring and maintaining beneficial uses are: (i) the establishment of specific environmental goals or "desirable" standards using measurable indicator variables; (ii) the determination of the optimal remedial actions vis-à-vis the targeted beneficial use impairments; and (iii) the

* Corresponding author.
E-mail address: georgea@utsc.utoronto.ca (G.B. Arhonditsis).

[1] Delisting of an AOC refers to the situation in which all remedial actions have been completed and monitoring data indicate that all beneficial uses are restored. Under this condition, the geographical region is no longer considered impaired. Generally, assessment for delisting involves the same criteria, guidelines, and thresholds that rendered the area as an AOC in the first place (EC and OMOE, 2007).

https://doi.org/10.1016/j.ecoinf.2018.08.005
Received 17 April 2018; Received in revised form 3 August 2018; Accepted 7 August 2018
Available online 11 August 2018
1574-9541/

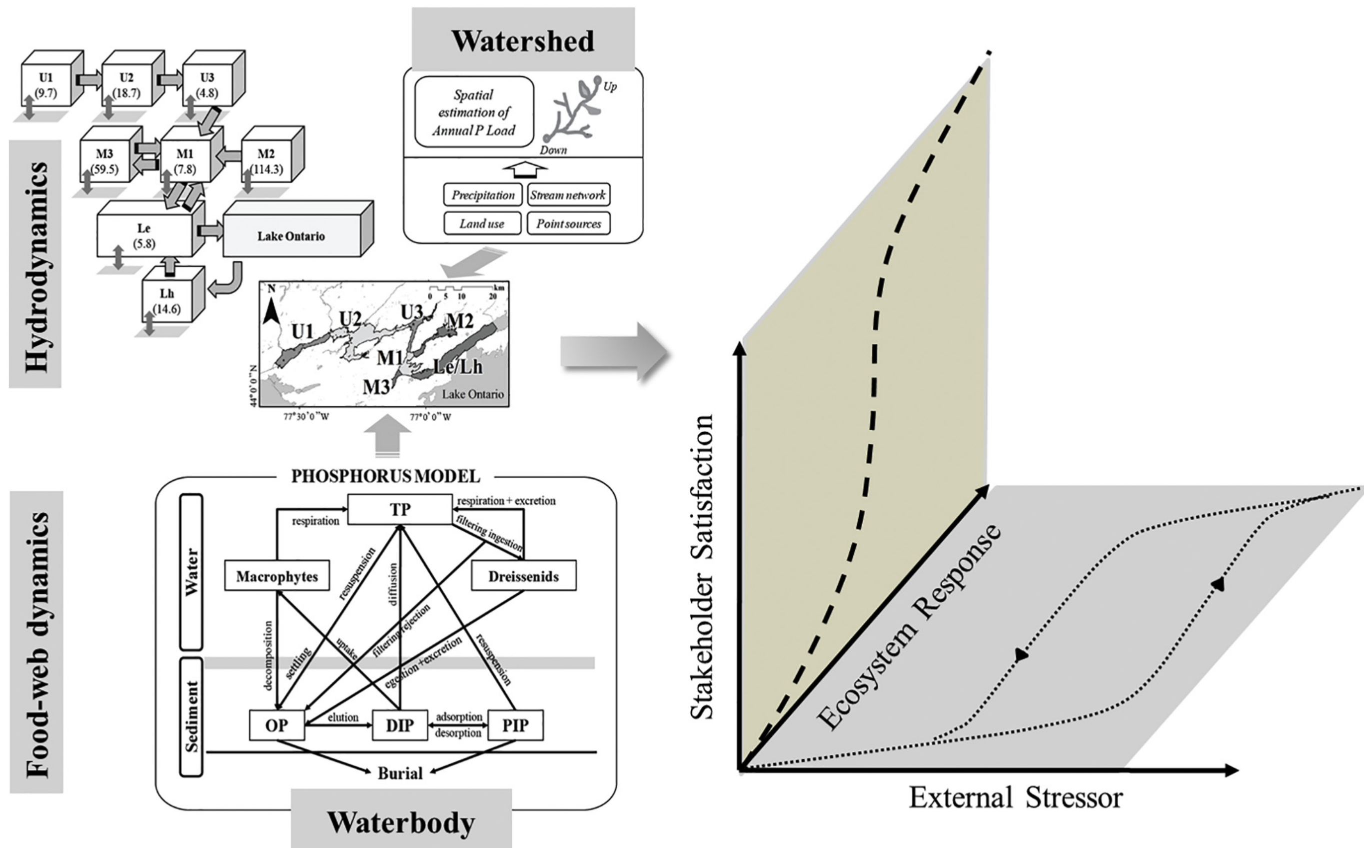


Fig. 1. Conceptual diagram of the integrated modelling framework proposed for delisting Areas of Concern for different Beneficial Use Impairments. Empirical estimates of the attainability of water quality targets are derived from a stakeholder elicitation exercise. Our study serves as a proof of concept of this integrative framework using the Bay of Quinte, Ontario, Canada, as a case study. Segments U1, U2, U3, M1, M2, M3, and Le/Lh correspond to the spatial compartments of the eutrophication model (Arhonditsis et al., 2016; Kim et al., 2013).

development of a decision making process that capitalizes upon actual monitoring data to assess the attainment of locally-derived delisting targets (USPC, 2001).

Despite several decades of active management actions, contentious issues related to the delisting criteria across all the AOCs in the Great Lakes area continue to persist. Restoration targets and associated guidelines for specific BUIs have been established independently in different AOCs (Chen and Mackay, 2004; USPC, 2001). This is highly evident for the water quality criteria, which significantly vary across AOCs, such as the Bay of Quinte (30 μg TP L^{-1} and 12–15 μg Chl*a* L^{-1}), Hamilton Harbour (initial targets: 34 μg TP L^{-1} and 15–20 μg Chl*a* L^{-1}; final targets: 20 μg TP L^{-1} and 5–10 μg Chl*a* L^{-1}), Saginaw Bay (15 μg TP L^{-1}), St. Louis River (30 μg TP L^{-1}), Toronto Harbour (30 μg TP L^{-1}), and White Lake (30 μg TP L^{-1}, 10 μg Chl*a* L^{-1}, Secchi Disk depth = 2.0 m, Trophic State Index by Carlson ≈ 50–55) (Chen and Mackay, 2004; MDEQ, 2008; Gudimov et al., 2011). Delisting objectives or guidelines can be characterized by a certain degree of subjectivity arising from the preferences of individual stakeholders.[2] In fact, vocal support and excessive lobbying from individual leaders among stakeholders can exert a large influence, the so-called "participation-leadership effect", on the water quality criteria setting process that may undermine environmental decision-making (Hall et al., 2006; Mullen, 1991). Therefore, the development of more comprehensive and analytic modelling frameworks for the evaluation of stakeholder opinions is essential for establishing impartial environmental standards that will be relevant across all jurisdictions (Fig. 1).

The management of natural systems often involves policy analysis and decision making in the face of considerable uncertainty arising from a number of sources (Morgan et al., 1992). Uncertainty is a generic term that consists of a multitude of concepts. For instance, the *variability* that environmental quantities demonstrate over time and space may consist of: *random error*, in direct measurements; *systematic error,* introduced by the measuring apparatus and/or experimental protocols; *inherent randomness* or *indeterminacy,* which is often considered the product of our incomplete knowledge of the world; *approximation uncertainty,* reflecting the assumptions made and imperfect knowledge used to understand the structure and inputs of the impaired environmental system; and *subjective judgments,* which are used to overcome the knowledge gaps and lack of empirical measurements related to the major ecological mechanisms and/or variables underlying the environmental problem at hand (Reckhow et al., 2005). Significant uncertainty also arises from potential *disagreements* among decision makers and stakeholders, reflecting their different perspectives and conscious (or unconscious) biases as to what represents a non-impaired state or –even more fundamentally– how closely delisting objectives are connected to an ecosystem's particular beneficial use (Borsuk et al., 2001; Lerner et al., 2011).

Regarding the uncertainty arising from decision makers and stakeholders, expert elicitation is a formal process for synthesizing qualitative and subjective judgments that are surrounded by uncertainty resulting from limited knowledge and/or lack of available resources (Meyer and Booker, 1991). This methodology is very useful in synthesizing opinions pertaining to the decision making process, especially

[2] Stakeholders in this study represent the proportion of the participants who were highly aware and/or directly involved in the restoration efforts of the Bay of Quinte. This subset of the participants were involved in the study without any added incentive (refer to Section 2.2. Stakeholder Preferences for more information)

when robust and objective socio-environmental models are unavailable or compromised by the lack of appropriate hard data (Burgman, 2005; Martin et al., 2012). With regard to water-policy analysis, expert elicitation has been used to combine different judgments related to the status of an ecosystem, thereby quantifying the attainment of a narrative designated use (Reckhow et al., 2005). Given the complex decisions typically involved in water resources management, an expert-based approach is suitable for collecting judgments from water quality experts and extracting objective relationships among subcomponents of a system (Van Houtven et al., 2014). In particular, the Delphi process, an expert-based method, has been used extensively in data-poor environments to develop consensus among experts over several rounds of investigation, under the assumption that combining the expertise of multiple individuals will provide more representative results than consulting a single individual (MacMillan and Marshall, 2006; Meyer and Booker, 1991).

Since multiple expert judgments, as noted above, often introduce conflicting information or opinions, much of the research in the field of expert elicitation has been directed toward articulating methodological guidelines to impartially integrate across expert opinions (Cooke, 1991). Also recognizing the problems that may arise from the use of multiple experts in risk analysis, Keith (1996) stressed that a key barrier to incorporating the elicited judgment is uncertainty derived from divergent judgments. However, this study underscored the importance of seeking alternative modes of policy analysis rather than pursuing improved methodologies for combining multiple expert judgments. Clemen and Winkler (1999), in contrast, reviewed a wide range of methods for combining experts' probabilty distributions in risk analysis and highlighted that the inclination to collect information to the utmost poses as one of the main driving forces behind the use of multiple experts. Whereas Keith (1996) noted that the results of combined expert judgments may be meaningless if experts' responses are inconsistent or fundamentally different, the study by Clemen and Winkler (1999) asserted that a Bayesian aggregation allows for careful control of the quality of expert judgments by adjusting individual expert distributions and phasing out the impact of the well-known phenomenon of "overconfidence" (Clemen and Reilly, 1999).[3] The nature of Bayesian inference also relates to the concept of adaptive management, in which decisions are sequentially modified as new sources of information become available (Ellison, 1996). In the context of risk analysis and decision-making, Kuhnert et al. (2010) also stressed that Bayesian modelling is capable of minimizing potential bias by using informative priors of existing knowledge (McCarthy and Masters, 2005), even if linguistic and epistemic uncertainties could lead to bias when eliciting information from experts (Regan et al., 2002). Moreover, Martin et al. (2012) addressed that Bayesian methods best accommodate updating judgments in light of new empirical information because they broadly define subjective probabilities. Thus, a Bayesian modelling framework can offer a natural platform for adaptively learning and assessing goal/target achievability based on the elicited expert judgments.

