UNIVERSITÉ DU QUÉBEC À MONTRÉAL

QUANTITATIVE FRAMEWORKS TO IMPROVE THE ANALYSES OF ECOLOGICAL COMMUNITIES IN THE CONTEXT OF AQUATIC ECOSYSTEM SERVICES

THESIS

PRESENTED

AS A PARTIAL REQUIREMENT

FOR THE DOCTORATE IN BIOLOGY

BY

WAGNER SANDRO DA COSTA MOREIRA

JANUARY 2019

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

CADRES QUANTITATIFS POUR AMÉLIORER LES ANALYSES DES COMMUNAUTÉS ÉCOLOGIQUES DANS LE CONTEXTE DES SERVICES D'ÉCOSYSTÈMES AQUATIQUES

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN BIOLOGIE

PAR

WAGNER SANDRO DA COSTA MOREIRA

JANVIER 2019

"If you can't fly then run, if you can't run then walk, if you can't walk then crawl, but whatever you do you have to keep moving forward"

-Martin Luther King, Jr.

ACKNOWLEDGEMENTS

First and foremost, I would like to express the deepest gratitude to my advisors Pedro Peres-Neto and Nigel Lester, for the opportunity given, all the knowledge shared and friendly support during this time.

Also, I would like to express my sincere appreciation to my past and present labmates, who received me with open arms when I first arrived in the group and continue to show their friendship: Bailey Jacobson, Renato Henriques da Silva, Mehdi Layeghifard, Frédéric Boivin, Andrew Smith, Jason Samson, Emily Tissier, Hedvig Nénzen, Henrique Giacomini, Ignacio Castilla-Morales, Louis Donelle, Marcia Marques, Pedro Henrique Pereira Braga, Sylvie Clappe, Viviane Monteiro, Bertrand Fournier and Hector Velazquez. Huge thanks as well to the friends Montréal gave me, who brought some joy during these years: Eric Côté, Sawssan Kaddoura, Emanuel Araújo, Julia Stringhetta, Emanuele Barley, Stéphane Boivin; and to the friends who live physically far, but close in my heart: Vincenzo Maiorana, Eric McCormick, Rafael José, Patrícia Mesquita and Thiago Silva.

Last but not least, a heartfelt thank you to my parents, Waldemir and Edilma Moreira, who always supported my decisions, even if they meant to live far from them. These were difficult years to have them so far away, specially under some delicate situations, but optimism wins them all.

This thesis was funded by both a FQRNT and NSERC research grants awarded to me and P.R. Peres-Neto, which were fundamental during these doctoral years.

“Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.” – Winston Churchill

To my family, for the
unconditional love and for
accepting who I am.

To my Montréal friends, who
were given to me as a precious
gift and helped me endure these
stormy, yet beautiful years.

TABLE OF CONTENTS

CHAPTER II

CHAPTER III

LIST OF FIGURES

LIST OF TABLES

Table Page

RESUMÉ

L'analyse de la relation espèces-habitat a toujours été un des objectifs centraux de l'écologie. C'est devenu un cadre pour explorer, comprendre et répondre à des questions spécifiques concernant la complexité et les mécanismes qui sous-tendent les patrons spatio-temporels de distribution d'espèces. Le savoir-faire généré par la modélisation écologique et la quantification des relations espèces-environnement est un élément clé pour la conservation et la gestion des écosystèmes et des populations. L'objectif général de cette thèse est d'améliorer les méthodes quantitatives en : (i) construisant une approche pour les donnée multi-espèces basée sur les modèles linéaires généralisés, appelée "Analyse de Redondance Généralisée" (gRDA) ; (ii) améliorant la modélisation de la distribution des poissons (présence/absence) en considérant les interactions biotiques entre espèces, et en utilisant les connaissances générées par ces modèles pour (iii) modéliser la biomasse des poissons, déterminer les principaux facteurs environnementaux qui influencent la biomasse des différentes espèces, ainsi qu'évaluer l'impact de la pression exercée par le pêche. Ces liens peuvent contribuer à renforcer la recherche sur la relation espèces-habitat, les services écosystémiques tout en apportant des informations sur les processus qui sous-tendent ces relations. La recherche sur des méthodes quantitatives interconnectées devrait permettre de mieux comprendre les agents qui structurent la biodiversité et comment ils interagissent pour fournir des services écosystémiques, tout en clarifiant les actions qui devraient être entreprises pour remédier à la perte de ces services liées à l'augmentation de l'impact de l'homme. Dans le Chapitre 1, une méthode multi-espèces appelée "Analyse de Redondance Généralisée" (gRDA) a été développée. Cette méthode est basée sur la distribution logistique et la distribution de Poisson, et elle a été étendue au Partitionnement de Variation. Le but du chapitre est de présenter cette méthode et de déterminer ces performances en utilisant une variété de méthodes de Monte Carlo. Nos résultats montrent que la méthode proposée est robuste et devrait remplacer le Partitionnement de Variation standard basé sur l'Analyse de Redondance ordinaire. Le chapitre II présente une comparaison de la performance et de la capacité prédictive de plusieurs méthodes de modélisation de la distribution des espèces, tout en identifiant les prédicteurs les plus importants pour expliquer la présence ou l'absence des espèces considérées. Ces modèles diffèrent par l'utilisation de données empiriques et ont été développés pour six espèces de poisson de l'eau douce dans les lacs d'Ontario. La modélisation dans le chapitre 2 prend en compte trois classes de modèles : (i) les modèles qui n'utilisent que l'information sur les conditions environnementales comme prédicteurs : (ii) ceux qui utilisent uniquement l'information sur les communautés de poissons pour prédire la présence d'une espèce particulière (c.à.d. variables abiotiques) et (iii) une combinaison de (i) et (ii) (c.à.d. variables abiotiques et biotiques). Finalement, en utilisant les mêmes espèces de poissons que pour le Chapitre II, le chapitre III présente des modèles de biomasse développés au moyen d'arbres de régression permettant d'identifier l'importance relative de différentes variables environnementales ainsi que la pression angulaire. De plus, les similarités et différences entre guildes de poissons ont été mises en évidence par l'influence de chacun des prédicteurs sur la biomasse de chaque espèce. Les résultats des chapitres II et III révèlent l'importance de la morphométrie des lacs et du climat pour l'estimation de la présence/absence et de la biomasse, et, dans le chapitre II particulièrement, il a été observé que les modèles de présence/absence ont de meilleures performances lorsqu'ils tiennent en compte de l'information biotique ainsi que des prédicteurs abiotiques. De plus, les résultats indiquent que la performance des modèles pourrait être largement affectée par la manière dont ces modèles sont développés et évalués. Choisir une méthode de modélisation appropriée, les variables explicatives, les méthodes de validation et les mesures de performances sont des étapes importantes pour obtenir des inférences ou prédictions plus fiables plutôt que des résultats spécifiques aux données ou des artéfacts statistiques. De ce fait, la présente thèse fournit une enquête complète sur l'amélioration des méthodes de modélisation populaires, tout en montrant leurs utilités pour évaluer l'importance des facteurs influençant la distribution et la biomasse des poissons qui ont une importance économique en Ontario, ce qui peut être utile du point de vue de la gestion des ressources naturelles.

Mots-clés: modélisation écologique; Analyse de Redondance; Partitionnement de Variation; modélisation de la distribution des espèces; modèles de biomasse; Poisson d'Eau Douce.

SUMMARY

The analysis of species – habitat relationships has always been a central goal in ecology. It has become a central framework to explore, understand and tackle specific questions about the intricacies and mechanisms underlying species distributional patterns in space and time. The *savoir-faire* generated by ecological modelling and its quantification of species-environment relationships is critical for conservation planning and ecosystem/population management. The general objective of this thesis is to improve quantitative methods to: (i) build a framework for multispecies data based on generalized linear models, called "Generalized Redundancy Analysis" (gRDA); (ii) access improvement in species distribution modelling approaches developed for fish presence-absence by inserting biotic information related to the fish species, and knowledge about these models are linked to the (iii) development of fish biomass models while unravelling the principal environmental determinants that influence the biomass of different species, when also evaluating the impact of fishing pressure. These links can contribute to major scientific underpinnings related to the research of species-habitat relationships, while consisting of ecosystem services when promoting information about the processes underlying these relationships. The investigation of interconnected quantitative frameworks to link environmental, spatial and biotic interactions should bring to light a greater understanding of the key agents structuring biodiversity and how they interact to provide the delivery of aquatic ecosystem services, while clarifying about the actions that should be taken to mitigate the loss of these services in face of increasing human impacts. In Chapter I, a multispecies framework called "Generalized Redundancy Analysis" (gRDA) was developed, based on logistic and Poisson distributions via GLMs, and extend this framework to variation partitioning. The goal was to present the framework as well as assess its performance using Monte Carlo approaches under a variety of scenarios. Our results showed that the proposed framework is very robust and should essentially replace the current standard variation partitioning based on ordinary Redundancy Analysis. Chapter II presents a comparison among a number of SDM approaches in terms of their predictive performance and explanatory power, while identifying the predictors that are most influent in explaining presence-absence of the species considered. These models were contrasted by using empirical data and developed for six species of freshwater fishes in north-temperate lakes of Ontario. The modelling routine for Chapter 2 was done taking into account three classes of models (i) incorporating only information about the environmental features as predictors; (ii) using solely the information about the fish community to predict the occurrence of a particular species (i.e. using biotic parameters) and (iii) represents a combination of (i) and (ii) (i.e. using abiotic + biotic parameters); Finally, using the same fish species from Chapter II, Chapter III presents biomass models developed via boosted regression trees, allowing to identify the relative importance of the different principal environmental determinants used as regressors, together with angling pressure. In addition, similarities and dissimilarities within and among fish thermal guilds were showcased regarding the biomass response of each species to every predictor used in the modelling process. Results from Chapter II and III point out the importance of lake morphometry and climate to the estimation of presence-absence and biomass, and in Chapter II particularly it was observed model improvements when biotic information was used together with abiotic predictors for presence-absence models. In addition results indicated that model performance could be largely affected by how models were developed and evaluated. How to choose appropriate modelling approaches, predictor variables, model validation methods, and performance metrics are important steps if we want to get more reliable inferences or predictions rather than data-specifics results or statistical artifacts. Thus, this thesis provides a comprehensive investigation on aspects consisting of improvements of popular modelling approaches, while showing their usefulness in assessing important information related to factors most influent in the distribution and biomass of economically important fish species in Ontario, which can be useful from the natural resources management point of view.

Keywords: ecological modelling, Redundancy Analysis, Variation Partitioning, species distribution models, biomass models, freshwater fish

INTRODUCTION

In this section, I provide details on the background that supports this thesis. The first part refers to a discussion about the role and increasing importance of species distribution modelling in Ecology, showing the different steps involved in the calibration (building) process, the importance of each of these steps, followed by the importance of proper estimation, addressing the most common limitations faced when building static and probabilistic models. The second part outlines ecosystem services and the importance of different services provided by fish populations. I will address the importance of the development and assessment of models and quantitative tools to better predict and understand the ecosystem services provided by fish, focusing on how this information can improve monitoring and management programs for fish populations.

0.1 Species modelling

The analysis of species-habitat relationships has always been a central goal in ecology. It has become a central framework to explore, understand and tackle specific questions about the intricacies and mechanisms underlying species distributional patterns in space and time. The *savoir-faire* generated by ecological modelling and its quantification of species-environment relationships is critical for conservation planning and ecosystem/population management. The statistical frameworks applied in ecological modelling are generally based on estimating parameters about the importance of environmental features (e.g., local habitat, regional climate, habitat connectivity) influencing the distribution of species and their communities (Guisan and Zimmermann 2000). This knowledge is of utmost importance to estimate habitat suitability for endangered species, discover new populations or previously unknown species, forecast effects of habitat change due to human interference, establish potential locations for species reintroduction, predict how community structure may be affected by the invasion of exotic species, predict the effect of ecological disturbances, climate change or how environmental conditions affect different communities across different spatial/temporal scales. Habitat models relating habitat characteristics, and species distributions and community structure allows one to derive/predict the habitat potential distribution within the modelled area, which is equivalent to modelling its potential habitat (Schuster 1994) and niche (Elith and Leathwick 2009). A plethora of statistical approaches for species modelling are available (e.g. Harvey 1978; Somers and Harvey 1984; Legendre and Fortin 1989; Jackson and Harvey 1993; Jenkins and Buikema 1998; Guisan and Zimmermann 2000; Peres-Neto et al. 2006; Sharma and Jackson 2007; Elith et al. 2008 and many others) and they differ in their ability to model environmental relationships. As such, an evaluation of different statistical techniques can provide insights into which approaches are most appropriate for the biological question being asked at both the species and community levels (Guisan and Zimmermann 2000; Elith et al. 2006; Sharma and Jackson 2008; Sharma et al. 2012). It is important to keep in mind that habitat models are expected to address at least two questions: (1) how well the distribution of a set of species is explained given a set of covariates? (2) Which covariates are unimportant in the sense of contribution to the explanation of patterns already accounted for by other variables present in the model (i.e., marginal and independent contribution).

The development of environmental niche models involves some steps that are central while generating a consistent tool that works with similar levels of accuracy across large landscapes, particularly in those in which the nature of the covariation among predictors and their contribution to species distributions change spatially (e.g., non-stationarity; Wenger and Olden 2012). The first step is the conceptual model formulation, or underlying conceptual framework. In this step, one usually faces the task of deciding which model properties are desirable to achieve. Levins (1966) points out three main model properties, *generality*, *reality* and *precision*, stating that only any two out of three can be improved simultaneously. In this sense, three classes of models can be designed: (i) accurate prediction within a limited or simplified reality. In this category, *analytical* (Sharpe 1990) models are suitable and they incorporate precision and generality; (ii) predictions attained on real cause-effect interactions are named *mechanistic* (Prentice, 1986), designed to be realistic and general, with a focus on theoretical correctness of the predicted response over predicted precision; (iii) *Empirical* models (Decoursey 1992) are based on precision and reality, condensing empirical facts instead of considering realistic "cause and effect" between response and explanatory variables. Central aspects to consider when conceptualizing a predictive habitat distribution model are the inclusion of direct *versus* indirect (e.g., proxy) predictors, the choice of modelling the fundamental *versus* the realized niche (Kearney and Porter 2009), to assume equilibrium between environment and observed species patterns *versus* non-equilibrium (Hof et al. 2012) and individual species modelling *versus* community approach (Ferrier and Guisan 2006). Note that the choice of an appropriate spatial scale (Wiens 1989), selection of a set of conceptually meaningful explanatory variables and designing an efficient sampling strategy are equally important when formulating a given model and its objectives.