Taking these concerns about the role of uncertainty in the decision making process into account, our study illustrates an integrated framework that combines stakeholder values, expert judgment, and mathematical modelling to predict achievability of water quality targets. Our case study is the Bay of Quinte, one of the Canadian AOCs in Lake Ontario, which has experienced significant water-quality improvements since its original designation as an impaired system in 1987 (BQRAP, 1987). Yet, much progress remains to be made as 10 out of 14 beneficial uses remain impaired (BQRAP, 2015). With these impairments in mind, the Bay of Quinte Remedial Action Plans strive for AOC-delisting along with the current water quality targets of lower than 30 μg TP L^{-1} and 10–12 μg Chl*a* L^{-1}. In this regard, our modelling framework comprises a two-pronged approach that first uses mathematical modelling to reproduce the fundamental relationships among external stressors and ecosystem response. We then statistically link the projected patterns with the likelihood of achieving acceptable water-quality conditions. In order to estimate the latter likelihood, we elicit expert and stakeholder opinions through a carefully choreographed series of interviews with government and university scientists familiar with the waterbody. A major undertaking of the present study has also been to engage local stakeholders and policy makers with the criteria setting process in order to include their perspectives on the identification (or possible change) of the optimal delisting criteria that can effectively balance environmental aspirations with socioeconomic priorities.

2. Methodological framework

The general scheme of our modelling framework, for assessing the goal achievability, is comprised of four main modules: an integrated watershed-receiving waterbody modelling based on our best scientific knowledge (Section 2.1), a questionnaire related to stakeholder preferences (Section 2.2), an expert elicitation (Section 2.3), and the projection of ecosystem responses using the integrated process-based model in conjunction with stakeholder/expert judgments. The first module emulates the Bay of Quinte by predicting water quality in response to climatic and hydrological conditions. The second module provides useful information regarding the stakeholders' opinions of the current status of Bay of Quinte based on the present conditions of beneficial use indicators of ecosystem health. The third module engages quantification of elicited expert judgments based on water quality values generated from observed data. Finally, the fourth module frames a prognostic tool through the connection between the first and third module. From the mechanistic process-based model (which represents our scientific knowledge) to the combination of multiple expert judgments (which corresponds to stakeholders' perception more quantitatively), our framework reflects the multiple facets of ecosystem assessment and therefore can be used to project achievability of BUI removal with much greater confidence.

2.1. Integrated watershed-receiving waterbody modelling

The role of the first module is to provide a predictive device for water quality variables of management interest that is founded upon the complex interplay among hydrodynamics, chemistry, and biology. Several process-based models have been developed to predict water quality in the Bay of Quinte (Minns and Moore, 2004; Kim et al., 2013; Zhang et al., 2013). We used the most updated eutrophication model that explicitly accommodates both external nutrient loading related to watershed attributes (e.g., land use and management practices) and internal phosphorus cycling associated with the role of macrophytes, dreissenids, and sediment diagenesis (Arhonditsis et al., 2016). To account for the impact of exogenous nutrient loading on the Bay, we used the SPAtially Referenced Regressions On Watershed attributes (SPARROW) non-linear regression strategy to estimate nutrient loads, yields, and deliveries at both reach and subwatershed levels (Arhonditsis et al., 2016; Kim et al., 2017). In the context of our integrative methodological framework, the first module allows us to project water quality conditions in the Bay of Quinte stemming from various external loading scenarios, hydrodynamic regimes, and internal nutrient recycling conditions. These predictions along with the associated uncertainty can then be used to estimate our confidence in achieving acceptable water-quality conditions based on the expert judgments.

[3] Overconfidence refers to the phenomenon in which experts state narrower confidence intervals than expected based on their level of knowledge and experience in the matter. This issue can significantly undermine expert judgments (Lin and Bier, 2008).

2.2. Stakeholder preferences

The second module provides a comprehensive assessment of the stakeholder sentiment in regard to the current progress and outstanding challenges with the management of the Bay of Quinte. The main objective of this module is to elucidate the facets of ecosystem functioning that are deemed impaired and to translate this information into quantifiable criteria. Stakeholders were selected from different organizations (i.e., Lower Trent Conservation, Department of Fisheries and Oceans Canada, and Ontario Ministry of the Environment and Climate Change) on the basis of their involvement with the Bay of Quinte Remedial Action Plan for the past five years. These individuals were identified through our professional network, and were contacted in person or via e-mails about voluntary participation in the study. They were told that they were under no obligation to participate and that if at any time they became uncomfortable in providing their judgment and would prefer not to answer the questions, they could remove themselves from participation in the interview. Although this may have limited our sample size, it allowed us to filter out candidates who may be unsuitable for the study and instead select for the portion of stakeholders who were the most interested (reflected through their willingness to participate in the study with no added incentives), and therefore the most engaged in matters pertaining to the current status of the Bay of Quinte. The survey consisted of 14 questions that aimed to assess a respondent's level of knowledge (i.e., expertise and familiarity) with our case study, pinpoint their opinion of the ecological mechanisms impacting water quality conditions, identify the BUIs that should be prioritized based on their broader socio-environmental impact, and to delineate the optimal management strategies for restoring and maintaining the integrity of the Bay of Quinte (Please refer to Section B of Supporting Information). Overall, a total of 42 stakeholders participated in the study and their response rate to the questions in the survey ranged between 35% and 60% (mean response rate: 45%, and standard deviation: 9%). Specifically, question 4 from the survey provided insight into the stakeholders' opinions on water-quality conditions in the Bay of Quinte and this information is illustrated in Fig. 2a. Survey question 10 provided information on the changes observed by the interviewed stakeholders within the last 5 years at the Bay of Quinte (Fig. 2b). The information offered by these two survey questions were corroborated with survey question 5, and resulted in the identification of the stakeholders' opinions on the most important BUI requiring the greatest attention in the Bay of Quinte within the next 5 years (Fig. 2c). This identified BUI was used to select the most relevant water-quality variables for the expert-elicitation exercise presented in the third module.

2.3. Expert elicitation

The overarching goal of our expert elicitation exercise was to quantify the judgments of individuals, who have in-depth knowledge of the water quality problems in the studied system. An underlying assumption of our study was that the group of experts in this exercise would be somewhat disconnected in their knowledge of management and policy priorities for the Bay of Quinte. Instead, these experts were thought to be much better and more knowledgeable than the public/laypersons and stakeholders in terms of quantifying the degree of water quality impairment by translating qualitative assessment into measurable values. Expert judgments are therefore assumed to be less divergent and more reliable sources of information than the judgments from either the public or the stakeholders. This exercise was intended to generate a statistical relationship that links several candidate water quality variables, e.g., total phosphorus (TP), chlorophyll *a* (Chl*a*) microcystin concentrations, and water transparency, to their empirical knowledge of eutrophic conditions. Given that no single variable consistently demonstrated to be the best predictor of a trophic state, our expert elicitation model could encompass one or more water quality variables as predictors. The predictand, on the other hand, was exclusively an estimate of the likelihood of non-impairment based on the experts' best judgment. Compared to the layperson (Ramin et al., 2018), expert limnologists/water quality specialists typically have significant experience and quantitative understanding of eutrophication surrogates, such as algal biomass, microcystin concentrations, water clarity, and nutrient concentrations. The functional relationship derived from the expert knowledge, as a result, can help in delineating the prevailing conditions of the desirable (or non-impaired) state in the Bay of Quinte. We conducted the elicitation with 11 experts familiar with freshwater eutrophication. Two modelling experts out of 11 were also part of the stakeholders group because they were highly involved with the Bay of Quinte projects. This implies some partial commonality in judgments between stakeholders and experts. After an information session in which we presented all the recent advances in our understanding of the Bay of Quinte ecosystem functioning along with the basic objectives of our exercise, the experts were asked to express their views on the likelihood of restoration based on different scenarios representing various levels of five water quality variables. Specifically, the experts were asked to consider total phosphorus, chlorophyll *a*, and ammonium concentrations as well as Secchi Disk depth and cyanobacteria toxins (microcystin concentrations).[4] All 11 experts consented to responding completely to 100 rows of data that were designed to reflect realistic combinations of the five variables; that is, individual snapshots of recorded levels from samples collected from the regular monitoring of the Bay of Quinte. For each row, the experts were asked the following question (Please refer to Fig. SI-1 for more details):

- "*How many times (out of 100) will the Bay of Quinte be in attainment of its Beneficial Uses, when the summer average levels of the five water quality variables (Total Phosphorous, Chlorophyll a, Ammonium, Secchi Disk depth, Microcystin) are those reported in each row?*"

The experts had one week to submit their completed assessments of the 100 data lines. Through this question and given the underlying correlation structure of the water quality data in the waterbody, we anticipated to capitalize on the experts' best professional understanding of eutrophication and aimed to solicit values for the probability of beneficial-use attainment. To this end, we used a Bayesian hierarchical framework to link water-quality conditions with goal achievability in the Bay of Quinte (as depicted by the judgments of the 11 experts). The hierarchical nature of our modelling was intended to accommodate the differences in the interviewees' independent assessments, while reflecting the fact that all the experts have good knowledge of the relationships among the fundamental limnological variables, as well as the Bay of Quinte ecosystem dynamics (Gelman and Hill, 2007). Therefore, some commonality in their responses is expected.

Our hierarchical model postulates a linear relationship between the expert judgments and the water-quality variables of interest and is mathematically formulated as follows:

$$P_{i,j} \sim N(p_{i,j}, \sigma^2)$$

$$p_{i,j} = \alpha_i + \beta_i \cdot X_j$$

$$p_j = \frac{\sum_{i=1}^{N} p_{i,j}}{N}$$

$$\alpha_i \sim N(\alpha, \omega_{1i}^2) \quad \beta_i \sim N(\beta, \omega_{2i}^2)$$

$$\alpha \sim N(\mu_\alpha, \omega_1^2) \quad \beta \sim N(\mu_\beta, \omega_2^2)$$

[4] These parameters were selected as water-quality indicators based on the public survey presented by Ramin et al. (2018), where the majority of both residents and tourists linked these five measurable variables to the water quality issues in the Bay of Quinte.

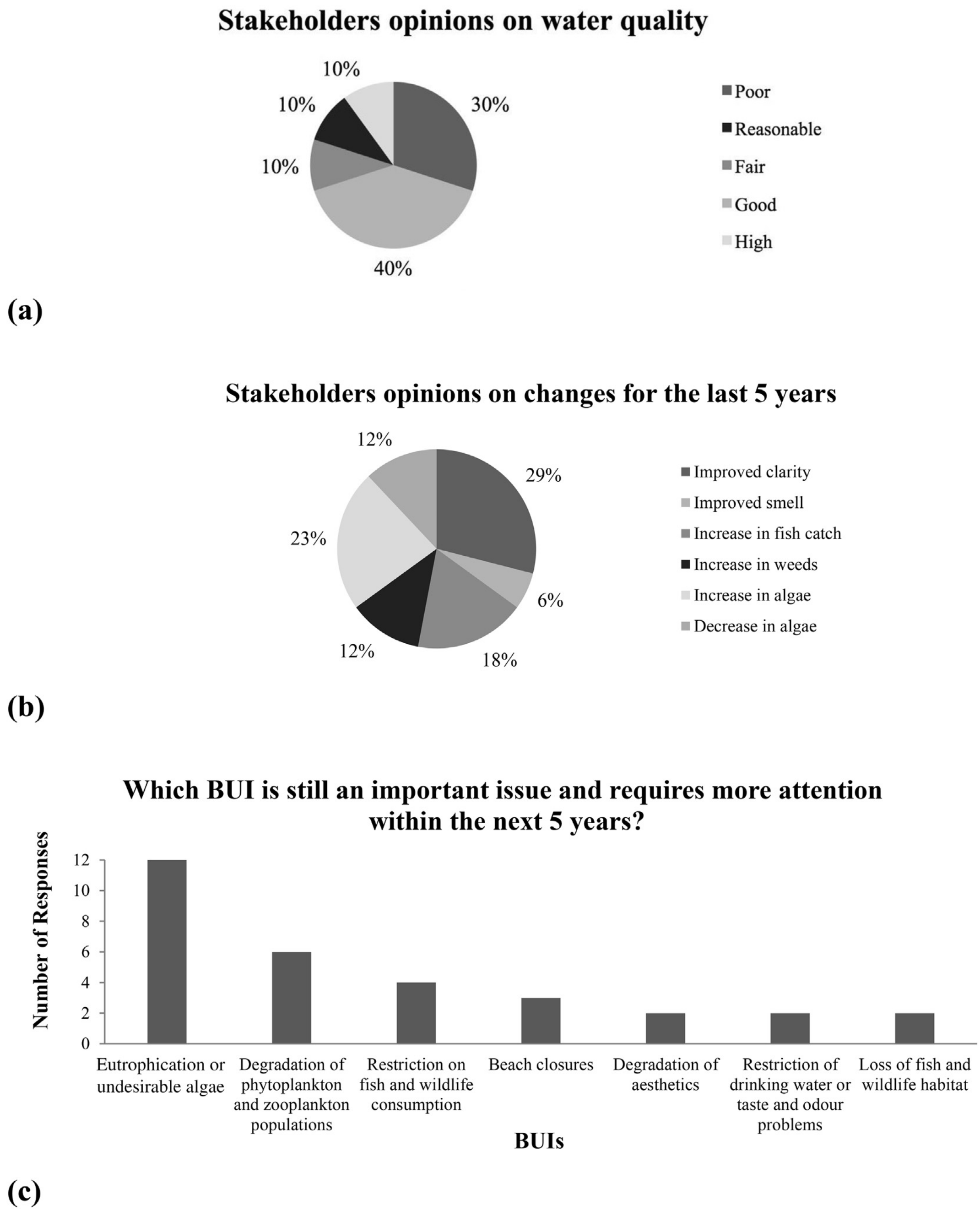


Fig. 2. Stakeholder opinions (a) on water quality ($n = 10$), based on results from expert elicitation survey question 4; (b) on changes that have been observed within the last five years ($n = 17$), based on results from survey question 10; and (c) on the improvement for specific BUIs ($n = 31$), constructed using results from survey question 5.

$$\mu_\alpha, \mu_\beta \sim N(0, 10000)$$

$$\sigma^{-2}, \omega_{1i}^{-2}, \omega_{2i}^{-2}, \omega_1^{-2}, \omega_2^{-2} \sim G(0.001, 0.001)$$

$N = 1$, ……,11 *experts*

where $P_{i,j}$, representing the response of expert i, indicates the delisting likelihood (%) of the BUI removal (i.e., non-impairment) for the Bay of Quinte based on the multivariate water-quality observation j; σ^2 represents the model error variance, $p_{i,j}$ the predicted delisting probability, as determined by the water-quality variables X_j presented to the experts in row j; α_i and β_i are the expert-specific slopes and intercepts; the expert-specific regression coefficients are treated as draws from normal distributions with global means (α and β) and expert-specific variances ($\omega_{1i}{}^2$ and $\omega_{2i}{}^2$), respectively; μ_α, μ_β and $\omega_1{}^2$, $\omega_2{}^2$ are the mean and variance of the hyper parameters, respectively; N(0, 10,000) is the normal distribution with mean 0 and variance 10,000, and G(0.001, 0.001) is the gamma distribution with shape and scale parameters of 0.001. These prior distributions are considered "non-informative" or vague. Finally, $p_{i,j}$ predictions were averaged across the experts (p_j) in order to draw unbiased inference of the attainability of impaired conditions in the Bay of Quinte.

Using Markov-chain Monte Carlo (MCMC) simulations (Gilks et al., 1998), we obtained sequences of realizations from the model posterior distributions. We used a general normal-proposal Metropolis algorithm that is based on a symmetric normal-proposal distribution, whose standard deviation is adjusted over the first 4000 iterations, so that the acceptance rate ranges between 20 and 40%. For each analysis, we used three chain runs of 50,000 iterations, keeping every 10th iteration (thin of

10) to minimize serial correlation. Convergence of the MCMC chains was checked using the Brooks-Gelman-Rubin (BGR) scale-reduction factor (Brooks and Gelman, 1998). The BGR factor is the ratio of between-chain variability to within chain variability. The chains have converged when the upper limits of the BGR factor are close to one. The accuracy of the posterior parameter values was inspected by assuring that the Monte Carlo error (an estimate of the difference between the mean of the sampled values and the true posterior mean) for all parameters was < 5% of the sample standard deviation (Spiegelhalter et al., 2003).

3. Results

3.1. Stakeholder perception survey

Stakeholders' opinions (from module 2) on the present water-quality conditions in the Bay of Quinte were generally positive (Fig. 2a). Among those interviewed, 40% believed that the water quality is “good”, while another 30% collectively felt that the prevailing water-quality conditions in the bay can be classified as “fair”, “reasonable”, and “high”. Additionally, only 30% believed the water quality to be "poor" (Fig. 2a). The majority of the interviewed stakeholders noticed positive changes to water clarity, smell and fish catch over the past five years, although they also noticed algae and weeds to be increasing (Fig. 2b). As a result, stakeholders believed that the BUIs “*Eutrophication or Undesirable algae*”, and to a lesser extent, “*Degradation of phytoplankton and zooplankton populations*” and “*Restriction on fish and wildlife consumption*” are still important issues that require attention during the next five years (Fig. 2c).

3.2. Expert elicitation modelling

Following this result, the elicitation model aimed to predict the likelihood of delisting the Bay of Quinte, based on the best judgment of the 11 experts in characterizing the impairment status of the system for the beneficial use “*Eutrophication or Undesirable algae*” (hereafter referred to as BUI #8) from the water-quality snapshots provided. According to the experts' elicited responses, TP and Chl*a* concentrations were the key surrogate variables to determine the likelihood of delisting the system. The mean predicted frequency (and the associated 95% credible intervals) of attaining satisfactory water-quality conditions as a function of TP and Chl*a* concentrations are shown in Fig. 3. The mean delisting probability is higher than 60% only when TP is lower than $10\,\mu g\,L^{-1}$, which has been recorded in < 10% of the samples collected from the Bay of Quinte over the past five years (Kim et al., 2013). The mean likelihood of attaining acceptable water-quality conditions gradually varies from approximately 60% to 45%, when ambient TP lies within the $10–25\,\mu g\,L^{-1}$ range, and is further reduced from 45% to 35% between 25 and $40\,\mu g$ TP L^{-1}. The same model predicts that the mean frequency of attainment could drop to 30% (or lower) when TP concentrations exceed the level of $50\,\mu g\,L^{-1}$ (left panel in Fig. 3a). Interestingly, the predictive uncertainty of the delisting probability increases as the TP concentrations increase, from a range of ± 10% at low TP $\sim 10\,\mu g\,L^{-1}$ to greater than ± 20% with TP > $50\,\mu g\,L^{-1}$ (see predictive credible intervals in the left panel of Fig. 3a). It is also worth mentioning that the predicted delisting probability at the current water-quality standard of $30\,\mu g$ TP L^{-1} was slightly higher than 40%, with substantial predictive uncertainty range (± 15%). When chlorophyll *a* concentrations are lower than $5\,\mu g\,L^{-1}$, the mean frequency of attainment of a non-impaired state is > 50%. With higher levels of phytoplankton biomass, the mean delisting probability gradually declines by an approximate rate of 1% per μg Chl*a* L^{-1}, until it asymptotically reaches a mean frequency of 35% at > 20 μg Chl*a* L^{-1}. Notably, according to the assessment of the experts, the mean predicted frequency of attainment at the current delisting target of 10–12 μg Chl*a* L^{-1} was lower than 50%, with considerable predictive uncertainty ± 20% (right panel in Fig. 3a).