The second step in the development of empirical ecological niche models is the statistical model formulation. In other words, the choice of the statistical technique to be applied in order to model the relationship between the response and explanatory variables. The type of response variable (quantitative, semi-quantitative, qualitative) and its probability distribution has also a great influence on the selection of an appropriate technique. Some of the most popular species distribution modelling approaches comprises:

(i) Linear regression methods, which can be extended to model complex data types (e.g. fixed versus random covariates) and allow the inclusion of additive combinations of predictors and/or terms representing interactions between predictors;

(ii) GLMs, extensively used by ecologists for its ability to deal with data possessing different error structures, particularly presence/absence modelled via logistic regression. They consist of mathematical extensions of linear models that do not force data into unnatural scales, also allowing for nonlinearity and nonconstant variance structures in data;

(iii) Generalised Additive Models (GAMs, Hastie and Tibshirani 1990), which consist of a powerful extension of GLMs, gaining increasing popularity due to the ability of defining non-parametric smoothers to describe nonlinear responses, contributing with useful flexibility for fitting ecologically realistic associations;

(iv) Multivariate Adaptive Regression Splines (MARS, Friedman 1991) combines the strengths of regression trees with piecewise linear basis functions, which allows the modelling of complex relationships while possessing exceptional analytical speed;

(v) Ordination techniques, more specifically direct gradient analysis, providing axes that are constrained to be a function of environmental factors, i.e., sample scores are constrained to be linear combinations of explanatory variables. Canonical Correspondence Analysis (CCA) and Redundancy Analysis (RDA) are noteworthy approaches;

(vi) Classification tree analysis (CTA, Breiman et al. 1984, also referred to as classification and regression trees - CART) are machine-learning methods for constructing prediction models from data. The models are obtained by recursively partitioning the data space and fitting a simple prediction model within each partition. Trees explain variation of a single or multiple response variable by repeatedly splitting the data into more homogeneous groups, being each characterized by a typical value of the response variable, the number of observations in the group, and the values of the explanatory variables that specify it;

(vii) Artificial neural networks (ANN, Olden et al. 2008), whose information processing system is composed of a large number of highly interconnected elements called "neurons", working together to solve specific problems. They are popular in Ecology because they are considered to be universal approximations of any continuous function, being quite popular when modelling nonlinear relationships. Their full classification procedure is a complex non-parametric process;

(viii) Random forests (RF, Prasad et al. 2006), which creates multiple boot-strapped regression trees without pruning and averages the outputs, with each tree being grown with a subset of predictors entered in the model in a random order to avoid bias due to the inter dependencies among predictors. Typically a large number of trees is grown (500 to 2000), creating a limited generalization error, and thus reducing overfitting;

(ix) Boosted regression tree (BRT, Elith et al. 2008) improves the performance of a single model by fitting many models and combining them for prediction, using two algorithms: regression trees (from the classification and decision tree group of models) and boosting builds, which combines a collection of models. For regression problems, boosting is a form of "functional gradient descent", a numerical optimization technique for minimizing the loss function by adding, at each step, a new tree that best reduces (steps down the gradient of) the loss function. The first tree is the one that maximally reduces the loss function, followed by a tree that is fitted to the residuals of this first tree, which can contain quite different variables and split points compared with the first. The model is then updated to contain two trees (two terms), and the residuals from this two-term model are calculated, and so on. The final BRT model is a linear combination of many trees that can be thought of as a regression model where each term is a tree. BRTs have been showing interesting results, possessing the ability to account for uncertainty in model structure;

(x) Maximum entropy (MaxEnt, Phillips et al. 2006) estimates a target probability distribution by finding the probability distribution of the maximum entropy (in other words, that is most spread out, leading to uniform), subject to a set of constraints that represent incomplete information about the target distribution. It has gained popularity among studies that entail presence-only data (e.g., museum data), but can also be applied to presence/absence data by using a conditional model.

After selecting a given statistical approach, the model is calibrated on real data. Rykiel (1996) defines calibration as "the estimation and adjustment of model parameters and constants to improve the agreement between model output and a data set". However, the selection of proper explanatory variables is a vital component of this process. In an ideal modelling world, a model should be parsimonious, *i.e.*, should accomplish its desired level of explanation or prediction with as few predictor variables as possible, and this can be a difficult task since nature is not completely parsimonious. Due to the recognition of (living) nature as an extremely complex phenomenon, driven by several factors interacting at the same time, and in different orders of magnitude. One particular challenge to keep in mind is that parsimony is often obtained by the trade-offs between predictive power and complexity, and not necessarily taking into account how well we understand variable contribution to the model. For instance, two models may have similar predictive power and amount of predictors, but one is much easier to explain based on current knowledge of how a predictor is likely to affect the response. The choice of predictors can be done arbitrarily (which is not recommended for its inconsistency), automatically (via stepwise procedures available in linear regression methods and GLMs), following physiological and other ecological or mechanistic principles or by shrinkage rules (Harrell et al. 1996). After variable selection, model parameters are estimated and its fit is characterized, most of the time, by a measure of the variance reduction (or deviance reduction in maximum likelihood techniques). The model optimization through deviance reduction is performed through an estimated D^2 (equivalent to R^2 in least-squares models), being defined as:

$$D^2 = \frac{Null\ deviance - Residual\ deviance}{Null\ deviance}$$

where the null deviance is the deviance of the model with the intercept only, and the residual deviance is the one that remains unexplained by the model after the insertion of all selected predictors. The ideal model has no residual deviance and its D^2 equals one. Weisberg (1980) argues that the deviance formulation is not representative of the real fit, and proposes an adjusted version based on the number of observations n and the number of predictors p, which has been largely adopted:

$$adjusted\ D^2 = 1 - \frac{n-1}{n-p} \times (1 - D^2)$$

The value of the adjusted D^2 increases with an increasing n or a decreasing p in the model, and it allows for comparisons among nested models that include different combinations of explanatory variables. In tree-based approaches, no fit needs to be characterized since the model will attempt to predict the data exactly. Diagnostic tests for significance of estimated model coefficients can be performed based on the related variance or deviance distribution in least-squares and GLM estimation.

After model calibration, the next step consists of evaluating the model, also called *model validation*, which is the measurement of adequacy between model predictions and field observations, depending mostly on the specific purpose of the research and the domain in which the model is supposed to be applicable (Fielding and Bell 1997). Two main approaches are worth mentioning: (i) using a single data set to both calibrate and validate the model via cross-validation (Van Houwelingen and Le Cessie 1990), *leave-one-out* Jackknife (Efron and Tibshirani 1993) or bootstrap (Efron and Tibshirani 1993); (ii) having two independent data sets, using one for calibration and the other for validation, this approach being optimal and attractive. The first approach is usually selected when the data set is too small to be split into separate data sets but often used for large data sets as well. Regardless of the approach, two types of measure can be used to quantify the fit between predicted and observed values of the validation data set: (i) using the same *goodness-of-fit* methods used to calibrate the model or (ii) using any discrete measure of association between predicted and observed values (Guisan and Harrell 2000). In the case of presence-absence data (binary), probabilities are truncated at an adjusted optimal threshold (below threshold being predicted absent and above as present) that provides the best agreement between predicted and observed values. A *confusion matrix* (Guisan et al. 1998) expressing the number of true positives (predicted and observed present), true negatives (predicted and observed absent), false positives (predicted present but observed absent) and false negatives (predicted absent but observed present) is then analysed through the proportion of area correctly classified, the percentage of commission and omission errors, or using Cohen's κ (Cohen 1960), among other metrics. Another option for binary data is the use of a threshold-independent measure like the receiver operating characteristic (ROC) plot methodology (Fielding and Bell 1997). For quantitative data, the evaluation of predictions can be done simply by using Pearson's product-moment correlation coefficient if the variable is normally distributed, or a non-parametric rank correlation coefficient like Kendall's τ or Spearman's ρ.

With the increasing number of studies on predictive habitat distribution modelling, some key topics related to their limitations appear frequently and they comprise potential areas of investigation:

- Multiple scales: all species function at specific spatial and related temporal scales. However, their joint localized activities mediate processes that are important at the landscape scale (Anderson 1993). A hierarchy of scale-dependent abiotic factors, biotic interactions, population processes, disturbances and legacies govern their distribution (Ettema and Wardle 2002). Given this fact, species distribution modelling should either focus on appropriate scales that are relevant to the research question, the system, data availability or simultaneously consider multiples scales (Beever et al. 2006). Conceptually, there is no single natural scale at which ecological patterns should be studied (Levin 1992).
- Biotic interactions: most SDMs are calibrated under the assumption that biotic interactions do not influence species range patterns (Huntley et al. 1995) or only affect patterns at small spatial scales (Pearson and Dawson 2003). Recent developments aim at considering abiotic interactions (e.g., Boulangeat et al. 2012), though these approaches remain in their early stages. In models for understanding or interpolation-style prediction, the consequences may not be too severe, except where the presence of a host species is critical and not predicted by the available covariates. Studies were published showing how the incorporation of biotic interactions into SDMs better models species distributions and responses to environmental change. The information about these interactions is included, for example,

in the form of occurrence (Heikkinen et al. 2007), counts or frequencies (Leathwick and Austin 2001) or as a competition coefficient (Strubbe et al. 2010). The importance of biotic interactions may vary depending on the spatial scale and position along environmental gradients. In models built for extrapolation, like in the case of the effects of climate change on species distributions, the effects of competitors, mutualists and conspecific attractions might have far-reaching effects (Elith and Leathwick 2009).

- Uncertainty: it results both from data deficiencies (like missing covariates, small samples, biased species occurrences, lack of data about absence or inadequate sampling strategy) and from errors in model specification (Barry and Elith 2006). Few studies have addressed uncertainty in SDMs and its effects in model calibration, predictions and associated decision making. In management applications, it is important to investigate the impact of uncertainty and how to reduce it, characterizing and exploring its effects in decision making. Heikkinen et al. (2006) provide some useful information in various aspects of SDMs that contribute to uncertainty.
- The use of presence-absence data *versus* abundance: the relative use of presence-only and presence-absence data has been widely discussed (Elith et al. 2006), abundance data are available for many taxa in some regions, and it can provide additional information that might be better related to conservation status (Johnston et al. 2013), extinction risk (O'Grady et al. 2004) and community structure and function (Davey et al. 2012). Moreover, SDMs derived from abundance data may reflect the importance of key demographic and environmental factors such as carrying capacity (Pearce and Ferrier 2001). Howard et al. (2014) points out the importance of abundance data in predictive modelling, deriving a more accurate assessment of habitat suitability in contrast to presence-absence data. However, the choice of using presence-absence over abundance comes, in most cases, from the focus on methodological development to enhance

model performance (Guisan and Thuiller 2005, Elith et al. 2006) or pure lack of abundance data available, the latter being the most restrictive. Comparisons between models based on abundance *versus* presence-absence data, while addressing other SDM limitations, are the subject of potential future research.

- Spatial autocorrelation (SAC): spatially explicit predictive models are generally built with little or no attention to spatial processes that drive species distributional patterns. SAC occurs when the values of variables sampled at nearby locations are not independent from each other (Tobler 1970). The causes underlying are manifold, but three are worthy of mention: 1) biological processes such as speciation, extinction, dispersal or species interactions; 2) non-linear relationships between environment and species modelled (erroneously) as linear; 3) spatially-structured environmental features impose a spatial structure in the response (Legendre and Fortin 1989, Legendre and Legendre 1998). While spatial autocorrelation can provide information about biotic processes such as population growth, geographic dispersal, differential mortality, social organization or competition dynamics (Griffith and Peres-Neto 2006), if not properly accounted for, it can also cause serious drawbacks for hypotheses testing and prediction as it affects type I errors rates and precision in the estimation of model parameters (with the exception of machine-learning methods such as Random Forests). This is because SAC violates important statistical assumptions such as independently and identically distributed (i.i.d.) errors in both parametric and non-parametric testing (Lennon 2000). There are several approaches to incorporate spatial autocorrelation in statistical models, being these autocovariate regression, spatial eigenvector mapping (SEVM), generalised least squares (GLS), conditional autoregressive models (CAR), simultaneous autoregressive models (SAR), generalised linear mixed models (GLMM) and generalised estimation equations (GEE), among others. Dormann et al. (2007) addresses a comparison among these methods, reporting their efficiency and flexibility, with some remarks regarding the use of autocovariate regression. That said, different models are likely to be affected differently by spatial autocorrelation (e.g., logistic regression *versus* regression trees) and systematic comparison of the effects of SAC on model estimation and prediction is largely missing. There is still considerable debate as to whether spatial autocorrelation results in (statistically) biased coefficient estimates, how to best use explicitly spatial methods with incomplete sample data and whether previous studies that used non-spatial methods with spatially autocorrelated data should be considered fraught with error.

0.2 Modelling in the context of aquatic ecosystem services

Ecosystems generate a range of goods and services to society, which in turn directly contribute to our well being and economic wealth (de Groot et al. 2012). Over the past two decades, progress has been made in understanding how ecosystems provide services and how service provisioning translates into economic value (Daily 1997). Ecosystem services - the processes whereby ecosystems render benefits to people - are becoming the principal means for communicating ecological change in terms of human benefits (Daily 1997). Understanding ecosystem services is fundamental to decision-making efforts that influence multiple human activities and components of ecosystems, informing management and planning decisions such as the appropriate scale and location of a number of activities. Wise and sustainable decisions of this nature will require a comprehensive understanding of how changes in human activities and ecosystem states will result in changes in ecosystem services and the associated benefits to people. Valuing the contribution of ecosystems to human well-being through economic, ecological and social accounting demands robust methods to define and quantify ecosystem services. Yet, it has proven difficult to move from general pronouncements about the tremendous benefits nature provides to people to credible, quantitative estimates of ecosystem service values. Without quantitative assessments, and some incentives for landowners to provide them, these services tend to be ignored by those making land-use and land-management decisions. Decision making and policy aimed at achieving sustainability goals can be improved with accurate and defendable methods for quantifying ecosystem services (McKenzie et al. 2011).

Currently, there seems to exist a gap between how the regulation of ecosystem services is perceived and how they are managed. At one hand, the effects of landscape patterns at different spatial scales on ecological processes and species distributions have long been recognized by scientists and managers as crucial to understanding how such processes function (Hobbs 1997) and regulate ecosystem services and their sustainability. On the other hand, management strategies for freshwater ecosystem services most often focus on local systems (place-based) such as individual water bodies. Therefore, an integrated multi-scale framework where conservation, management and development of ecosystem services are coordinated is likely to result in the best approach (Abell et al. 2007).

Although an integrated, landscape oriented framework would likely offer improved tools for the effective and sustainable conservation of regional aquatic biodiversity and related services (Lester et al. 2003), knowledge and agreement on how such an approach can best be implemented is lacking. It is of utmost importance to design sampling strategies (field measures, site selection for monitoring) and quantitative frameworks to link ecological indicators of ecosystem health and provide the best estimates of an ecosystem's capacity to sustainably deliver ecological services. The focus on increasing the understanding of how ecosystem function is related to the delivery of aquatic ecosystem services, at several spatial scales, and how to assess the health of these ecosystems is of considerable relevance. Assessing whether ecosystems and their functional ability to deliver services have been impaired (or are at risk or are recovering) requires the possession of robust metrics to determine ecosystem status and trends. Moreover, reliable ecological knowledge regarding aquatic ecosystem services needs to incorporate ways to measure and understand the effect of cumulative impacts on ecosystem health (Duinker and Greig 2006). Finally, by comparing health conditions of different systems and their delivery regarding aquatic ecosystem services, it is possible to establish (1) how multiple natural and anthropogenic stressors interact and affect aquatic ecosystem services; and (2) an understanding of how resilience of biodiversity-ecosystem services is linked to environmental conditions and ecosystem health. This knowledge makes possible the design of landscape-oriented approaches that can provide much more effective information about the status of local and regional aquatic ecosystems and their related services. This is especially important given the growing support for aligning conservation efforts to ecosystem services (Goldman et al. 2008).

Fish populations in aquatic ecosystems benefit human societies in numerous ways, providing: (i) regulating services (e.g., regulation of food web dynamics, recycling of nutrients, regulation of ecosystem resilience, redistribution of bottom substrates, regulation of carbon fluxes from water to atmosphere, maintenance of sediment processes and maintenance of genetic species, ecosystem diversity); (ii) linking services (e.g., linkage within aquatic ecosystems, linkage between aquatic and terrestrial ecosystems, transport of nutrients, carbon, minerals and energy); (iii) cultural services (e.g., production of food, aquaculture production, production of medicine, control of hazardous diseases, control of algae and macrophytes, reduction of waste, supply of aesthetic values and recreational activities) and (iv) information services (e.g., assessment of ecosystem stress and resilience, revealing evolutionary histories, provision of historical information, scientific and educational information) (Holmlund and Hammer 1999). Certain ecosystem services generated by fish populations are also used as management tools, *e.g.* enhancing rice production (Tilapia, carp), mitigating diseases in tropical zones (mosquito control) or mitigating algal blooms (pike *Esox Lucius*). However, increasing fishing pressure, pollution, habitat destruction, introduction of exotic species and other factors continue to exert strong pressure on fish populations around the world (Malakoff 1997). Human-induced direct and indirect degradation of common fisheries resources might cause impacts at the ecosystem level, putting the fundamental and demand-derived ecosystem services generated by fish at risk with consequences for biodiversity and ecosystem resilience (Perrings et al. 1995).

Ecological modelling can provide a way of clarifying which factors and processes drive ecosystem services related to fish, and it is urgently needed. Modification and loss of aquatic habitat is recognized as the primary factor threatening the conservation of fish populations and their communities (Ricciardi and Rasmussen 1999). Species distribution models have a number of important applications to the conservation and management of fish populations and their related ecosystem services, playing an important role in prioritising surveys and monitoring programmes for fish populations because limitations to resources often preclude exhaustive and continual sampling of sites and that extensive sampling is needed to accurately sample lake fish communities (Jackson and Harvey 1997). Some applications include: (i) forecasting or measuring the effects of habitat alteration and changing land-use patterns (Oberdorff et al. 2001); (ii) providing first-order estimates of habitat suitability to establish potential locations for re-introduction (Evans and Oliver 1995); (iii) assess the impacts of the most recent climate change scenarios (Buisson et al. 2008); (iv) understand the factors that regulate the spread of invasive species and identifying their potential distributions (Wang and Jackson 2014); (v) predicting the likelihood of local establishment and spread of exotic species that may help set conservation priorities for preserving vulnerable species and populations that might be lost locally (Peterson and Vieglais 2001). (vi) predicting hotspots of species persistence for the conservation of biodiversity (Williams and Araújo 2000); and (vii) revealing additional populations of threatened species, or alternatively revealing unexpected gaps in their range. The first step towards a comprehensive assessment of ecosystem goods and services involves translating ecological complexity (abiotic factors and ecological processes) into a more limited number of ecosystem functions, which in turn provide ecosystem services (de Groot et al. 2002). The development of broad-scale perspectives to understand the nature, function, vulnerability and threats to fisheries-based aquatic ecosystem services are now deemed essential when designing and implementing scientifically-sound management strategies.

0.3 Thesis outline

The goal of this thesis was to conduct an investigation into aspects currently used modelling practices (more specifically species distribution models), seeking improvement of these and ultimately provide analytical tools to be applied in ecological assessments related to freshwater fish populations. To this end this thesis is comprised of three chapters, the first one conducted with simulated data and the others applied to freshwater fish data:

Chapter I: Generalized linear models for direct gradient analysis and variation partitioning of species data matrices

Chapter II: Assessing the role of community composition and abiotic factors in predicting fish species distributions

Chapter III: Effects of abiotic factors on the biomass of six fish species in Ontario lakes

Within Chapter I, I developed a framework called gRDA, and evaluated its performance via simulations for modelling multispecies data based on generalized linear models (logistic and Poisson). To date, the most used tool for multi-species modelling is based on linear regression (Gaussian; Legendre and Legendre 1998), which is known to have undesirable properties for presence-absence and abundance data. Moreover, my research within Chapter 1 found that, despite its popularity, logistic regression has extremely bad behaviour in terms of parameter estimation properties under a high number of covariates and when covariates are missing. To remediate this issue, I proposed a modified Poisson model that can model presence-absence data and that provides reliable estimates when covariates are missing or the number of covariates is high. Moreover, I evaluated the performance of a robust estimate for the coefficient of determination (R^2) for GLMs, which allows for the first time an appropriate variation partitioning scheme, a widely used tool in ecology (Peres-Neto et al. 2006), which is currently based on multiple regression.

Chapter II presents a comparison among a number of SDM approaches in terms of their predictive performance and explanatory power in predicting occurrence of six freshwater fish species in north-temperate lakes of Ontario, including the GLM based on Poisson distribution addressed in Chapter 1. These models were contrasted by using empirical data based on the Aquatic Habitat Inventory provided by the Ontario

Ministry of Natural Resources and Forestry. The modelling routine for Chapter 2 was performed taking into account three classes of models: (i) incorporating only environmental parameters; (ii) using solely the information about the fish community to predict the occurrence of a particular species (i.e. using biotic parameters) and (iii) using both environment and the information about the occurrence of other fish species as predictors (i.e. using abiotic + biotic parameters), seeking to identify the influence of biotic information to the predictive ability of these different models. Finally, variable relative importance assessments were also made across models, for each fish species and classes of models, in order to identify which variables were important in predicting presence-absence of a particular species.

Chapter III aimed at the development of biomass models using biomass-per-unit effort (BPUE) information from the same six freshwater fish species mentioned in Chapter 2, using environmental predictors based on principal environmental determinants (light, heat, lake morphometry, nutrients), together with fishing pressure. The species were chosen in order to disentangle the effects of climate and exploitation when evaluating the model results. To this end, four sets of environmental predictors were used: the first one consisted of the full model, which considered all main environmental predictors from the Broad Scale Monitoring program dataset, including angling pressure; the second one kept the same environmental predictors used in Chapter II to develop occurrence models, in order to evaluate the degree of agreement between occurrence (developed in that chapter) and biomass models; the third, a model with secchi depth, total dissolved solids (TDS) and pH encompassing the pool of variables representing water chemistry, in order to compare with the fourth set; the fourth set included total phosphorus (TP) and dissolved organic carbon (DOC) by replacing secchi depth and TDS from set 3, in order to evaluate which model better optimizes the full model, achieving model parsimony. In addition, variable relative importance was assessed for all models built, and partial dependence plots were obtained for the optimized model, seeking to evaluate the influence of each environmental feature to the biomass of the six freshwater fish species. These models brought a better understanding related to the importance of the different environmental determinants in the biomass composition of each species, including existing interactions among main variables influencing biomass.

In summary, Chapter I results provide ecologists with a new GLM-based tool for variation partitioning schemes, consisting of an improvement over the OLS-based ones; Chapter II presents occurrence models considering both abiotic and biotic features, showcasing the importance of adding biotic information when seeking to improve species distribution models developed for six freshwater fish species, and Chapter III investigates the main environmental determinants influencing the biomass of the same six species, while allowing us to assess similarities among the models developed in the last two chapters, and finally making use of one of the modelling approaches that best explained fish occurrence variability in Chapter II.

CHAPTER I

GENERALIZED LINEAR MODELS FOR DIRECT GRADIENT ANALYSIS AND VARIATION PARTITIONING OF SPECIES DATA MATRICES

W.S.C. Moreira and P.R. Peres-Neto

1.1 Summary

Describing and understanding species-environment relationships is one of the major goals in ecology. It has broad applicability to conservation planning and management such as uncovering the effects of ecological disturbances, climate change or how environmental conditions affect different communities across different spatial/temporal scales. One of the most common classes of approaches for modelling multiple species is Canonical analyses of which redundancy analysis is likely the most common form. Variation partitioning, an extension of canonical analysis, has become the de facto quantitative framework to estimate the relative contributions of predictors and groups of predictors to community assembly patterns, and it is based on ordinary multiple linear regression which assumes that the response variables (species distributions) are normally distributed. This assumption is clearly not appropriate for most species distribution data. It is interesting to note that although generalized linear models (GLM) are extremely used for modelling single-species distributions due to its flexibility in allowing the response (species) to follow different distributional families, variation partitioning has not been yet generalized to accommodate the different families within the GLM framework. In this study, we developed a multispecies framework that we refer to as "Generalized Redundancy Analysis" (gRDA) based on logistic and Poisson distributions via GLMs, and extend

the proposed framework to variation partitioning. Our goal is to present the framework as well as assess its performance using Monte Carlo approaches under a variety of scenarios including the correlation structure of different sets of predictors in determining species distributions as well as the influence of predictors without any contributions. Our results show that our proposed framework is very robust and should essentially replace the current standard variation partitioning based on ordinary Redundancy Analysis, contributing to variation partitioning's popularity and importance when analyzing species data matrices.

1.2 Introduction

Describing and understanding species-environment relationships is one of the major goals in ecology and has broad applicability to conservation planning and management such as uncovering the effects of ecological disturbances, climate change or how environmental conditions affect different communities across different spatial/temporal scales (Sharma et al. 2012, Stefan et al. 2001, Stauffer 2014). Species respond differently to complex environmental relationships and they often occupy only a fraction of all suitable environments in a landscape due to population demographics (e.g., not enough individuals produced to disperse to all sites containing suitable habitats), dispersal limitation (species cannot get to all suitable sites) and complex interactions with other species (e.g., competition and predation forcing species out of sites that are suitable). Although species and community models (Guisan and Zimmermann 2000; Peres-Neto et al. 2006; Sharma and Jackson 2007; Elith et al. 2008) have been proven as extremely valuable tools to explore and estimate the relative importance of different mechanisms (biotic and abiotic) underlying species distributions and their biodiversity patterns (e.g., similarities and differences among species regarding their environmental affinities), statistical approaches differ dramatically in their ability to model species-environment relationships particularly given the assumptions (or lack of assumptions) made about species distributions (Guisan and Zimmermann 2000; Elith et al. 2006).

One of the most common class of approaches for modelling multiple species is Canonical analyses in which redundancy analysis (RDA; Rao 1964) is likely the most common form. Canonical analyses provide the means of conducting direct analyses (e.g., direct gradient analysis) in which multiple species across communities (or time periods) can be studied with respect to their commonalities or differences in their relationships with multiple predictors (e.g., environmental, spatial). Variation partitioning, an extension of canonical analysis, has become the *de facto* quantitative framework to estimate the relative contributions of predictors and groups of predictors to community assembly patterns (Borcard et al 1992; Cottenie 2005; Legendre et al 2005; Peres-Neto et al 2006; Soininen 2016). In this context, variation partitioning allows us to estimate: (1) How well species distributions are explained (predictive ability) by the independent contribution of different sets of predictors? (2) How different sets of predictors co-vary in their predictive ability? Because variation partitioning estimates unique and common fractions of variation in species distributions due to sets of predictors grouped together, it provides a way to determine the relative contributions of broad mechanisms explaining patterns of community structure instead of the relative importance of multiple individual predictors within each preditctor set. Examples include the relative importance of environmental versus spatial factors (e.g., Soininen 2016) and the importance of environmental factors grouped into different spatial scales (e.g., local land cover versus regional climate; Heino and Alahuhta 2014).

Variation partitioning based on canonical analyses uses ordinary multiple linear regression which assumes that the response variables (species distributions) are normally distributed. This assumption is not appropriate for species distribution data (presence-absence, abundance), Peres-Neto et al. (2006) showed that without proper species distributions transformations (e.g., Hellinger-transformation), variation partitioning lead to biased estimates of variable (sets) contributions. However, transformations are not likely to make data conform well to the different statistical properties that species distributions may take. It is interesting to note that although generalized linear models (GLM) are extremely used for modelling single-species distributions due to its flexibility in allowing the response (species) to follow different distributional families (Elith and Leathwick 2009), variation partitioning has not been yet generalized to accommodate the different families within the GLM framework. Generalized linear modelling (McCullagh and Nelder 1989) allows one to specify the distribution of the data, thus implicitly defining the relationship between the mean and the variance of each observation (e.g., abundance of a species in a site) to be a function of its predicted value (i.e., variance is not constant as in ordinary regressions).