3.3. Linking the expert elicitation model with the integrated watershed-receiving waterbody model

The delisting likelihood was evaluated at two locations (U2, U3) in the upper Bay of Quinte, based on the expert elicitation model, using the predicted Chl*a* and TP concentrations from the integrated watershed-receiving waterbody model as inputs. These model inputs were in turn forced with an assortment of tributary TP concentrations and flushing rates (Fig. 3b and Fig. SI-2). The two locations have been experiencing high TP (> $50\,\mu g\,L^{-1}$) and Chl*a* (> $30–40\,\mu g\,L^{-1}$) concentrations, while internal nutrient recycling (mediated by macrophytes, dreissenids, and sediment diagenesis processes) contributes a significant amount of phosphorus (Arhonditsis et al., 2016; Kim et al., 2013; Zhang et al., 2013). Recent empirical studies have noted that more frequent cyanobacterial blooms, such as toxic *Microcystis*, were observed in these locations despite a great deal of effort to reduce TP loading over the past four decades (Shimoda et al., 2016). Our previous modelling studies (Arhonditsis et al., 2016; Kim et al., 2013) were based on two years, 2005 and 2008, which exhibited contrasting patterns of summer TP levels, with maximum $\sim 60\,\mu g\,L^{-1}$ in the former and $\sim 40\,\mu g\,L^{-1}$ in the latter year.

We first used the Chl*a* concentrations to draw delisting predictions from the expert elicitation model (Fig. 3b). In the middle-upper Bay of Quinte (U2), the likelihood of achieving non-impaired conditions was fairly constant (~40–42%) throughout the range of inflow TP concentrations and flushing rates examined, with the rest of the conditions resembling those that prevailed in 2005 (left-top panel, Fig. 3b). In 2008, the mean predicted probability slightly increased (42–45%), depending on the interplay between inflow TP concentrations and estimated flushing rates (left-bottom panel, Fig. 3b). In the eastern part of the upper Bay of Quinte (U3), the predicted delisting likelihood was lower than 40% during the conditions experienced in 2005 (right-top panel, Fig. 3b). The same probability was distinctly higher (42–45%) when the model was forced with the 2008 conditions (right-bottom panel, Fig. 3b). Similar conclusions can be reached when using the bay's predicted TP concentrations to draw delisting predictions from the expert-elicitation model (Fig. SI-2). Overall, we found that the hydrodynamic regime was a particularly influential factor of the joint predictions of the expert elicitation-integrated watershed/receiving waterbody model. In particular, the delisting predictions are distinctly more favorable when the simulated spatial segments are flushed with frequency > 2.5–3.5 times during the growing season (May–October). By using this information, decision makers can establish criteria that are more realistic and representative of the water quality issues pertaining to the Bay of Quinte ecosystem.

4. Discussion

Viewed from the perspective of decision making in the face of uncertainty, environmental-standard setting typically involves a dynamic interactive process between scientific/professional knowledge and stakeholder/public opinion. The determination of beneficial uses, in principle, reflects the public perception regarding the realization of desirable physical, chemical, and biological characteristics of the waterbody. However, the selection of criteria to assess beneficial uses is a choice based largely on science; that is, a good criterion is one that is easily and reliably measured and is an objective indicator of the beneficial use of interest. In this context, our methodological protocol for BUI removal (ultimately leading to AOC delisting) explicitly accommodates the multi-dimensional nature of policy analysis. The various facets of our framework draw upon our scientific knowledge and understanding, public preferences, stakeholder values, and rigorous uncertainty assessment.

To date, many past studies have involved environmental management and policy making in various elicitation frameworks. Expert elicitation has been employed to deal with several environmental

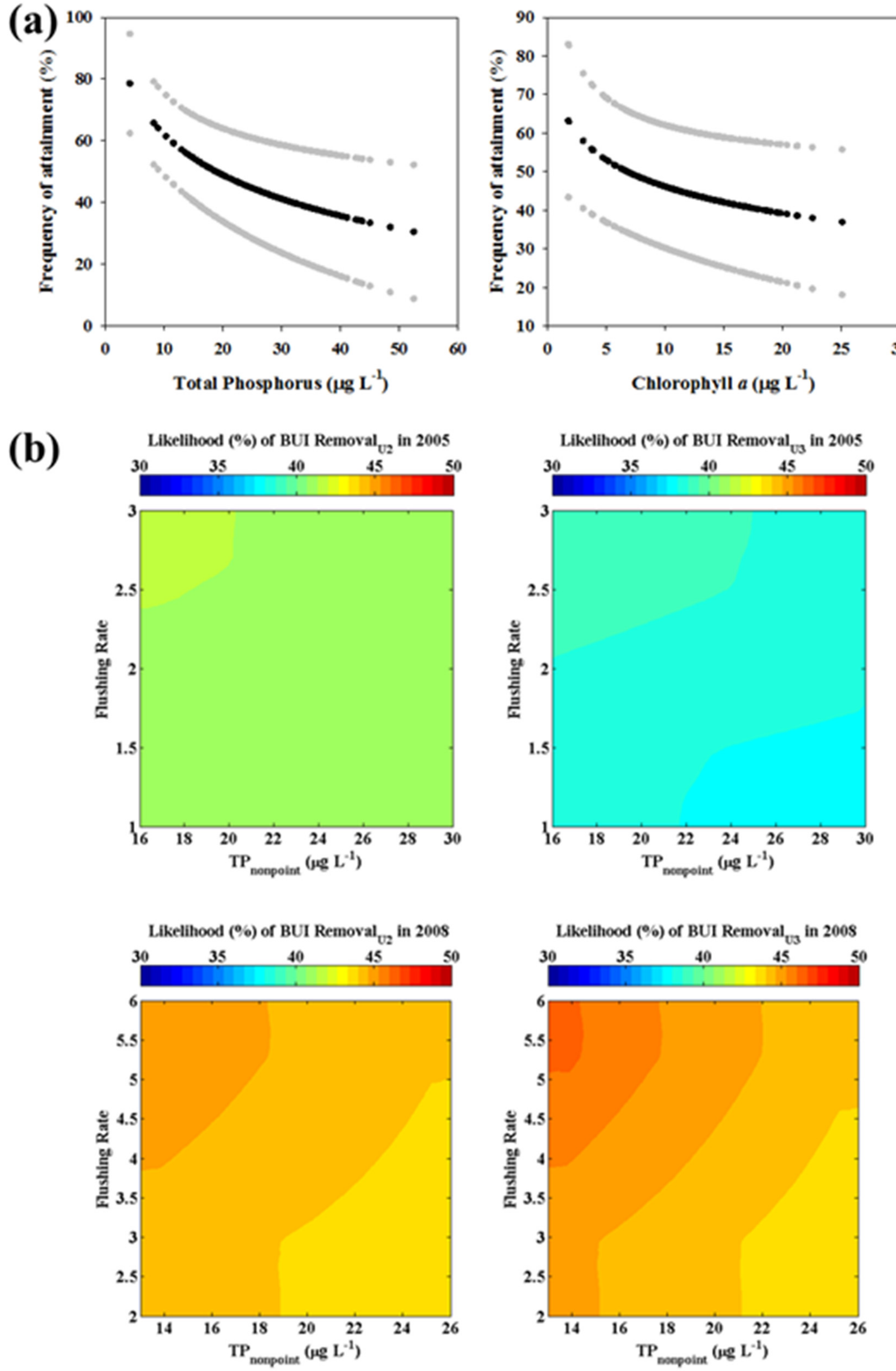


Fig. 3. (a) Predicted frequency -mean prediction and 95% credible intervals- of attaining acceptable water quality conditions as a function of TP (left panels) and chlorophyll *a* (right panels) concentrations in the Bay of Quinte, based on expert elicitation responses ($n = 11$). (b) Likelihood (%) of delisting the upper Bay for the BUI "*Eutrophication or Undesirable Algae*". The predicted chlorophyll *a* concentrations of the Bay in response to inflow TP concentrations and flushing rates (or number of times the system flushes during the growing season between May and October) at two locations of the upper Bay in 2005 (upper panels) and 2008 (lower panels), are linked with the expert elicitation model.

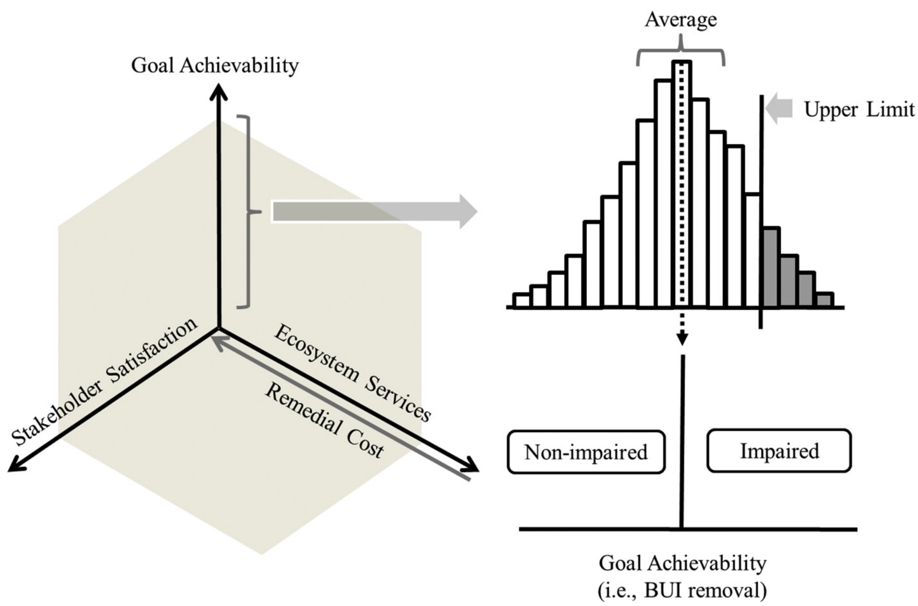


Fig. 4. Conceptual diagram of the proposed shift in our ecosystem restoration paradigm. The traditional perspective represents a two-dimensional space, with axes accounting for the ecosystem integrity and human perception of ecological conditions. In the management of the Great Lakes, the BUIs represent our knowledge of the ecosystem state along with the values of the associated services, while the second dimension is the stakeholder satisfaction, as captured by the variability among the different AOCs and may be ranked by the level of degradation (e.g., degree or type of contamination issues, number of BUIs, or agricultural versus urban impacts). Histograms represent the distribution of indicator values used to track the progress with a particular BUI. The introduction of uncertainty with respect to goal achievability, as a third dimension, aims to shift the focus of management from the "average conditions", establishing the dichotomy between impaired and non-impaired conditions (bottom histogram), to the violation likelihood of a particular upper limit target (top histogram).

problems, such as fish-health assessment (Borsuk, 2004), attainment of designated uses (Kenney et al., 2009; Reckhow et al., 2005), or climate-change adaptation (Doria et al., 2009). In this context, water-quality research has mainly aimed to elicit public willingness-to-pay (WTP) or willingness-to-accept (WTA), which are predominantly influenced by the trade-off between environmental and economic benefits (Carson and Mitchell, 1993; Cooper et al., 2004; Del Saz-Salazar et al., 2009). In the United States, Carson and Mitchell (1993) used the contingent-valuation method to determine the national benefits (e.g., recreational, commercial, and industrial uses) via elicitation of people's WTP. In Europe, the Water Framework Directive applied this method to assess the public WTP and WTA, thereby assisting with the design and implementation of effective water-management policies (Del Saz-Salazar et al., 2009). Similar to these studies, Reckhow et al. (2005) and Kenney et al. (2009) added an "*ecological flavour*" to the identification of water-quality criteria by introducing an additional layer of intertwined ecological relationships, as depicted by structural equation models that are connected with expert judgment values. Furthermore, Van Houtven et al. (2014) proposed a new protocol for linking ecological processes with public preferences, encompassing water-quality modelling, opinion elicitation, and contingent valuation (i.e., WTP estimation). Although their modelling framework accommodated water quality from watersheds, receiving waterbodies, such as lakes, were not included. Given that the management and policy depend upon waterbody (despite our recognition of high correlation between watershed and receiving waterbody), our study is conceptually on par with these efforts to accommodate multi-facets of ecosystem evaluation. In fact, our study expands on the work of past elicitation studies by offering an enhanced framework that accounts for the presence of uncertainty stemming from scientific knowledge, judgment divergence, and natural variability associated with an AOC (in this case, the Bay of Quinte). As a result, our study presents an integrated modelling framework that can be used to to establish optimal delisting criteria of AOCs.

Given that elicitation is conducted on the basis of qualitative judgments, the preparation of questionnaires and the training of interviewees can be critical in accurately determining the respondents' judgments. Several previous studies addressed important issues regarding the scope and format of expert elicitation. These issues include: (*i*) group versus individual and (*ii*) interview versus survey. Clemen and Winkler (1999) stated that group interaction has the potential benefit of sharing knowledge from different perspectives within and among disciplines. However, individual interviews are strongly preferred because participants may feel more responsible for providing informed judgment to interviewers than to anonymous questionnaires. On the other hand, Knol et al. (2010) generally recommended interviews that may allow for more targeted questions and explanations. The same study, however, also highlighted several advantages of internet or postal questionnaire-based survey. For instance, the study highlighted that surveys are less expensive than interviews, and the content of surveys can be better standardized than the content of individual interviews. In addition, Knol et al. (2010) also stressed the drawback of group interaction, which could lead to the inappropriate dominance of *influential experts*. Based on this literature review, the current study opted for a two-stage expert-elicitation approach (comprising both one-on-one interviews and questionnaire-based surveys) in attempt to capitalize on the advantages of the two methods and record the uncensored viewpoints of individual experts.

Another controversial issue regarding expert elicitation is the adequacy of the sample size. Keith (1996) proposed that the minimum representative sample size should be several times larger than 15 experts, due to the difficulty in defining who is considered an expert. Keith (1996) cautioned specifically about inaccurately combined judgments in case that multiple judgments from each individual were too divergent or biased to the judgment from a more influential leader of a group. Our use of the judgments elicited from freshwater eutrophication modelling specialists could reduce divergence of expert judgments in comparison with those from stakeholders. Counter to this concern, Knol et al. (2010) asserted that there is no benefit to including additional experts beyond 12 and that there is no absolute guideline for the number of experts to be invited. Viewed from this point, our elicitation results from 11 individuals could conceivably introduce some bias in our effort to synthesize expert opinions. Nonetheless, we considered that the bias stemming from the small sample size ($N = 11$) could be reduced by using the experts' combined probabilistic assessment of a statistically rigorous dataset with 100 multivariate water-quality observations. Our results corroborate this approach in the sense that the individual expert models were remarkably similar (see expert-specific regression coefficients provided in Table SI 1), even though the elicitation exercise was conducted independently with each interviewee, with no information sharing among the experts.

An interesting finding of our study is the discrepancy stemming from the collective views of the expert group relative to the public (largely laypeople), with regard to each group's awareness of the conditions currently experienced in the Bay of Quinte. Compared to a recent public survey (see Figs. 4 and 8 in Ramin et al., 2018), the stakeholder group was inclined to more favorable judgment (50% ≈ 10%

"high" + 40% "good" in Fig. 2a). We surmise that this difference may stem from the experts' knowledge of the significant progress made over the past four decades, relative to the prevailing conditions in the 1970s. Having a better quantitative understanding of the historical improvements in all water-quality variables could be a factor to better appreciate the present status of the Bay of Quinte. At the same time, however, the stakeholders displayed some ambivalence in characterizing the delisting likelihood of the system, a response that has also been documented elsewhere (Hovardas and Poirazidis, 2007). In particular, counter to the results of the stakeholders' interviews (Fig. 2), the expert elicitation models seemingly provided more conservative predictions for the likelihood of BUI #8 removal (< 50%), even when the two delisting targets (TP < 30 μg L^{-1} and Chl*a* < 10–12 μg L^{-1}) were met. Similar to our previous assertion, we reason that the experts' conservatism relative to public opinion reflects the deeper knowledge of the former group about the intricate ecological interplay shaping eutrophication severity in the system (e.g., harmful algal blooms, internal loading; see the following discussion). Concerning our study's demonstration of divergent responses, when comparing those of the public to those of the experts, there is ample evidence that the level of public awareness is a critical factor in explaining differences in opinion or judgment. For example, Rogers (2013) offered empirical evidence that higher public awareness caused public opinions to converge with the experts' values. Therefore, to facilitate a consensus in the environmental policy-making process, federal and provincial government, municipal authorities, stakeholders, and related private-sector companies should strive to enhance public awareness and to educate the local community with respect to the outstanding challenges and existing "*ecological unknowns*" (sensu lato Gudimov et al., 2011).

Despite the significant progress made in reducing both point and non-point loading discharges in the Bay of Quinte over the past four decades, recent modelling work has highlighted internal recycling as one of the key drivers of phosphorus dynamics and therefore seems to lend credence to experts' conservatism (Arhonditsis et al., 2016; Kim et al., 2013). The sediments in the upper-middle bay area act as a net source of phosphorus, and the macrophyte and dreissenid activities likely amplify the corresponding fluxes. The presence of this significant positive feedback loop acts as a major challenge in current management efforts to further improve water-quality conditions. Active nutrient-regeneration mechanisms present in the system suggests that the emergence of hysteresis patterns may compromise the efficiency of on-going external-loading reduction efforts, and thereby also significantly delay the emergence of an improved state of the waterbody (Scheffer et al., 2001; Minns et al., 2004; Gudimov et al., 2011). In fact, in a recent analysis of nutrient-loading scenarios, Arhonditsis et al. (2016) showed that the restoration pace of the bay could be slow even if the riverine TP concentrations reach levels significantly lower than their present values, < 25 μg TP L^{-1} (see Fig. 6 in Arhonditsis et al., 2016). Coupled with the conservative predictions of our expert elicitation model, it is not surprising that our integrative modelling study suggests delisting-likelihood values < 45% throughout the plausible range of hydraulic regimes and external nutrient loading levels examined.

Consistent with the skepticism of our current expert elicitation exercise, Kim et al. (2013) argued that the presence of an active feedback loop (internal nutrient recycling) in the system makes it compelling to avoid overly confident statements about the future response of this impaired system. As a result, the most prudent strategy is to recognize explicitly an acceptable level of violation of the delisting goals. Specifically, the historical delisting criterion of a seasonal average TP concentration lower than 30 μg L^{-1} was challenged, since it neither reflects the considerable intra-annual variability in the upper bay, nor does it represent the water-quality conditions in nearshore areas of high public exposure (e.g., beaches). It seems very unlikely that a single-valued water quality standard, monitored in a few offshore sampling stations, can capture the entire range of dynamics in the system (e.g., the extremes seen in the nearshore sites) or the magnitude of the end-of-summer TP peaks, predominantly mediated by internal regeneration mechanisms (Kim et al., 2013). In a follow-up study, Arhonditsis et al. (2016) advocated the pragmatic stance that the delisting objectives should revolve around extreme (and not average) values of variables of management interest and must explicitly accommodate all the sources of uncertainty by permitting a realistic frequency of standard violations. Namely, the critical threshold level should be set at a value of 40 μg TP L^{-1}, which cannot be exceeded > 10–15% in both time and space. Under the assumption that the TP concentrations in the Bay of Quinte follow a log-normal distribution and that TP values < 15 μg L^{-1} are likely to occur only 10% of the time during the growing season, then 10–15% exceedances of the 40 μg TP L^{-1} level are approximately equivalent to a targeted seasonal average of 25–28 μg TP L^{-1}. This still meets the historical delisting criterion and therefore suggests that the adoption of a pragmatic/probabilistic approach to water-quality criteria does not intend to make the delisting decision easier (and so, does not intend to compromise essential water-quality goals for the bay). Instead, it offers a more comprehensive method for assessing the prevailing conditions in the bay.