In this study, we developed a multispecies framework that we refer to as "Generalized Redundancy Analysis" (gRDA) based on logistic and Poisson distributions via GLMs. Here we concentrate in these two models given their popularity and because robust and consistent estimators for their coefficients of determination (R^2) are available. One of the advantages of variation partitioning based on R^2 is its interpretability as the proportion of variation in the response variable explained by a set of predictors. Note, however, that although R^2 estimators for the logistic and Poisson regression have existed now for quite some time, their properties remain relatively unknown even in the statistical literature. The proposed framework, however, is flexible enough that it can be readily extended to other distributional families (e.g., Gamma, Exponential) if consistent coefficients of determination (R^2) are developed for other distributions. In essence, we present a framework in which direct gradient analysis can be performed via GLMs (gRDA) and extend the proposed framework to variation partitioning. Our goal is to present the framework as well as assess its performance using Monte Carlo approaches under a variety of scenarios including the correlation structure of different sets of predictors in determining species distributions as well as the influence of predictors without any contributions. Our results show that our proposed framework is very robust (yield good performance) and should essentially replace the current standard variation partitioning based on ordinary Redundancy Analysis.

1.3 Methods

1.3.1 gRDA - Redundancy Analysis and Variation Partitioning via Generalized Linear Models

Redundancy Analysis is an extension of multiple regression (linear and non-linear; Rao 1964; Makarenkov and Legendre 2002) in which each response variable (species) is regressed independently against a predictor matrix (e.g., environmental variables, spatial predictors). Then the vectors of predicted values are juxtaposed into a single matrix that undergoes a principal component analysis (PCA) to establish how species partition themselves across environmental gradients. As such, a GLM extension for an RDA (gRDA) is straightforward in which predicted values are estimated for each species based on particular distributional properties of the response (e.g., logistic, Poisson). This is a simple extension that, to our knowledge, it has not been made into the literature yet. Although RDA (and other types of canonical analyses) is commonly used in the ecology, its computational details remain somewhat cryptic (but see Legendre and Legendre 1998). A more intricate extension is the use of gRDA in the context of variation partitioning because the relative total and unique contributions of sets of predictors are based on adjusted coefficients of determination (R^2) and their semi-partial contributions. Peres-Neto et al. (2006) have shown that the Wherry (1931) correction for multiple regression with one response variable behaves well for variation partitioning in the case of normally distributed responses (RDA on multiple response variables) as it is an unbiased sample estimator of the true population R^2 (see Yin and Fan 2001 for a review on adjusted R^2 for the case of ordinary regression). In this study we investigate the robustness of adjusted R^2 for GLMs based on one response variable to the case of multiple responses (gRDA).

Among GLM families, logistic regression is likely the most widely used approach in species distribution modelling given that presence-absence data are relatively easy to collect and compare across different sampling designs (Rushton et al. 2004; Pearce and Ferrier 2000b). Notwithstanding its popularity, the logistic model may have undesirable properties (accuracy of parameter estimation) when data do not fit well a logistic curve, under overdispersion, in the presence of multicollinearity, with small sample sizes, high number of covariates and when relevant predictors in the model are missing (Begg and Lagakos 1990; Robinson and Jewell 1991; Pearce and Ferrier 2000a; Sharma and Jackson 2008). Given these issues, it has been suggested that a Poisson regression could potentially serve as a better approximation for modelling dichotomous outcomes than the logistic regression (Mittlböck and Heinzl 2001), though this has not been tested in the literature. This extension is particularly interesting because the Poisson distribution has been considered a good underlying basis for developing approaches to model overdispersed and zero-inflated presence-absence data (Yee and Dirnböck 2009).

Variation partitioning for two sets of predictor matrices (say **X** and **W**) are based on three sets of regressions, though extensions to more than two sets are relatively straightforward. Here we concentrate on the two-predictor set case as the results are generalizable to any number of sets. The total amount of variation explained by the two-predictor set (fraction [abc]) is estimated as the R^2 adjusted with all predictors from both sets into one regression model; the amount explained by predictor matrix **X** (fraction [ab]) and **W** (fraction [bc]) are estimated by R^2 adjusted from two separate regression models only containing the predictors in each set. Unique (i.e., independent of one another) and shared (collinear) contributions of each set are estimated via subtraction: the unique contribution of **X** independent of **W** is [a]= [abc]-[bc]; the unique contribution of **W** independent of **X** is [c]= [abc]-[ab]; the shared contribution between **X** and **W** is [b]=[abc]-[a]-[c]; and the residual fraction is [d]=1-[abc]. Below we detail the R^2 adjusted metrics we used to estimate the [abc], [ab] and [bc] fractions and the Monte Carlo simulations we used to assess their robustness under different scenarios.

1.3.2 Measures of adjusted explained variation

We consider the following adjusted measures of explained variation:

(i) Adjusted R^2 for logistic regression

Given a sample of n observations (patches, sites, communities) ($\mathbf{y}=[y_i]$, $\mathbf{X}=[x_{i1}, x_{i2}\ldots x_{ip}]$), $i=1,\ldots,n$, $y_i \in \{0, 1\}$ denotes the response variable and x_i is the vector containing values across p predictors for the i^{th} observation. The estimates from a logistic regression are $\mathbf{Prob}(\mathbf{y}_i=\mathbf{1}\,|\,\mathbf{x}_i)=\hat{\mathbf{p}}_i=\dfrac{\mathbf{1}}{\mathbf{1}+\boldsymbol{exp}^{-(\hat{b}_0+\hat{b}_1\mathbf{x}_{1i}+\hat{b}_2\mathbf{x}_{2i}+\ldots+\hat{b}_p\mathbf{x}_{pi})}}$, where $\hat{\mathbf{p}}_i$ is the estimated probability for the i^{th} observation based on the estimated vector of p parameters. The proportion of one's in the sample is given by $\mathbf{Prob}(\mathbf{y}_i=\mathbf{1})=\bar{\mathbf{p}}=\sum_{\mathbf{i=1}}^{\mathbf{n}}\mathbf{y}_i\,/\,\mathbf{n}$. Let $D(y_i)=(y_i-\bar{p})^2$ and $D(y_i\,|\,x_i)=(y_i-\hat{p}_i)^2$

denote the squared distance (residual) between observed (y_i) and predicted ($\overline{p}$ and $\hat{p}_i$) values under the null model (only with intercept b_0) and under the full model (including all covariates), respectively. Mittlböck and Schemper (1996) suggested the following adjusted measure of explained variation R^2:

$$R^2_{adj} = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}$$

where $SSE = \sum_i^n D(y_i|x_i)$, $SST = \sum_i^n D(y_i)$, n is the number of local communities (sites) and k the number of predictors. The interpretation of this measure is intuitive due to the fact that it is based on the use of squared residuals, varying between 0 and 1 unlike AIC and other metrics alike, which cannot be interpreted as the relative amount of variation explained by a model. Other adjusted measures have been suggested by Mittlböck and Schemper (1996), but as they are based on likelihoods, they do not have a clear interpretation when compared to the use of squared residuals.

(ii) Adjusted R^2 for Poisson Regression

Cameron and Windmeijer (1996) and Waldhör et al. (1998) defined R^2 for Poisson regression based on deviances, while Mittlböck and Waldhör (2000) proposed bias adjustments for small and/or many covariates (adjusted R^2). In practice, however, data are often over- and sometimes underdispersed as compared to the standard Poisson model. As such, Heinzl and Mittlböck (2003) presented a generalization based on the adjusted R^2 measure of Mittlböck and Waldhör (2000) that has a dispersion parameter.

Consider

$$D(y;\hat{\mu}) = \log L(y) - \log L(\hat{\mu}) = \sum_i [y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)]$$

and

$$D(y;\bar{y}) = \log L(y) - \log L(\bar{y}) = \sum_i \left[y_i \log(y_i/\bar{y}) - (y_i - \bar{y}) \right].$$

Where $D(y;\hat{\mu})$ is the deviance of the full model containing the covariates of interest and $D(y;\bar{y})$ is the deviance of the intercept-only model, $y = (y_1, \ldots, y_n)'$ are the observed values of the dependent variable and $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_n)'$, the corresponding predicted values under the full model.

The adjusted measure of explained variation for Poisson regression models (Heinzl and Mittlböck 2003) has the form:

$$R^2_{adj} = 1 - \frac{D(y;\hat{\mu}) + k\hat{\phi}_D}{D(y;\bar{y})}$$

where $\hat{\phi}_D = D(y;\hat{\mu})/(n-k-1)$ represents the dispersion parameter for over- or underdispersed Poisson regression.

1.3.3 Simulation study considering one single matrix of regressors

To evaluate the accuracy of the adjusted R^2 measures described above, a Monte Carlo study was carried out using simulated presence-absence data and abundance data. The basis for generation of presence-absence data was a logistic model, using a binomial random number generator and abundance data was generated on the basis of a Poisson model, using a Poisson random number generator, each considering k=4 covariates. Population values consisted of a **X** (100000 x 4) matrix containing 4 random normally distributed variables $N(0,1)$ and **Y** was generated according to:

$$E[\boldsymbol{Y}|\boldsymbol{X}] = p = \frac{1}{1 + e^{-\boldsymbol{BX}}}$$

for the logistic regression case, and

$$E[\boldsymbol{Y}|\boldsymbol{X}] = p = e^{\boldsymbol{BX}}$$

for the Poisson case. **B** is a (50 x 5) matrix containing the slopes for 50 species on each of the 4 covariates plus an intercept (first column of **B**). The slopes were randomly originated from set intervals for each variable, which had numbers defining the lower and upper limit of each interval in such way that rates of presence as low as 10% and as high as 90% were generated, common features in ecological data. The generated **Y** in the model population is a (100000 x 50) matrix, in which each column represents the presence-absence (logistic-generated) data or abundance (Poisson-generated) of a given species. Model estimates for each data matrix were computed using the respective generalized linear model approach (logistic or Poisson) and the adjusted R^2 measures for each model were estimated at both population and sample levels. For each GLM type, samples were performed according to two scenarios: (i) 1000 samples with varying number of observations (20, 50, 100, 150 and 200) were randomly drawn from the population [**Y**, **X**] to assess the influence of sample size on the estimation; (ii) 1000 samples of size n = 100 were randomly drawn from the population and a number of normally distributed variables N(0,1) (5, 10, 15 and 20) were added to the sample covariate matrix **X** to evaluate the impact of unimportant predictors on the estimation of adjusted R^2 values.

1.3.4 Variation Partitioning - simulation study considering two groups of regressors

This second simulation study evaluated the adjusted R2 measures under variation partitioning involving three data matrices [**Y**, **X**, **W**], where **X** and **W** are two separate sets of predictors. In common applications of variation partitioning in ecology, these matrices often represent a set of local environmental predictors **(X)** and spatial predictors (**W**). Here, different levels of correlation between these two matrices were considered to allow estimating how the adjusted R2 measures behave in estimating unique and common fractions of variation. The population generation and simulation structure followed the one described in above (1.3.3). In order to generate correlation structure between **X** and **W**, a matrix **XW** was created containing 8 normally distributed variables N(0,1) and was post-multiplied by the Cholesky decomposition of a (8 x 8) correlation matrix containing all cross-correlation values set to a pre-specified number, resulting in matrix **XW**. The first 4 columns of **XW** were then defined as **X**, and the last 4 as **W**, while **Y** was created according to the generating model as before. Here the matrix of slopes **B** is (50 x 9) and it contains the slopes of both **X** and **W**, as well as an intercept (first columns of **B**). As for the case of one set of predictors, the influence of sample size (20, 50, 100, 150, 200) and number of unimportant predictors (5,10,15,20) were also assessed. Note, however, that unimportant predictors were only inserted in matrix **W**. For each type of data (presence-absence or abundance), variation partitioning was performed as described earlier (session gRDA - Redundancy Analysis and Variation Partitioning via Generalized Linear Models) using the adjusted R2 associated with each generating model for the data.

As noted earlier, the Poisson model may provide a better specification to presence-absence data than the logistic model. However, given that term $\log(y_i/\hat{\mu}_i)$ cannot be defined when y_i equals zero, we added a constant (here 1) to all presence-absence values; i.e., instead of modeling zeros and ones, we modeled ones and twos. Because variance is invariant with respect to changes in a location parameter, any constant added to all y_i values does not affect the adjusted R^2 for the Poisson model.

1.3.5 Missing predictors scenario – logistic regression case

Given that logistic regression is known to affect parameter estimation of model parameters (slopes) when relevant predictors are missing (Begg and Lagakos 1990), we investigated the performance of the adjusted R^2 estimators when relevant variables are missing in the logistic model. A species data matrix **Y** (presence-absence) was generated as before containing 100000 sites and 50 species, according to $\hat{p} = 1/(1+\exp(-(0+3x_1+0.8x_2+0.8x_3+2x_4)))$. Predictors ($x$) were sampled from a normal distribution $N(0,1)$. 1000 samples of size n = 100 were taken from the model population ([**Y**,**X**]) and adjusted R^2 estimation was performed considering the exclusion of one to three covariates in the model. The choice of these slope values allow the evaluation of how missing variables having both large and small contributions to the model (e.g., x_1 versus x_2) affect the adjusted R^2 estimation. Due to fact that the number of relevant predictors in a model should obviously affect (increase or decrease) the true (population) adjusted R^2, sampling variation in estimates need to be contrasted considering samples from a population containing only the missing predictor in question and a population containing all four predictors. For example, adjusted R^2 estimates are obtained from a population generated with all four variables (i.e., x_1, x_2, x_3, x_4) but sample models only containing x_1 (for example; i.e., x_2, x_3, x_4 missing from the model; here referred as "missing" case) were contrasted with samples from a population generated with only x_1 as a predictor (referred here as ALL case). In the case of linear ordinary least squares (OLS), adjusted R^2 is not affected by missing predictors that do not co-vary with the other relevant predictors considered in the model. As such, sample variation would not vary considering the two types of samples above (i.e., ALL versus missing cases). However, given that logistic regression slope estimates are affected by missing relevant predictors that do not covary with the predictors considered in the model, we suspect that adjusted R^2 should be affected as well.

1.4 Results

1.4.1 Simulation results considering one group of predictors

Mean absolute errors for sample adjusted R^2 estimates in relation to the true population value are presented in Appendix A and Figure 1.1 for both presence-absence (logistic and Poisson implementation) and abundance data (Poisson implementation). The Poisson model applied to presence-absence data yielded greater precision in contrast to the logistic model for all sample sizes except for the smallest number of observations considered in the simulation (20 observations; Figure 1a). The estimation precision of adjusted R^2 considering irrelevant predictors is adversely affected in the case of presence-absence data using the logistic implementation but the estimator is quite reliable when using the Poisson implementation (Figure 1b). The Poisson model for abundance data show that the adjusted R^2 is not affected by increased numbers of irrelevant predictors (Figure 1b), though the mean absolute error is slightly greater than the presence-absence for the logistic implementation.

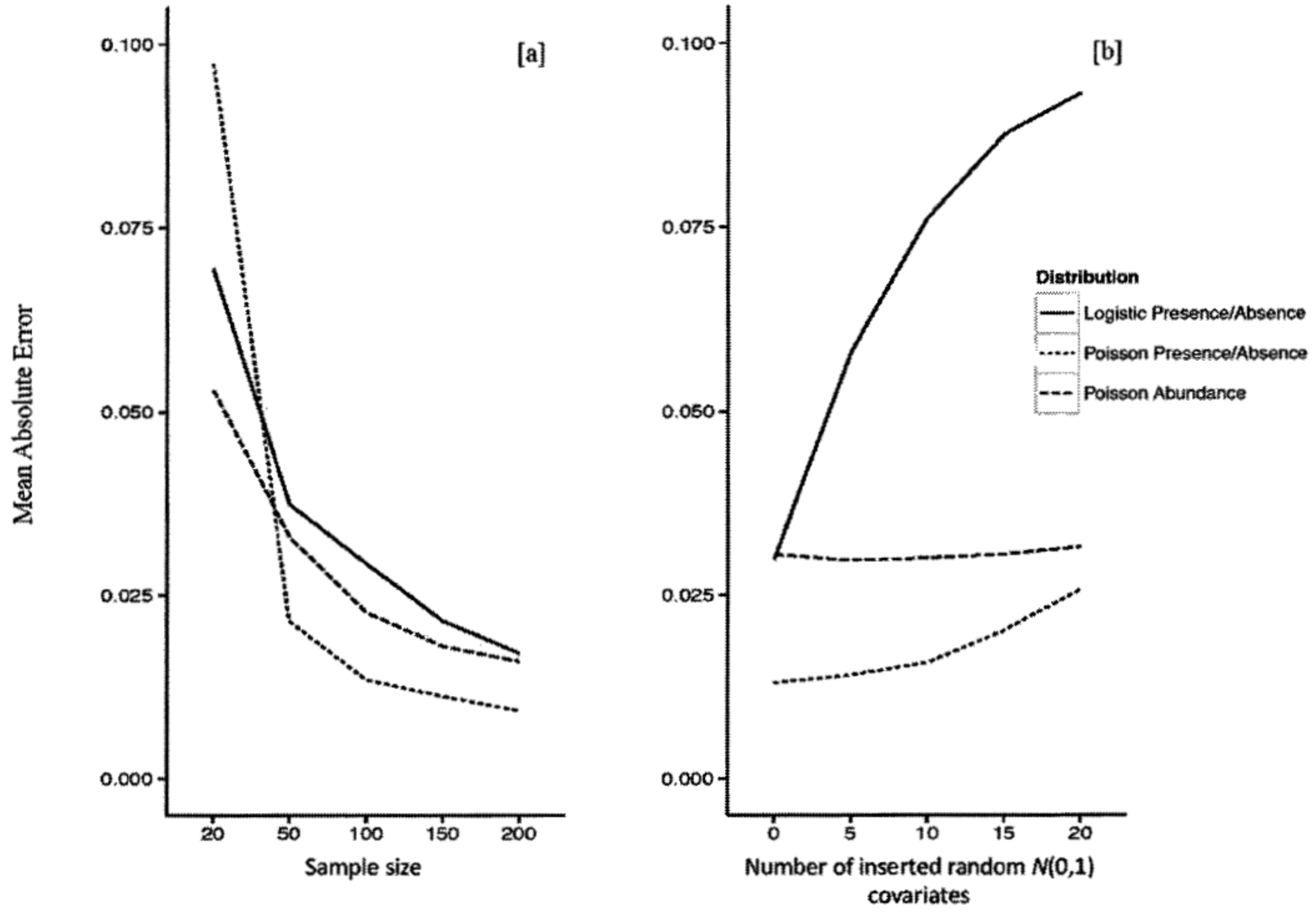

Figure 1.1: The influence of sample size and insertion of random $N(0,1)$ covariates on R^2 accuracy, expressed as mean absolute error. No covariate was included to the model in the sample size influence assessment, while samples were based on 100 observations in the number of inserted random covariates influence case.

Figure 1.2 presents the average adjusted R^2 across 1000 samples (sample size = 100) and one standard deviation for presence-models missing different number of relevant predictors contrasting ALL and missing cases (see 1.3 Methods for more details). The results clearly indicate that the logistic model heavily bias adjusted R^2 estimates when variables are missing and that the adjusted R^2 estimates based on the Poisson model are much less affected by missing variables.

1.4.2 Simulation study considering two groups of predictors and assessment of GLM Poisson and logistic measures of explained variation through variation partitioning

Average fraction values based on variation partitioning estimated from GLM logistic and Poisson models across different correlation values for the two predictor matrices (**X** and **W**) are presented in Appendix B. Estimation accuracy was assessed via mean absolute differences between the true population and samples values (Figure 1.3) for each sample size, model type and correlation level between predictor matrices.

In the case of presence-absence data the results from the Poisson based implementation leads to smaller sampling errors of fractions when contrasted to the logistic model (Figure 1.3). Both models are quite sensitive to small sample sizes (20 observations) but become quite robust when considering sample sizes of 50 observations or greater. Within presence-absence models (logistic and Poisson implementation), all fractions ([a], [b], [c], and [d]) have the same amount of sampling error and the correlation (level of collinearity) among predictor matrices does not affect sampling error (Figure 1.3). Variation partitioning based on Poisson regression

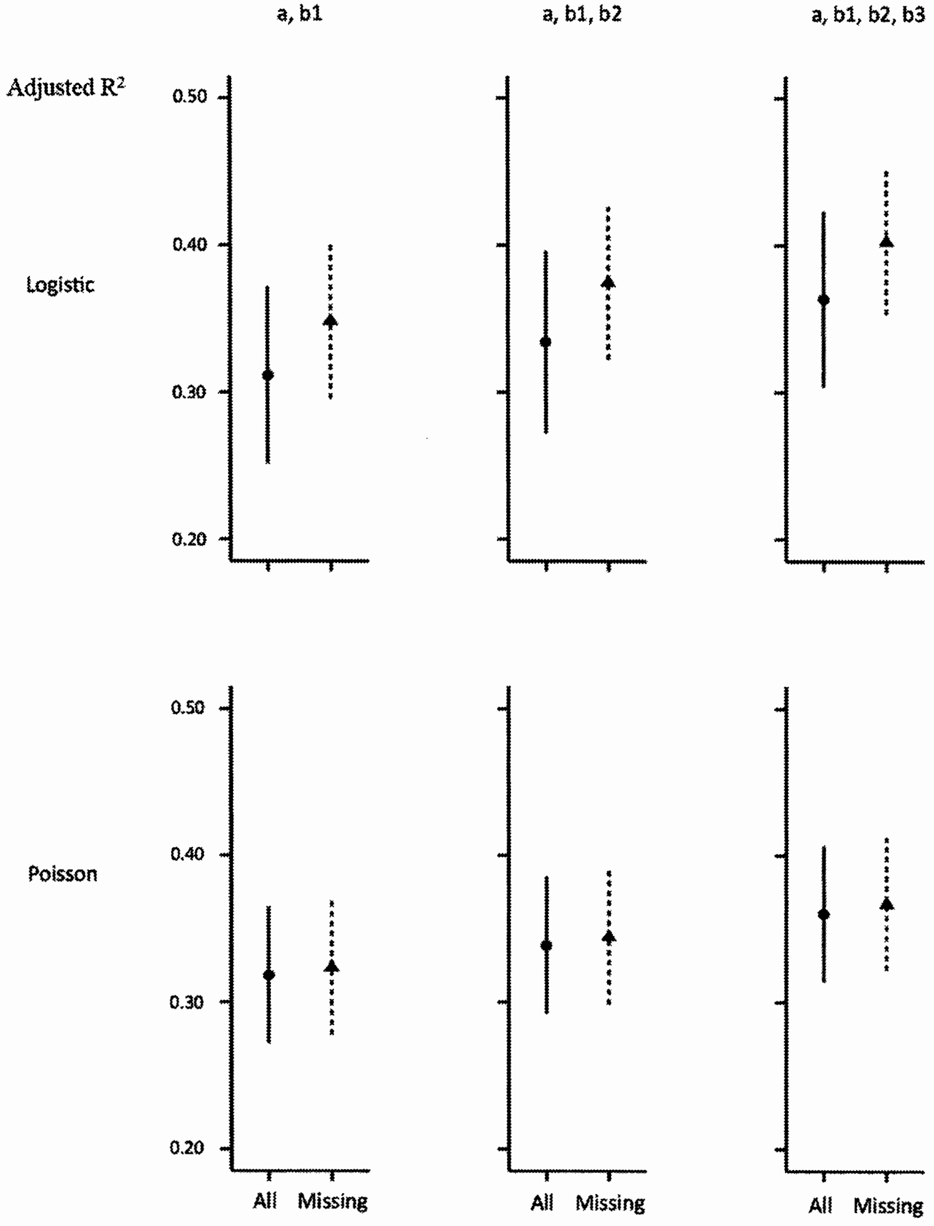

Figure 1.2: The influence of missing covariates on R^2 estimation. Values are expressed as mean ± standard deviation. “a” stands for intercept while “b1”, “b2” and “b3” stand for the first, second and third slopes, respectively.

(abundance data) generates slightly larger sampling errors for fraction [b] in contrast to other fractions. Sampling error for all fractions is affected by the correlation structure where relatively smaller sampling errors are achieved when predictor matrices are more correlated (Figure 1.3).

Appendix C and Figure 1.4 presents fraction estimation accuracy results when an increasing number of random $N(0,1)$ of predictors are added to **W** in order to estimate the impact of the number of irrelevant predictors on estimation of variation partitioning fractions. The Poisson implementation for presence-absence data is much less affected by irrelevant predictors than logistic regression. Interestingly, fraction [d] (residual variation) is greatly affected in contrast to the other fractions for the Poisson implementation for the presence-absence data, whereas for the Poisson model for abundance data, it is fraction [b] that is largely affected. Sampling error increases with number of irrelevant predictors in the case of the Poisson implementation but it remains unchanged for the Poisson regression based on abundance data.

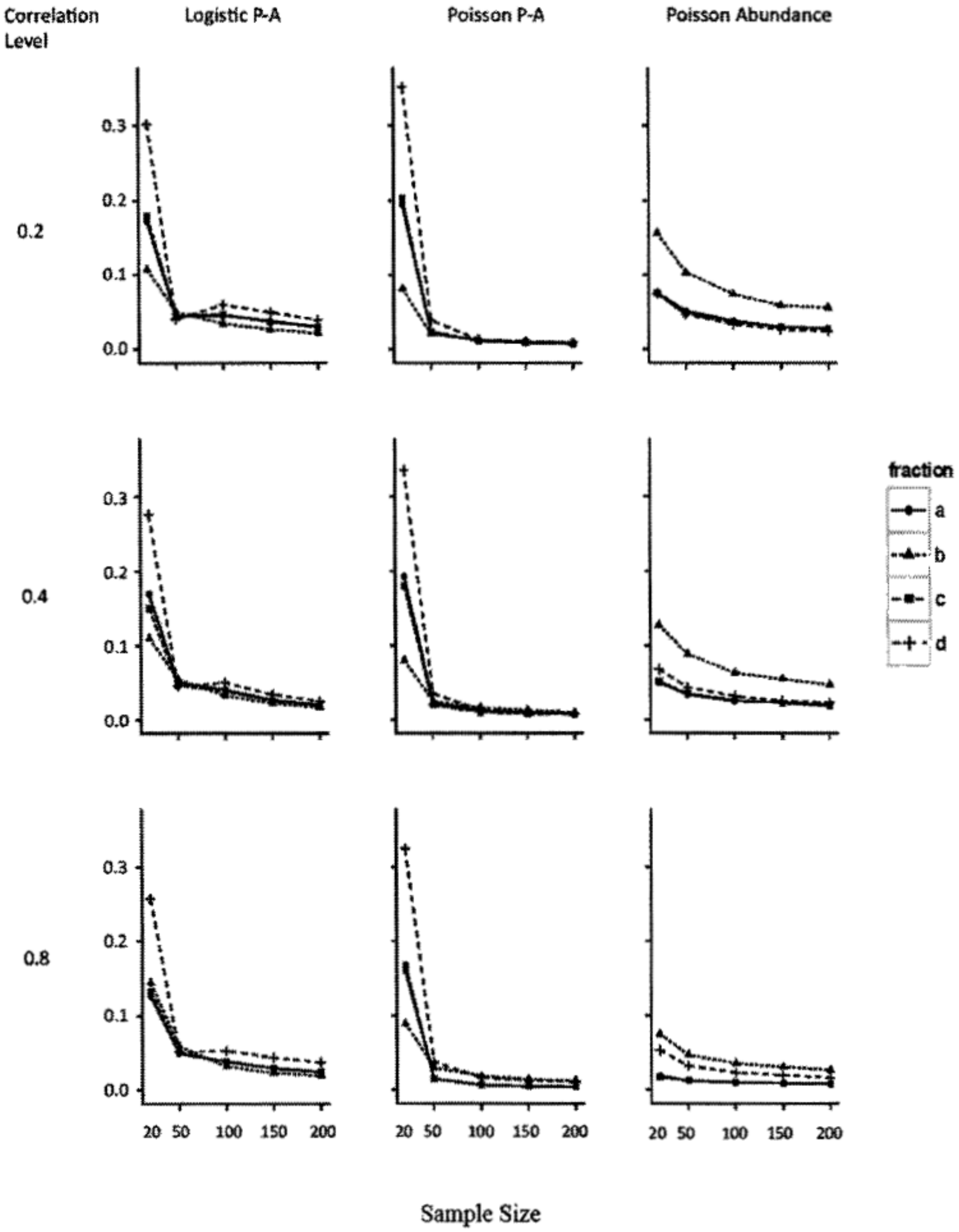

Figure 1.3: The influence of sample size on fraction estimation accuracy, expressed as mean absolute error (Y axis), in R^2 variation partitioning according to varying levels of correlation between **X** and **W**.