Bearing in mind that phosphorus control merely represents a "means to an end" and not "the end itself," in terms of the BUI "*Eutrophication or Undesirable Algae*", the critical question that our expert-interviewees appear to ponder is to what extent the reduced ambient TP levels could also trigger a significant decrease in algal-bloom frequency? In this regard, Nicholls et al. (2002) showed that the total phytoplankton biovolume declined in the bay after the reduction of point-source phosphorus in the 1970s, but did not change significantly after the establishment of dreissenids in the system since then. The arrival of dreissenid mussels may also be associated with both desirable (e.g., *Aphanizomenon* and *Anabaena* decline) and undesirable (e.g., *Microcystis* increase) changes in the integrity of the Bay of Quinte ecosystem (Shimoda et al., 2016). The recent occurrence of *Microcystis* blooms has significant implications for the aesthetics and other beneficial uses in the bay; these blooms form "scum" on the water surface, and some strains of *Microcystis* are toxin producers. Given their importance for overall ecosystem integrity, the frequency of extreme states (i.e., total phytoplankton biomass exceeding the maximum allowable levels or the occurrence of toxic algal blooms) must be an integral factor of the system delisting process. Similar to the skepticism associated with the TP criterion, we believe that the existing delisting target of a seasonal average of 10–12 μg Chl*a* L^{-1} cannot effectively project the likelihood of these types of extreme states. Recognizing that it is practically impossible to eliminate the occurrence of extreme events in the foreseeable future, the comprehensive investigation of the underlying drivers of algal blooms and the effective communication of the actual trends and on-going risks to the end-users (stakeholders and public) should be two essential steps of the local water-management efforts.

5. Concluding thoughts

Delisting decisions of impaired waterbodies resemble statistical null hypothesis testing, in which the prevailing conditions in the system at hand are compared against a reference (or non-impaired) state. This binary comparison stipulates the existence of a robust delineation of what constitutes a non-impaired state, along with establishing threshold values for key environmental variables, which act as surrogates for the degree of system impairment. In a system such as the Bay of Quinte, in which the severity of eutrophication phenomena is driven by both external and internal factors, there will inevitably be some uncertainty in the overall assessment of its water-quality status. For example, establishing predictive linkages between water-quality status and toxin concentration is particularly challenging, because the mechanisms associated with toxin production are complex, and not all *Microcystis* species produce toxins (Vanderploeg et al., 2001). In view of this scientific uncertainty, establishing the dichotomy between impaired and non-impaired conditions can be a very challenging exercise and, therefore, the input of public and stakeholders can be critical in order to weigh the trade-offs between environmental priorities and socioeconomic values.

In the Great Lakes region, the growing appreciation of the complexity of eutrophication control and the need to address the combined effects of a suite of tightly intertwined stressors has triggered the replacement of the Water Quality/Fisheries Exploitation paradigms with the Ecosystem Management paradigm (Krantzberg, 2012; Minns and Kelso, 2000). This paradigm shift has been perceived as a thrust toward adopting a multi-causal way of thinking in order to accommodate the complexity of ecosystems and maintain or resurrect their physical, chemical or biological integrity. In essence, the 14 beneficial uses stipulate the first dimension of the "*Ecosystem Approach*," reflecting the notion that ecosystem services should be viewed as a continuum, in which the remedial costs are increasing with the severity of ecosystem degradation; e.g., lower value of ecosystem services offered (Fig. 4). The second dimension of Great Lakes management is the stakeholder satisfaction (or perception), as captured by the variability among the different AOCs with respect to their levels of degradation; e.g., degree or type of contamination issues, number of BUIs, agricultural versus urban impacts, as well as their independence in determining their socioeconomic priorities or environmental stewardship practices. For each pair of the two-dimensional ecosystem integrity-human perception (or BUI-AOC) space, the third dimension is the uncertainty of goal achievability associated with (*i*) the important drivers of ecological degradation; (*ii*) the potential ramifications of the different sources of controversy, and (*iii*) our desire to have contingency plans to deal with the unexpected. In the context of integrative management, assessing this uncertainty is a core research topic, and our proposed framework is specifically designed to include the wide array of socio-ecological factors involved in the policy making process and to promote multi-level engagement as a hedge against forecasting errors.

Acknowledgments

This project was undertaken with financial support from Environment and Climate Change Canada. Additional funding for Dr. Maryam Ramin was provided by a MITACS Elevate Postdoctoral Fellowship. The authors are grateful to Weitao Zhang (City of Toronto), Peter Doris (Ministry of Agriculture, Food and Rural Affairs), Bryon Keene, Christine McClure (Quinte Conservation), Andrew Morley (Ontario Ministry of the Environment and Climate Change), Kristin Geater, Rimi Kalinauskas, Sue Watson, Ram Yerubandi, Veronique Hiriart-Baer, and Agnes Blukacz-Richards (Environment and Climate Change Canada), Kenneth Minns, Marten Koops, Sue Doka, Scott Millard, Mark Fitzpatrick, and Mohiuddin Munawar (Fisheries and Oceans Canada) for providing invaluable feedback throughout this project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2018.08.005.

References

Arhonditsis, G.B., Kim, D.-K., Shimoda, Y., Zhang, W., Watson, S., Mugalingam, S., Dittrich, M., Geater, K., McClure, C., Keene, B., Morley, A., Richards, A., Long, T., Rao, Y.R., Kalinauskas, R., 2016. Integration of best management practices in the bay of Quinte watershed with the phosphorus dynamics in the receiving water body: what do the models predict? Aquat. Ecosyst. Health Manag. 19, 1–18.
Bay of Quinte Remedial Action Plan (BQRAP), 1987. Bay of Quinte Remedial Action Plan.
Bay of Quinte Remedial Action Plan (BQRAP), 2015. Project Quinte Annual Report 2013, Bay of Quinte Remedial Action Plan Monitoring Report #24. Bay of Quinte RAP Restoration Council, Kingston, ON, Canada, pp. 84.
Borsuk, M.E., 2004. Predictive assessment of fish health and fish kills in the Neuse River Estuary using elicited expert judgment. Hum. Ecol. Risk. Assess. 10, 415–434.
Borsuk, M., Clemen, R., Maguire, L., Reckhow, K., 2001. Stakeholder values and scientific modeling in the Neuse River watershed. Group Decis. Negot. 10, 355–373.
Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. 7, 434–455.
Burgman, M., 2005. Risks and Decisions for Conservation and Environmental Management. Cambridge University Press, New York.
Carson, R.T., Mitchell, R.C., 1993. The value of clean water: the public's willingness to pay for boatable, fishable, and swimmable quality water. Water Resour. Res. 29, 2445–2454.
Chen, K., MacKay, S., 2004. A Summary of Delisting Criteria for Canadian Areas of Concern. Environment Canada, pp. 25.
Clemen, R.T., Reilly, T., 1999. Correlations and copulas for decision and risk analysis. Manag. Sci. 45, 208–224.
Clemen, R.T., Winkler, R.L., 1999. Combining probability distributions from experts in risk analysis. Risk Anal. 19, 187–203.
Cooke, R.M., 1991. Experts in Uncertainty: Opinion and Subjective Probability in Science. Oxford University Press, New York.
Cooper, P., Poe, G.L., Bateman, I.J., 2004. The structure of motivation for contingent values: a case study of lake water quality improvement. Ecol. Econ. 50, 69–82.
Del Saz-Salazar, S., Hernández-Sancho, F., Sala-Garrido, R., 2009. The social benefits of restoring water quality in the context of the water framework directive: a comparison of willingness to pay and willingness to accept. Sci. Total Environ. 407, 4574–4583.
Doria, M.D.F., Boyd, E., Tompkins, E.L., Adger, W.N., 2009. Using expert elicitation to define successful adaptation to climate change. Environ. Sci. Pol. 12, 810–819.
Ellison, A.M., 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecol. Appl. 6, 1036–1046.
Environment Canada (EC), Ontario Ministry of the Environment (OMOE), 2007. Canada-Ontario Agreement: Respecting the Great Lakes Basin Ecosystem.
Environment Canada (EC), United States Environment Protection Agency (USEPA), 2013. Great Lakes Water Quality Agreement. pp. 56.
Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York.
George, T.K., Boyd, D., 2007. Limitations on the development of quantitative monitoring plans to rack the progress of Beneficial Use Impairment restoration at Great Lakes Areas of Concern. J. Great Lakes Res. 33, 686–692.
Gilks, W.R., Roberts, G.O., Sahu, S.K., 1998. Adaptive Markov chain Monte Carlo through regeneration. J. Am. Stat. Assoc. 93, 1045–1054.
Gudimov, A., Ramin, M., Labencki, T., Wellen, C., Shelar, M., Shimoda, Y., Boyd, D., Arhonditsis, G.B., 2011. Predicting the response of Hamilton Harbour to the nutrient loading reductions: a modeling analysis of the "ecological unknowns". J. Great Lakes Res. 37, 494–506.
Hall, J.D., O'Connor, K., Ranieri, J., 2006. Progress toward delisting a Great Lakes area of concern: the role of integrated research and monitoring in the Hamilton harbour remedial action plan. Environ. Monit. Assess. 113, 227–243.
Hovardas, T., Poirazidis, K., 2007. Environmental policy beliefs of stakeholders in protected area management. Environ. Manag. 39, 515–525.
International Joint Commission (IJC), 1985. Report on Great Lakes Water Quality. Kingston, Ontario, Canada.
International Joint Commission (IJC), 2003. Status of Restoration Activities in Great Lakes Areas of Concern: A Special Report.
Keith, D.W., 1996. When is it appropriate to combine expert judgments? Clim. Chang. 33, 139–143.
Kenney, M.A., Arhonditsis, G.B., Reiter, L.C., Barkley, M., Reckhow, K.H., 2009. Using structural equation modeling and expert elicitation to select nutrient criteria variables for south-central Florida lakes. Lake Reserv. Manage. 25, 119–130.
Kim, D.-K., Zhang, W., Rao, Y.R., Watson, S., Mugalingam, S., Labencki, T., Dittrich, M., Morley, A., Arhonditsis, G.B., 2013. Improving the representation of internal nutrient recycling with phosphorus mass balance models: a case study in the Bay of Quinte, Ontario. Canada. Ecol. Model. 256, 53–68.
Kim, D.-K., Kaluskar, S., Mugalingam, S., Blukacz-Richards, A., Long, T., Morley, A., Arhonditsis, G.B., 2017. A Bayesian approach for estimating phosphorus export and delivery rates with the SPAtially Referenced Regression On Watershed attributes (SPARROW) model. Ecol. Inform. 37, 77–91.
Knol, A.B., Slottje, P., van der Sluijs, J.P., Lebret, E., 2010. The use of expert elicitation in environmental health impact assessment: a seven step procedure. Environ. Health 9, 19.
Krantzberg, G., 2012. The remedial action plan program, historical and contemporary overview. In: Grover, V.I., Krantzberg, G. (Eds.), Great Lakes: Lessons in Participatory Governance. CRC Press, Boca Raton, FL, pp. 245–256.
Kuhnert, P.M., Martin, T.G., Griffiths, S.P., 2010. A guide to eliciting and using expert knowledge in Bayesian ecological models. Ecol. Lett. 13, 900–914.
Lerner, D.N., Kumar, V., Holzkämper, A., Surridge, B.W.J., Harris, B., 2011. Challenges in developing an integrated catchment management model. Water Environ. J. 25, 345–354.
Lin, S.W., Bier, V.M., 2008. A study of expert overconfidence. Reliab. Eng. Syst. Saf. 93, 711–721.
MacMillan, D.C., Marshall, K., 2006. The Delphi process – an expert-based approach to ecological modelling in data-poor environments. Anim. Conserv. 9, 11–19.
Martin, T.G., Burgman, M.A., Fidler, F., Kuhnert, P.M., Low-Choy, S., McBride, M., Mengersen, K., 2012. Eliciting expert knowledge in conservation science. Conserv. Biol. 26, 29–38.
McCarthy, M.A., Masters, P.I.P., 2005. Profiting from prior information in Bayesian analyses of ecological data. J. Appl. Ecol. 42, 1012–1019.
Meyer, M.A., Booker, J.M., 1991. Eliciting and Analyzing Expert Judgement: A Practical Guide. Academic Press Ltd., London.
Michigan Department of Environmental Quality (MDEQ), 2008. Guidance for Delisting