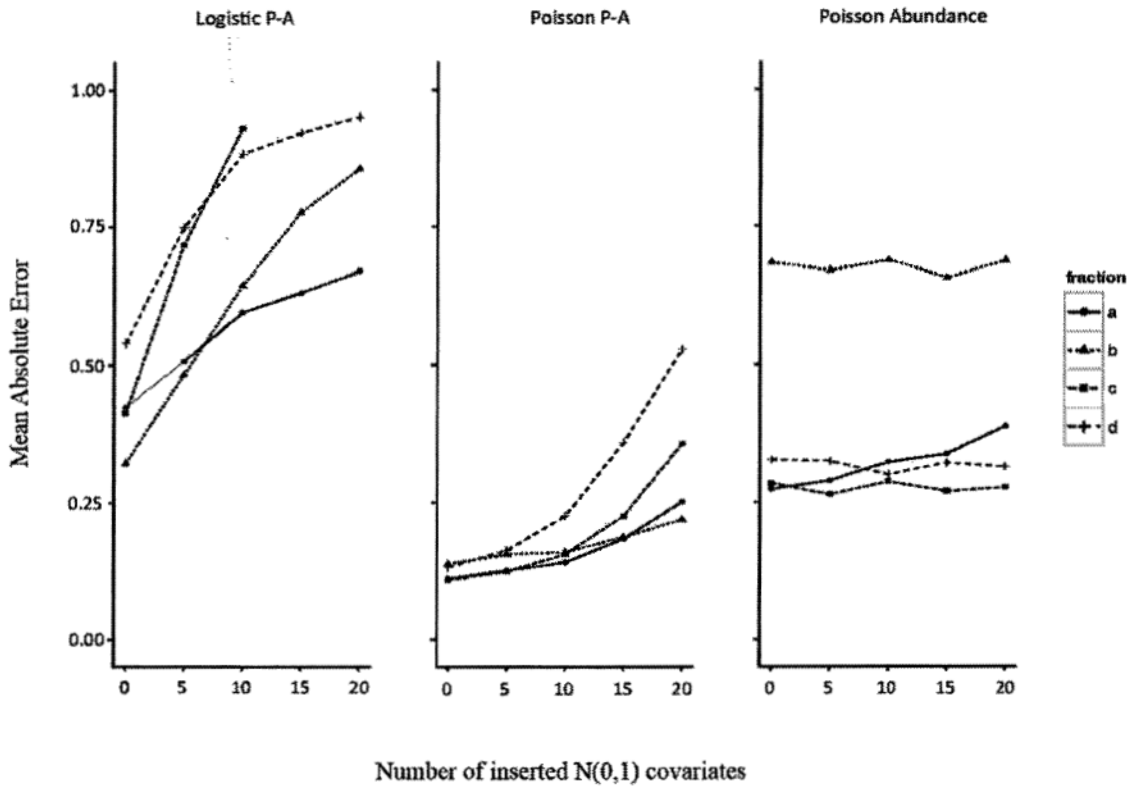

Figure 1.4: The influence of insertion of random $N(0,1)$ covariates on fraction estimation accuracy, expressed as mean absolute error, in R^2 variation partitioning. Samples were based on 100 observations. The degree of correlation between **X** and **W** is equal to 0.4.

1.5 Discussion

Our main goal in this study were (i) to introduce a multispecies framework named “Generalized Redundancy Analysis” (gRDA) based on logistic and Poisson distributions via GLMs; (ii) extend gRDA to perform variation partitioning which has become a central quantitative tool in studying the underlying factors structuring community structure. We have used simulations to demonstrate its performance and how different model specifications perform and compare. To date there has been no attempt to extend variation partitioning to GLM-based approaches. The approach proposed here widens the scope of applicability of variation partitioning, contributing to its popularity and importance when analyzing species data matrices.

Since canonical analyses have become a central quantitative framework to explore and identify patterns and relationships in community data, it is important to investigate alternate methods to be applied in cases where linear relationships cannot be assumed between species distributions and environmental/spatial data (i.e., presence-absence and abundance data). Moreover, our results showed that a Poisson implementation to presence-absence data provides a better model specification (small sampling variation) than the traditional logistic regression. The Poisson distribution is commonly applied as a model for number of events of a given dependent variable, however, as Mittlböck & Heinzl (2001) reported, there is an approximate equivalency between logistic and Poisson regression regarding modeling dichotomous outcomes.
GLM logistic accuracy was seriously impacted by irrelevant predictors, while GLM Poisson had only a slight decrease in estimation quality (Figure 1 [b]). This is an important characteristic as in real cases, we never know which predictors are truly relevant or irrelevant.

The sensitivity of logistic regression to large number of covariates (relevant or irrelevant) is a known problem (Le Cessie & Van Houwelingen 1992) and here we showed that it can impact proper estimation of adjusted R^2 values. GLM Poisson applied to presence-absence data also performed better when relevant predictors were missing in contrast to logistic regression (Figure 2). Zou (2004) reported the robustness of Poisson estimates in cases of omitted covariates when compared to logistic estimates, the latter being usually chosen due to its ease of application and popularity. The good performance of GLM Poisson when examining presence-absence data advocates its use particularly under a multispecies framework where multiple parameters across species need to be estimated and is likely to accumulate greater levels of error in contrast to single species models. Not surprising, the GLM Poisson applied to abundance data showed good performance (i.e., smaller sampling error for adjusted R^2) when compared to the logistic model for presence absence.

To place the errors we observed into perspective, the average absolute error for sample means based on 100 observations from a $N(0,1)$ population is about 0.08. The errors we obtained based on much more complex errors (GLM based models considering multiple predictors and species) are greater than 0.08 but well under accepted limits for the Poisson implementation for presence-absence data and Poisson regression for abundance data. Results for the simulation study considering two groups of predictors and to assess the performance of GLM-based adjusted R^2 in a variation partitioning scheme (Figure 3) showed that the accuracy of fractions estimates for the two models were quite satisfactory when $n \geq 50$ for both the Poisson implementation for presence-absence data and Poisson regression for abundance data. These results clearly show the robustness of our variation partitioning scheme using GLM-based implementations, computed for multispecies presence-absence and abundance data. Note, however, that the adjusted R^2 metrics used here were quite sensitive to small sample sizes ($n<50$) and should not be used in these cases. Our simulation results show that our GLM-based variation partitioning scheme is influenced somewhat by irrelevant predictors (Figure 4) with GLM logistic being more affected from added covariates. Note that only four relevant predictors were considered in matrix **W** where a relative large number of irrelevant predictors were added to the important predictors.

Although error estimates varied across fractions as a function of sample size, model specification and number of irrelevant predictors, adjusted R^2 estimates are unbiased (i.e., average sample value equals population value) particularly when sample size is greater than 50 and the correlation structure between predictors is not extremely large ($r=0.8$) (Appendix B). Although we have not covered model selection procedures in this study, future investigation may prove useful in determining how estimates of GLM-based R^2 adjusted values are affected by variable selection procedures. This is particularly important if one wants to reduce the sampling errors associated with potential irrelevant predictors, though Peres-Neto et al. (2006) showed that forward selection biased upwards adjusted R^2 values for normally distributed errors (i.e., GLM based on an identity link). Omitting important variables results in systematic bias in the estimation of regression coefficients and predicted values, and including too many predictors results in loss of precision in the estimation of regression coefficients and the predictions of new responses (Murtaugh 1998). Variable selection methods have been proposed for both linear and logistic regression, but not much effort has been put towards Poisson regression of abundance data.

Our results demonstrate the robustness of GLM-based procedures under an RDA and variation partitioning framework. For several decades now ecologists have resorted to OLS solutions that are not likely to be robust for non-linear species-environmental associations. We hope that this study compels ecologists for a shift in analytical tools towards our developments and continue to improve GLM-based tools for variation partitioning schemes which have become core in the study of communities.

CHAPTER II

ASSESSING THE ROLE OF COMMUNITY COMPOSITION AND ABIOTIC FACTORS IN PREDICTING FISH SPECIES DISTRIBUTIONS

W.S.C. Moreira, P.R. Peres-Neto and N. Lester

2.1 Summary

Effective conservation and management of freshwater fish species would be greatly aided by understanding the factors shaping fish species' spatial distributions and their patterns of biodiversity. The number of SDM studies applied to different fish species has increased in past years with a special attention to the prediction of range shifts caused by climate change. Understanding the importance of environmental factors affecting different fish species and having a distribution model for each species of interest should yield more effective conservation plans. The majority of SDMs make use of only environmental variables to predict species distributions and ignore other key determinants of species ranges, most pertinently, biotic interactions. Examining both the impacts of environmental factors and species interactions on the distribution of any given species is essential to understand how climate change, or other major environmental changes, will affect biodiversity. The objective of this chapter consisted of developing fish-habitat models for the occurrence of six species of economically important freshwater fishes in north-temperate lakes of Ontario, providing a comparison of: (i) variable importance among species and (ii) predictive performance among six popular SDM approaches. We expected that the models possessing both abiotic and biotic predictors would outperform models including only abiotic information. Our results confirmed that for all approaches and classes of

models. Promising results found were mostly related to random forest and boosted regression trees, with the latter always displaying the highest explained deviances values. The results presented by this study corroborates the idea of model improvement through the addition of biotic variables, in this particular case, including occurrence of other fishes from the same community as predictors. Large lakes allows the development of abundant vegetation in the littoral zone, which facilitates the spawning of species such as smallmouth bass, supporting the fact that surface area of the lake was an important predictor in all six fish occurrence models produced in this study. Climate was also important in shaping occurrences of the same species. The results related to the importance of lake size to all species serve as a basis for conservation measures that incorporate large lakes as part of the conservation strategy, given their importance in providing adequate habitat for several species.

2.2 Introduction

Effective conservation and management of freshwater fish species can be improved by understanding the factors shaping their spatial distributions. Species distributions are driven by abiotic (e.g. suitable climate) and biotic conditions (e.g. competitive interspecific interactions), as well as movement and dispersal (Soberon and Nakamura 2009). Species distribution models (SDMs) are widely used in ecology and routinely used to relate species occurrences to environmental predictor to generate models of environmental suitability. SDMs provide a snapshot estimation of species-environment relationships and generate predictions of the distribution of a given species across geographic scales. The number of SDM studies applied to different fish species has increased in past years with a special attention to the prediction of range shifts caused by climate change (Chu et al. 2005, Sharma and Jackson 2008, Bond et al. 2011, Filipe et al. 2013). Fish distributions could be strongly affected by global warming as most species have no physiological ability to regulate their body temperature (Wood and McDonald 1997). Also, climate-change drivers such as temperature and precipitation can operate directly to affect range, or indirectly through affecting habitat, feeding resources (e.g. Pratchett et al. 2011) and reproduction (Pankhurst and Munday 2011). Additionally, because the dispersal of freshwater fish is constrained by the network structure of drainage basins (Grant et al. 2007), their distributions and associated services may be more vulnerable to broad-scale environmental changes than terrestrial species. A generalized conservation plan is likely to be less effective because the species in a fish assemblage have differing life history strategies, habitat requirements, and sensitivities to stressors (Maloney et al. 2006). As such, species-specific SDMs should yield more effective conservation plans.

The majority of SDMs make use of only environmental variables to predict species distributions and do not include other key determinants of species ranges, most importantly, biotic interactions (see Elith and Leathwick 2009; Beale et al. 2014), as including these interactions have tended to require a priori knowledge and data on species interactions. Although macroclimate drives species distribution patterns at large scales (Whittaker 1975; Pearson and Dawson 2003; Thuiller et al. 2005), there is increasing evidence of the important role of biotic interactions in determining species' range extents (Jablonski 2008; Wiens 2011; Wisz et al. 2013). Biotic interactions can take place at local scales (e.g. predation, parasitism, competition and disturbance) or regional scales (e.g. dispersal, speciation, extinction, expansions or contractions of species ranges) (Cornell and Lawton 1992; Amarasekare 2003). Examining both the impacts of environmental factors and species interactions on the distribution of multiple species is essential to understand how climate change, or other major environmental changes, will affect biodiversity.

Despite limitations with current SDM approaches, their popularity in the realm of prediction of species distributions will remain certain. In addition to their application in conservation planning and management, SDMs can provide insights into the past and future distributions of species and the factors that shape their biogeography (Elith et al. 2006). That said, without explicitly considering biotic factors, SDMs could potentially misinform conservation management decisions. Accurate representations of species distributions are vital for the design and implementation of appropriate conservation measures, e.g. protected areas, but SDMs are unlikely to produce reliable predictions if they rely only on abiotic factors (Hof et al. 2012; Kissling et al. 2012).

Studies that use SDMs are beginning to consider biotic interactions by including the occurrence of other species' as additional covariates (e.g. Pellissier et al. 2010) and this has been central in improving performance (Araujo and Luoto 2007; Mod et al. 2014). Moreover, many recent theoretical and empirical studies have addressed the importance of including competition, predation or resource-consumer interactions in the development of SDMs (Guisan and Zimmermann 2000; Araújo and Guisan 2006; Zimmermann et al. 2010; Wisz et al. 2013) since their role in shaping patterns at broad spatial extents has been largely dismissed. Understanding the role of biotic interactions in determining species distributions, however, is difficult and often hindered primarily by the availability of data and then by the difficulties in inferring causation, multicollinearity, the complexity of species interactions and spatial and temporal variation in those interactions (Svenning et al. 2014).

The objective of Chapter 2 is to develop fish-habitat models for the occurrence of six species of economically important freshwater fishes in north-temperate lakes of

Ontario, providing a comparison of: (i) variable importance among species and (ii) predictive performance among six popular SDM approaches, including the GLM based on Poisson distribution considered in Chapter 1. Comparison of predictive models is relevant because the performance behaviour of different modelling approaches often depend on the characteristics of the study system of interest and it will determine which approach is most appropriate for that given system. As such, regional SDMs (here north-temperate lakes of Ontario) are relevant as they generate insights into the underlying mechanisms and relationships among species while addressing their errors and uncertainties useful for regional conservation policies, which are often regional themselves. Three model specifications will be developed for each fish species: (i) adopting only environmental parameters as predictors; (ii) using solely the information about the fish community to predict the occurrence of a particular species (i.e. using biotic parameters); (iii) incorporating both environmental and biotic information about the occurrence of other fish species as predictors (i.e. using abiotic + biotic parameters). Performance comparisons will be performed among these three classes of models as well. To date no study considering the different target species has covered such an extensive comparison of different SDM methods. We expect that the models possessing both abiotic and biotic predictors outperform models including only abiotic predictors.

2.3 Methods

2.3.1 Fish-environment data source

Presence-absence data on the fish species targeted in this study were gathered from the Aquatic Habitat Inventory (AHI) conducted in the province of Ontario (Canada).

The inventory is a compilation of lake surveys conducted by the Ontario Ministry of Natural Resources between 1970 and 1984. All surveys were conducted using standard methods as described in Dodge et al. (1984). The total number of lakes surveyed was 9885 (Figure 2.1), and the fish community surveys were conducted concurrently with the collection of lake environmental characteristics (see Mandrak and Crossman (1992) for the history of the data base and Goodchilde and Gale (1982) for sampling methods). The abiotic parameters considered here were surface area of the lake (L10Area) log transformed using base 10, growing degree days above 5°C for 1961-1990 (DD5), mean annual air temperature for 1961-1990 (MAT), maximum and mean depth of the lake (Depth_Max and Depth_Mean, respectively), secchi depth, total dissolved solids (TDS), pH and shoreline development factor (SDF). SDF is a measure of lake shape complexity; it compares the outline to that of a circle (SDF = 1 implies a circle), as follows:

$$SDF = \frac{Shoreline}{2 * \sqrt{\pi * Area}}$$

where shoreline (which includes islands) and area are measured in the same units (e.g. km and km^2). Figure 2.2 shows the strong latitudinal gradient in climate across the province. Table 2.1 summarizes descriptive statistics related to each abiotic parameter considered here and Table 2.2 provides information on the correlations among all abiotic parameters to be used in the species distribution models. Overall, the correlations among abiotic parameters are between low and moderate, with the exception of the ones found between DD5 and MAT, and mean and max depth. Mean and max depth are both measurements of depth, which justifies the correlation found between them; both are important for habitat characterization of different thermal guilds. Because species often use different environmental cues to maximize different life-history stages (e.g. young *versus* old individuals) and needs (e.g. reproduction *versus* growth), different types of variables that relate to similar factors (e.g., different ways to represent lake temperature) are important because they may influence species differently across these life stages and needs (Lee et al. 2016).

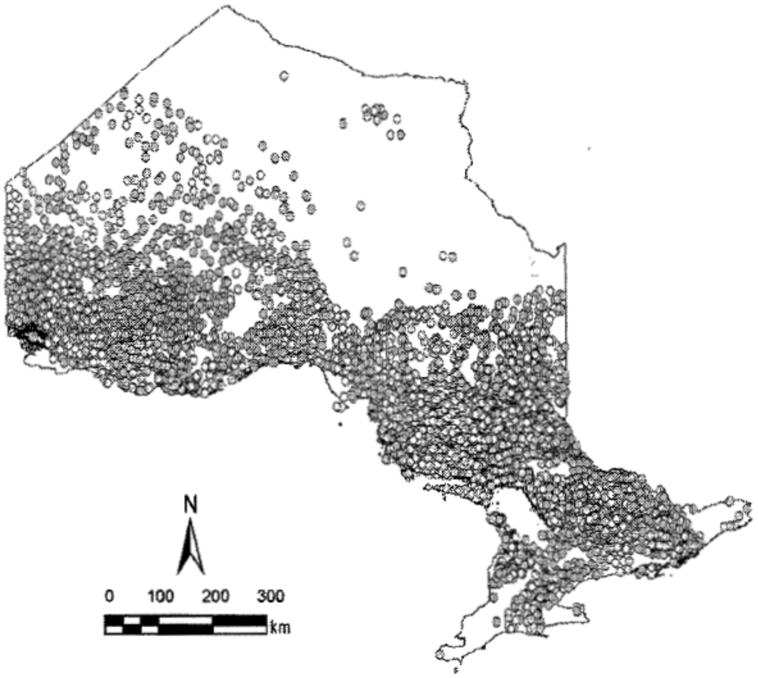

Figure 2.1: Distribution of the lakes surveyed across Ontario and compiled in the Aquatic Habitat Inventory.

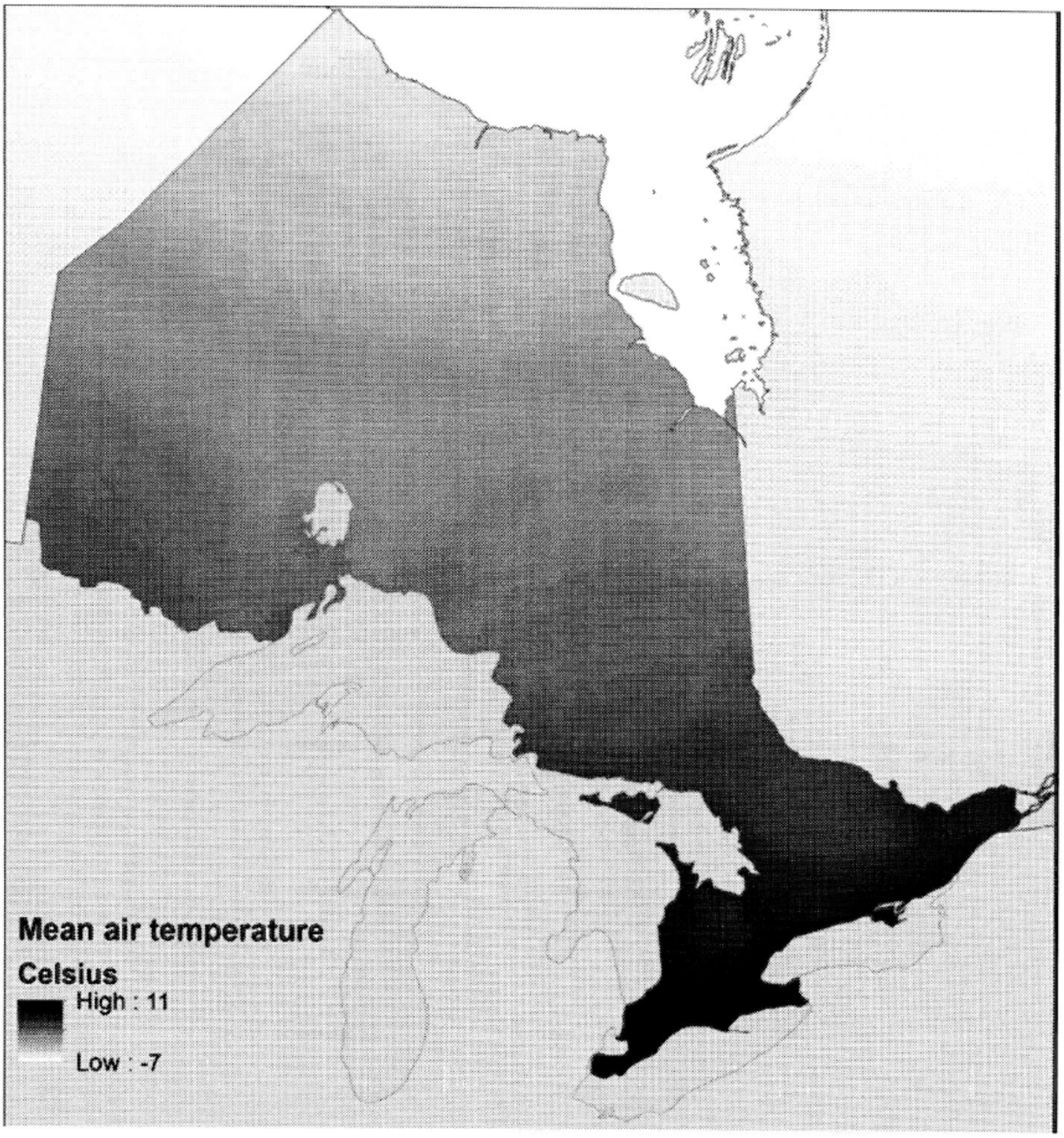

Figure 2.2: Latitudinal gradient in climate, represented by mean annual air temperature measurements across Ontario.

Table 2.1: Descriptive statistics of all abiotic parameters

	Units	n	Mean	Standard Deviation	Range
DD5	growing degree-days	9435	1438.8	222.5	1679
MAT	°C	9437	2.745	1.952	13.66
L10Area	ha	9437	1.762	0.805	6.15
SDF		9378	2.60	1.97	81.2
Depth_Max	m	9395	16.86	14.09	213.2
Depth_Mn	m	9318	5.54	4.41	47.4
TDS	mg/L	8955	59.39	56.62	611.7
Secchi	m	8313	3.67	2.04	36.4
pH		4989	7.08	0.67	5.2

Table 2.2: Pearson correlations among abiotic parameters

	DD5	MAT	L10Area	SDF	Depth_Max	Depth_Mn	TDS	Secchi	pH
DD5	1.00								
MAT	0.90	1.00							
L10Area	-0.15	-0.27	1.00						
SDF	-0.02	-0.09	0.52	1.00					
Depth_Max	0.12	0.14	0.29	0.23	1.00				
Depth_Mn	0.20	0.20	0.18	0.06	0.89	1.00			
TDS	0.04	0.07	0.01	-0.09	-0.11	-0.13	1.00		
Secchi	0.24	0.30	-0.09	0.00	0.50	0.56	-0.06	1.00	
pH	-0.21	-0.26	0.17	-0.07	-0.15	-0.15	0.63	-0.18	1.00

2.3.2 Target fish species

Fish species considered in this study were: smallmouth bass (*Micropterus dolomieu*), walleye (*Stizostedion vitreum*), northern pike (*Esox lucius*), white sucker (*Catostomus commersonii*), lake trout (*Salvelinus namaycush*) and lake whitefish (*Coregonus clupeaformis*). They all represent species that are commercially and recreationally important and much is known about their habitat preferences (Scott and Crossman 1998). Figure 2.3 shows the distribution of each freshwater fish species across Ontario. Table 2.3 presents some key ecological temperature metrics data on these species, which helps defining their thermal requirements and are strongly related to their fundamental niche since water temperature plays an important physiological role and influence reproduction (Hasnain et al. 2010).

Canadian freshwater fishes are classified into three thermal guilds: cold, cool and warm, which are assigned in accordance to their preferred summer water temperature. Thermal guild is assigned as "cold" if the preferred temperature is < 19°C, "cool" if it is between 19-25°C and "warm" if it is greater than 25°C. Lake trout and lake whitefish are classified as cold-water fishes, walleye, northern pike and white sucker are classified as cool-water fishes, and smallmouth bass, as warm water fish. In lakes, young lake whitefish are typically found at the surface (depths smaller than 1 m) in association with emergent vegetation and woody debris over substrates of boulder, cobble, and sand. Juveniles are found over the same substrates as young lake whitefish and also associate with woody debris and emergent vegetation. After spawning, adults move to deeper water (>10 m) to overwinter. They do not appear to have a substrate preference and have been found over boulder, gravel, cobble, sand, and clay. They are primarily bottom dwelling but may also be found in the pelagic zone (Sawatzky et al. 2007). In Ontario, lake trout are generally found in the pelagic zone of lakes at depths greater than 10 m, with temperatures of 10°C approximately. During the summer, such temperature can only be found in deeper, cooler portions below the thermocline in lakes that stratify (Sawatzky et al. 2007), and during spring and fall, they can be found in surface waters. Walleye prefer large, turbid lakes but will inhabit clear lakes, large streams and rivers if sunken trees or boulder shoals provide shelter from daylight (Scott & Crossman, 1998). Northern pike occurs in weedy slow rivers and, more frequently, weedy bays of lakes, requiring cover (e.g., logs, weeds, stumps) to capture prey and are more active during the day. Lacustrine adults typically occur in water smaller than 5 m deep, except during winter, when they are found in deeper water (Sawatzky et al. 2007). Lacustrine young white sucker remains in the gravel substrate for 1-2 weeks after hatching, inhabiting shallow waters along the shore over substrates of rock and sand and are often found in association with vegetation. They move to deeper offshore waters later in the summer. Lacustrine adults inhabit depths of 7-13 m (Sawatzky et al. 2007). Optimum lacustrine habitat for smallmouth bass is characterized by large, clear lakes and reservoirs with average depth greater than 9m with rocky shoals. They exhibit strong, cover-seeking behavior and prefer protection from light in all life stages (Scott & Crossman, 1998).