Michigan's Great Lakes Areas of Concern. Report OGL-002. Lansing, MI, pp. 66.
Minns, C.K., Kelso, J.R., 2000. NO! It is time for a Great Lakes ecosystem management agreement that SUBSUMES the Great Lakes water quality agreement. J. Great Lakes Res. 26, 1–2.
Minns, C.K., Moore, J.E., 2004. Modelling Phosphorus Management in the Bay of Quinte, Lake Ontario in the Past, 1972 to 2001, and in the Future, Canadian Manuscript Report of Fisheries and Aquatic Sciences No. 2695. Great Lakes Laboratory for Fisheries and Aquatic Sciences, Bayfield Institute, Burlington, Ontario, pp. 42.
Minns, C.K., Moore, J.E., Seifried, K.E., 2004. Nutrient Loads and Budgets in the Bay of Quinte, Lake Ontario, 1972 to 2001, Canadian Manuscript Report of Fisheries and Aquatic Sciences 2694. Great Lakes Laboratory for Fisheries and Aquatic Sciences, Bayfield Institute, Burlington, Ontario, pp. 49.
Morgan, M.G., Henrion, M., Small, M., 1992. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, New York.
Mullen, B., 1991. Group composition, salience, and cognitive representations: the phenomenology of being in a group. J. Exp. Soc. Psychol. 27, 297–323.
Nicholls, K.H., Heintsch, L., Carney, E., 2002. Univariate step-trend and multivariate assessments of the apparent effects of P loading reductions and zebra mussels on the phytoplankton of the Bay of Quinte, Lake Ontario. J. Great Lakes Res. 28, 15–31.
Ramin, M., Cheng, V.Y.S., Kim, D.-K., Midlane-Jones, S., Mugalingam, S., Arhonditsis, G.B., 2018. A Bayesian methodological framework for coupling public perception with the water quality criteria setting process. Ecol. Econ. 147, 298–311.
Reckhow, K.H., Arhonditsis, G.B., Kenney, M.A., Hauser, L., Tribo, J., Wu, C., Elcock, K.J., Steinberg, L.J., Stow, C.A., McBride, S.J., 2005. A predictive approach to nutrient criteria. Environ. Sci. Technol. 39, 2913–2919.
Regan, H.M., Colyvan, M., Burgman, M.A., 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecol. Appl. 12, 618–628.
Rogers, A.A., 2013. Public and expert preference divergence: evidence from a choice experiment of marine reserves in Australia. Land Econ. 89, 346–370.
Scheffer, M., Straile, D., van Nes, E.H., Hosper, H., 2001. Climatic warming causes regime shifts in lake food webs. Limnol. Oceanogr. 46, 1780–1783.
Shimoda, Y., Watson, S., Palmer, M.E., Koops, M., Mugalingam, S., Morley, A., Arhonditsis, G.B., 2016. Delineation of the role of nutrient variability and dreissenids (Mollusca, Bivalvia) on phytoplankton dynamics in the bay of Quinte, Ontario, Canada. Harmful Algae 55, 121–136.
Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS User Manual, Version 1. 4, Available at http://www.mrc-bsu.cam.ac.uk/bugs.
United States Policy Committee (USPC), 2001. Restoring United States Areas of Concern: Delisting Principles and Guidlines. pp. 29.
Van Houtven, G., Mansfield, C., Phaneuf, D.J., von Haefen, R., Milstead, B., Kenney, M.A., Reckhow, K.H., 2014. Combining expert elicitation and stated preference methods to value ecosystem services from improved lake water quality. Ecol. Econ. 99, 40–52.
Vanderploeg, H.A., Liebig, J.R., Carmichael, W.W., Agy, M.A., Johengen, T.H., Fahnenstiel, G.L., Nalepa, T.F., 2001. Zebra mussel (*Dreissena polymorpha*) selective filtration promoted toxic *Microcystis* blooms in Saginaw Bay (Lake Huron) and Lake Erie. Can. J. Fish. Aquat. Sci. 58, 1208–1221.
Zhang, W., Kim, D.-K., Rao, Y.R., Watson, S., Mugalingam, S., Labencki, T., Dittrich, M., Morley, A., Arhonditsis, G.B., 2013. Can simple phosphorus mass balance models guide management decision?: a case study in the Bay of Quinte, Ontario, Canada. Ecol. Model. 257, 66–79.

AN INTEGRATIVE METHODOLOGICAL FRAMEWORK FOR SETTING ENVIRONMENTAL CRITERIA: EVALUATION OF STAKEHOLDER PERCEPTIONS

[SUPPORTING INFORMATION]

Dong-Kyun Kim[1], Maryam Ramin[1], Vincent Cheng[1], Aisha Javed[1], Samarth Kaluskar[1], Noreen Kelly[1], Demetrios Kobiliris[1], Alex Neumann[1], Felicity Ni[1], Tianna Peller[1], Gurbir Perhar[1], Yuko Shimoda[1], Ariola Visha[1], Christopher Wellen[1], Cindy Yang[1], Shan Mugalingam[2], George B. Arhonditsis[1*]

[1]Ecological Modelling Laboratory,
Department of Physical & Environmental Sciences, University of Toronto,
Toronto, Ontario, Canada, M1C1A4

[2]Lower Trent Conservation,
Trenton, Ontario, Canada, K8V 5P4

* Corresponding author
E-mail: georgea@utsc.utoronto.ca, Tel.: +1 416 208 4858; Fax: +1 416 287 7279.

Section A – Figures and Table

"***How many times (out of 100) will the Bay of Quinte be in attainment of its Beneficial Uses, when the summer average levels of the five water quality variables (Total Phosphorus, Chlorophyll a, Ammonium, Secchi Disk Depth, Microcystin) are those reported in each row?***"

Variables	*Chla*	*NH_4*	*TP*	*Secchi Disk depth*	*Microcystin*	
Unit	µg/L	µg/L	µg/L	m	µg/L	(Out of 100)
Present Condition	12.5	25.5	31.5	1.75	0.75	
1	14	46	33	1.80	0.18	3
2	16	11	25	1.94	0.39	5
3	11	35	36	1.82	0.59	10
4	15	20	34	1.60	0.19	5
5	14	13	21	1.74	0.84	20
6	11	41	31	1.73	0.50	20
7	17	23	27	1.49	1.60	20
8	12	5	26	1.44	0.27	30
9	10	25	29	1.95	0.54	40
10	16	42	24	2.07	0.13	30
11	10	54	33	1.52	0.62	20
12	9	18	25	1.55	2.20	20
13	7	46	37	1.90	0.30	15
14	13	24	25	1.75	0.22	15
15	14	12	35	1.87	1.00	15

Figure SI-1: Sample of an expert's responses on the probability of achieving acceptable water quality conditions based on five lake trophic indicators: Total Phosphorus (*TP*), Chlorophyll *a* (Chl *a*), Microcystin concentrations, Secchi Disk Depth, and Ammonium concentration (*NH_4*).