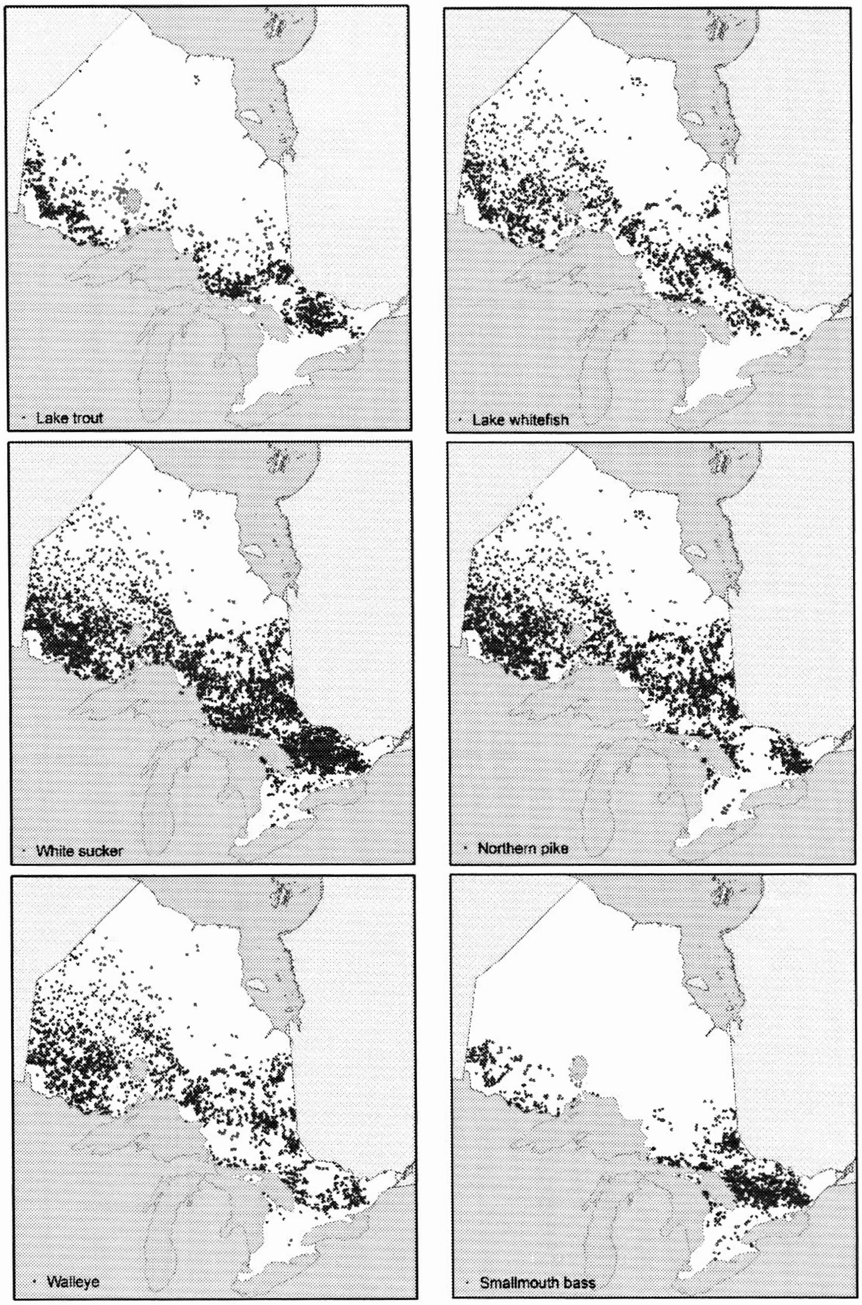

Figure 2.3: Distribution of the six freshwater fish species across Ontario, compiled in the Aquatic Habitat Inventory.

Table 2.3: Thermal metrics associated to growth, survival and reproduction of the fish species considered in this study (data extracted from Hasnain et al. 2010). Temperature preference class: cold (<19°C), cool (19 to 25°C) and warm (> 25°C); OGT: optimum growth temperature, the temperature that supports the highest growth rate; CTMax: critical thermal maximum, indicator of thermal resistance, the temperature at which a fish loses its ability to maintain a "normal" upright posture in the water; OS: optimal spawning temperature, the one at which spawning reaches its peak; OE: optimum egg development temperature, the one at which the rate of successful egg development is highest.

Common Name	Scientific Name	Spawning Season	Temperature Preference Class	OGT (°C)	CTMax (°C)	OS (°C)	OE (°C)
Lake trout	*Salvelinus namaycush*	Fall	cold	10	-	-	-
Lake whitefish	*Coregonus clupeaformis*	Fall	cold	14.7	-	3.05	4.95
Walleye	*Stizostedion vitreum*	Spring	cool	22.1	23.4	7.73	12.2
Northern pike	*Esox lucius*	Spring	cool	23	-	11.5	12.05
White sucker	*Catostomus commersoni*	Spring	cool	25.5	31.6	15.83	15
Smallmouth bass	*Micropterus dolomieu*	Spring	warm	26	36.3	18	21

- No data available

2.3.3 Species distribution modelling approaches

Predictive models were built using the following approaches: Generalised linear model (GLM) based on Poisson distribution; Generalised additive model (GAM); Multivariate adaptive regression splines (MARS); Random forests (RF); Boosted regression trees (BRT)and Maximum entropy approach (MaxEnt). These models were chosen due to their popularity in the literature.

Refer to section 0.1 to obtain more details about the approaches considered here. All techniques were implemented using functions and packages from the R programming environment (R Development Core Team). GLM was implemented using the "glm" function from the "stats" package. GAM was implemented using the "gam" function from the "mgcv" package. MARS was implemented using the "earth" function from the package of same name. RF was implemented using the "randomForest" function, and its related package has the same name. BRT was implemented using the "gbm.step" function and MAXENT was implemented using the "maxent" function, both from the "dismo" package.

Three classes of models were built for each SDM technique: one containing only abiotic variables (called "A"); one containing only biotic information in the form of presence-absence of the fish species considered in this study (e.g. walleye occurrence predicted by northern pike, white sucker, lake trout and lake whitefish occurrences), this model class called "B"; and one containing both abiotic variables and biotic information via presence-absence of fish species, called "A + B". Model calibration and validation was conducted via five-fold cross-validation: at each iterative fold, 80% of the whole dataset was withheld for model calibration and the other 20% used for model validation. Lakes were randomly assigned to each group at every iteration, preserving the proportion of presences and absences. Model residuals were computed through the differences between the observed values from the dataset and fitted values estimated from calibration process for each modelling technique, and expressed as mean squared error (MSE). MSE is a highly informative metric about model performance and consist of a well-known diagnostic measure to detect discrepancies between the data and fitted values (Lin et at. 2002). The adjusted deviances were computed according to the formula shown in section 0.1.

In this study, model performance was also evaluated using both threshold-independent (AUC) and threshold-dependent (sensitivity, specificity and TSS) measures of accuracy. Models producing non-binary scores (i.e., probabilities) were evaluated by applying a threshold to transform the scores into a dichotomous set of presence-absence predictions, and generating the corresponding confusion matrix necessary to calculate TSS. The threshold chosen was the one that maximizes sensitivity + specificity. The description of each evaluation metric follows:

- Sensitivity, the proportion of observed presences that are predicted as such, quantifying omission errors. Calculated as TP/(TP + FN), where TP = true positives and FN = false negatives. In general, FN can result from data or model inaccuracies or even a threshold chosen with a high value;
- Specificity, the proportion of observed absences that are predicted as such, quantifying commission errors. Calculated as TN/(TN + FP), where TN = true negatives and FP = false positives. FP may result from detection issues or biotic or abiotic factors that prevent a given species from occupying suitable habitat, such as dispersal limitations;

- The area under the curve (AUC) of the receiver operating characteristic (ROC, Fielding and Bell 1997), constructed by using all possible thresholds to classify the scores into confusion matrices, obtaining sensitivity and specificity for each matrix/thresholds, and then plotting sensitivity against the corresponding proportion FP. AUC ranges from 0 (no discrimination ability) to 1 (perfect discrimination). It is the most popular measure of accuracy reported by the literature;
- True skill statistic (TSS, Allouche et al. 2006), a synthetic index defined as TSS = sensitivity + specificity – 1. TSS ranges from -1 (performance no better than random) to +1 (perfect agreement).

The best-performing models were determined considering the residual metrics computed during the calibration process, and the quantitative assessment of their predictive performance using the above described indexes in the validation step. In other words, since we have six calibration and validation metrics in total, models that show the best performance in most (or all) of them are elected as best models. Sensitivity, specificity, AUC and TSS were calculated for all SDM models using the "evaluate" function from the "Dismo" package, R programming environment (R Development Core Team).

Variable importance was computed for all six SDM approaches and three model classes built for all six species. In order to standardize the computation of this metric, each variable received a rank varying from 1-9 for A models, 1-5 for B models and 1-14 for A + B models. The higher the rank number, higher the importance of a given variable in a given model. The ranks were established on the basis of both the estimated coefficients for regression models and variable importance proportion estimated from machine-learning methods. After variable importance estimation, the degree of agreement among models in terms of their ranked variables' importance was evaluated using Spearman's rank correlation coefficient (also called "Spearman's rho"), a nonparametric measure of rank correlation that assesses how well the relationship between two variables can be described using a monotonic function. Spearman's rho assumes values between -1 (a negative relationship between the two variables) and +1 (a positive relationship), with 0 denoting absence of relationship.

2.4 Results

Calibration and validation metrics results computed for all six fish species are compiled in Appendix D. These metrics values were based on averaging the results from five subsets of the whole data, according to the five-fold cross-validation procedure. Residual inspection of MSE values (Appendix D) revealed that models were improved when both abiotic and biotic variables were taken in account. The one exception was GLM which performed better with only environmental variables in most cases, but still showed high values of MSE in comparison to the other models, indicating a poor calibration performance overall. The random forests model (RF) performed better in terms of precision (i.e., MSE), but quite similar in performance to the generalized additive model (GAM), which also showed good precision performance.

2.4.1 Fish species distribution models' quality-of-fit assessment via deviance estimation

A comparison of goodness-of-fit results based on deviance is presented in Figure 2.4, according to the three classes of models (A, B, A+B). Overall, adjusted deviance values were improved when abiotic and biotic variables were included together in all six modelling approaches, showing greater levels of variation explained in contrast to model specifications A (only abiotic) and B (only biotic) models, with the latter showing the lowest fit to date. Across all models, the most successful in reducing the residual deviance was BRT. Deviance results for MAXENT are not shown because the function related to this model ("calc.deviance" from "Dismo" package) computes the deviance based on observed and predicted values, however there were cases in which this function presented deviance results above 1, indicating a possible problem since the maximum value that deviance can assume is 1 (100%). Deviance results for smallmouth bass, walleye, northern pike and lake trout showed good variance explanation, especially the A+B model specification for lake trout, which showed an optimal deviance result of 66.4% (Figure 2.4), indicating that the abiotic and biotic set of variables used to model this species' presence-absence is a good candidate for explaining distribution patterns. The lowest deviance results were seen for white sucker models, with values below 30%, while the lake whitefish deviance results were mid-low in the 20-50% range.

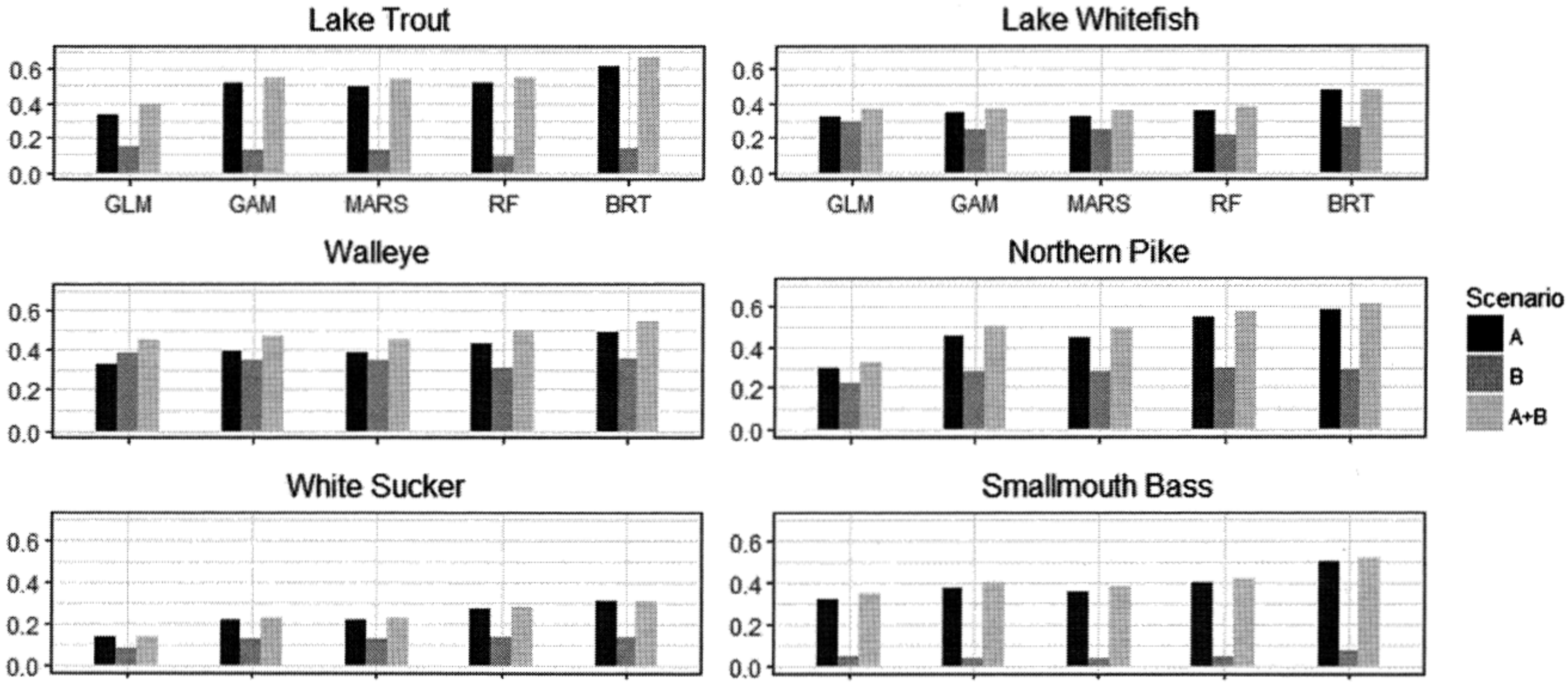

Figure 2.4: Models' adjusted deviance for all six freshwater fish species target of this study, according to the three model specifications: (A) models including only abiotic variables, (B) models using only biotic information, and (A+B) models incorporating both abiotic and biotic variables. GLM: generalized linear model based on Poisson distribution; GAM: generalized additive model; MARS: multivariate adaptive regression splines; RF: random forests; BRT: boosted regression trees.

2.4.2 Assessment of predictive accuracy

The predictive accuracy indicate good performance for all models (Figures 2.5 – 2.11). The same trend of an increase in performance was observed for A+B models in comparison to A and B models separately. Figure 2.5 shows that multivariate adaptive regression splines, random forests and boosted regression trees showed the best predictive accuracies through the inspection of both AUC and TSS, however the differences among all models are very small (numerical values presented in Appendix D). The sensitivity + specificity values computed for MARS, RF and BRT models reflect their good performance across all fish species (Appendix D). MAXENT and GLM were the ones with the lowest performance, generating slightly lower predictative accuracy of predictive accuracy among all species.

Lake trout models showed the highest values of predictive accuracy among all species (Figure 2.6, Appendix D), which reflects the good calibration performance observed through the estimated deviance values. The predictive accuracy obtained for walleye models including only biotic information (B models), in some cases, was slightly higher than the one based on models using only abiotic variables (A) (Figure 2.7). This indicates that the biotic variables included in walleye B models possess approximately the same importance as the abiotic ones used to predict this fish species. This statement was further examined by looking at models' variable importance for each species.