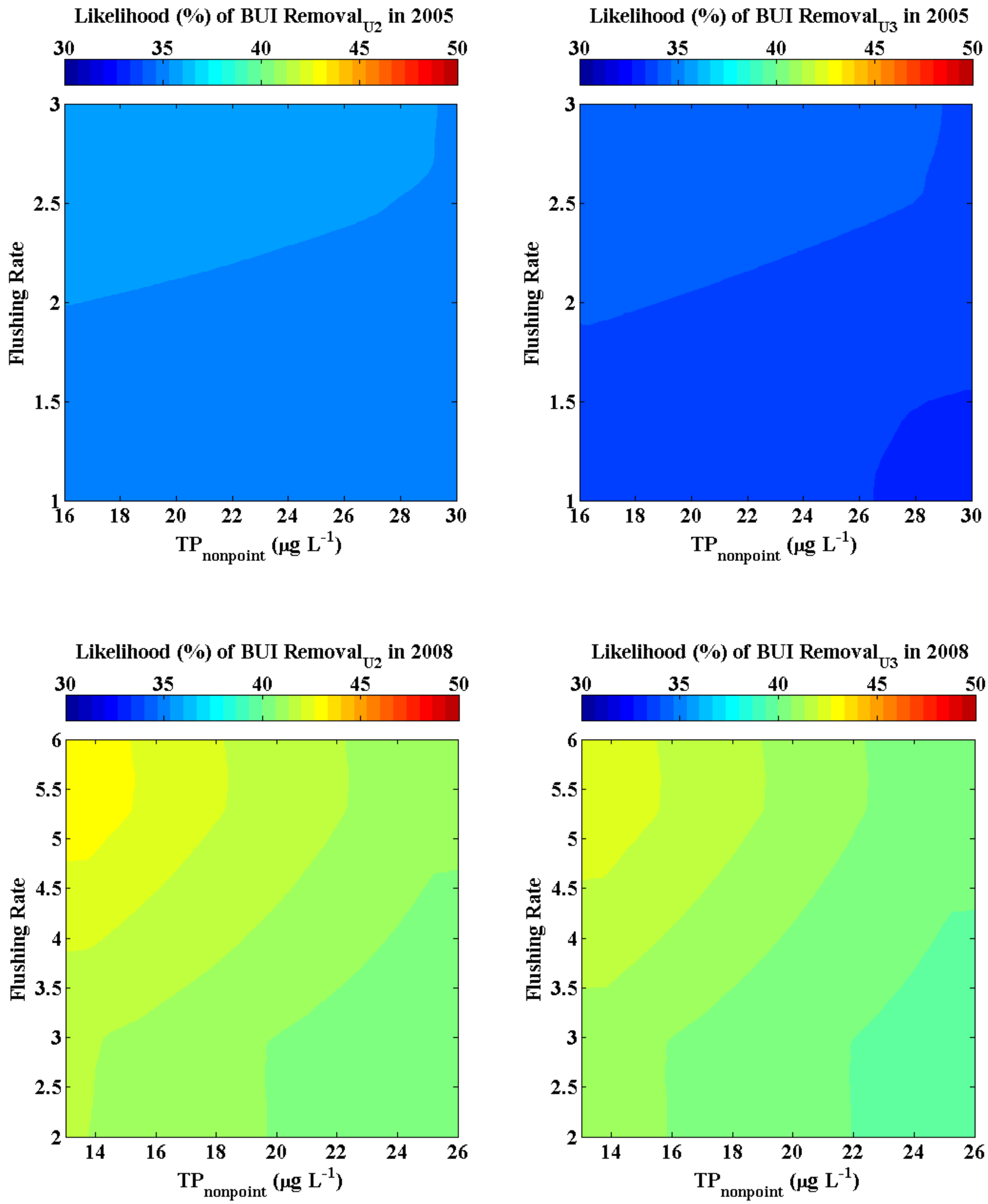


Figure SI-2: Likelihood (%) of delisting the upper Bay of Quinte for the BUI #8 "*Eutrophication or Undesirable Algae*". The predicted *TP* concentrations of the Bay in response to inflow *TP* concentrations and flushing rates (the number of times the system flushes during the growing season between May and October) at two locations of the upper Bay in 2005 (upper panels) and 2008 (lower panels), are linked with the expert elicitation model.

Table SI-1: Posterior parameter estimates of the two expert-elicitation models

	TP				*Chl a*			
Parameter	Mean	S.D.	2.5%	97.5%	Mean	S.D.	2.5%	97.5%
α	119.9	7.5	101.3	134.3	60.0	5.5	49.3	70.6
α_1	113.3	7.8	96.7	127.9	58.8	5.5	48.4	70.1
α_2	79.8	13.5	55.0	104.9	35.4	11.8	11.2	53.5
α_3	83.4	11.7	61.8	107.0	89.2	7.0	72.6	101.6
α_4	124.4	8.8	111.8	148.0	118.3	10.3	98.4	138.7
α_5	48.0	10.5	27.2	68.3	57.2	7.4	38.7	69.6
α_6	96.6	11.6	70.4	114.6	53.8	6.4	41.2	66.3
α_7	124.1	7.7	107.2	138.6	70.7	7.3	56.0	84.2
α_8	121.2	6.6	108.5	135.5	61.4	5.3	50.1	71.8
α_9	137.1	9.8	117.5	152.7	89.9	6.3	77.3	102.9
α_{10}	120.2	6.5	106.3	134.4	62.6	5.3	52.2	73.7
α_{11}	120.1	8.6	101.7	137.2	60.1	7.1	47.4	74.7
β	-24.1	1.8	-28.07	-20.0	-8.8	1.9	-12.9	-4.9
β_1	-24.7	2.3	-29.21	-19.6	-10.3	2.2	-14.8	-6.0
β_2	-16.9	4.1	-24.51	-9.2	-3.9	4.7	-10.9	5.9
β_3	-4.3	3.6	-11.62	2.3	-8.2	2.7	-13.1	-1.4
β_4	-25.5	2.7	-32.94	-21.6	-31.3	4.2	-39.5	-23.1
β_5	4.3	3.2	-2.043	10.7	1.9	3.1	-3.3	9.2
β_6	-21.3	3.5	-26.73	-13.3	-10.3	2.5	-15.6	-5.3
β_7	-22.1	2.4	-26.55	-16.9	-7.2	2.9	-12.6	-1.2
β_8	-24.7	2.1	-29.09	-20.8	-8.0	2.1	-12.3	-3.3
β_9	-21.6	3.0	-26.44	-15.5	-9.1	2.5	-14.4	-4.2
β_{10}	-24.5	2.0	-28.78	-20.3	-8.7	2.1	-13.1	-4.5
β_{11}	-29.1	2.7	-34.47	-23.2	-13.6	3.0	-20.1	-8.2

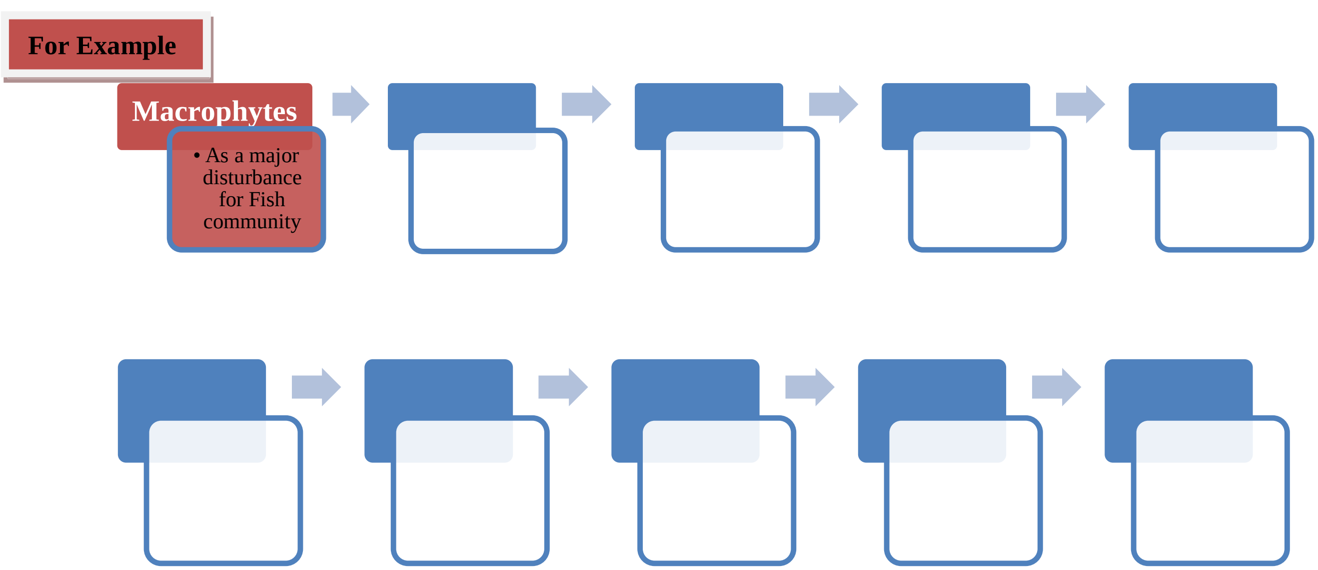
Section B - Questionnaire of the Expert Elicitation Survey

1. **Which of the following types of stakeholder involvement best describes the organization you represent in the BofQ?**
   (a) Local government
   (b) Provincial government
   (c) Federal government
   (d) Industry
   (e) Non-profit
   (f) Other

2. **How do you define your level of expertise and familiarity with the BofQ?**
   Scale of 1 (familiar with little knowledge) – 10 (expert with high knowledge).

3. **What are the mechanisms that lead to eutrophication in the BofQ?**
   Use the keywords below and put them in order from the most important to the least important:

   <u>Keywords:</u>
   Nitrogen, Phosphorus, Chlorophyll a, Total Suspended Solids, Turbidity, Secchi Disk Depth, Dissolved oxygen, pH, Temperature, Macrophytes, Dreissenid mussels, Benthic organism, Toxic algae, Sport fish population, Fecal coliform, Land use, Fish biodiversity.
   (Add comments if any to the box below).



4. **How would you characterize the Bay of Quinte with respect to the following eutrophication indicators? Please choose one word from the keywords provided below.**

Water clarity	Algae	Nutrient levels	Oxygen	Odour	Aquatic life

<u>Keywords:</u>
Excellent, Good, Very little, Very low, Very high, Little, Low, High, Abundant, Fair, Moderate, Poor, Noticeable, Scarce.

5. **Which Beneficial Use Impairment is still an important issue and requires more attention in the Bay of Quinte within the next 5 years?**
(a) Restriction on fish and wildlife consumption, (b) Eutrophication or undesirable algae, (c) Degradation of aesthetics, (d) Degradation of Phytoplankton and Zooplankton Populations, (e) Restrictions on drinking water or taste and odour problems, (f) Beach closures, (g) Loss of fish and wildlife habitat, (h) Others (Please specify).

6. **How would you define the Beneficial Use that you have selected? If multiple variables provided for attainment of beneficial use, what would be the single best variable and why?**

7. **Given the variable that you just identified as ideal to measure the aspects of the beneficial use, what do you believe is the attainment vs. non-attainment change point level for this variable?**

8. **How do you understand the broader roles of TP and Macrophytes in the Bay of Quinte?**

9. **Do you think the Bay of Quinte Remedial Action Plan was successful in achieving its goals? Please elaborate.**

10. What changes have you noticed within the last 5 years at the Bay of Quinte?

(a) Clarity better
(b) Clarity worst
(c) Smell better
(d) Smell worst
(e) More fish catch
(f) Less fish catch
(g) More weeds
(h) Less weeds
(i) More algae
(j) Less algae
(k) Other (Specify)

11. What do you think are the biggest problems from the public using the Bay of Quinte at the time being?

(a) Water Quality
(b) Macrophytes
(c) Toxins
(d) Other (Specify)

12. What is the optimal management strategy to restore and maintain the integrity of the Bay of Quinte?

13. Please identify the barriers to implement the optimal management practices at the Bay of Quinte.

<u>Examples</u>:
Time Constraint, Lack of Resources, Lack of Knowledge, Cost, Unknown Ecological Ramifications, ...

14. Other comments: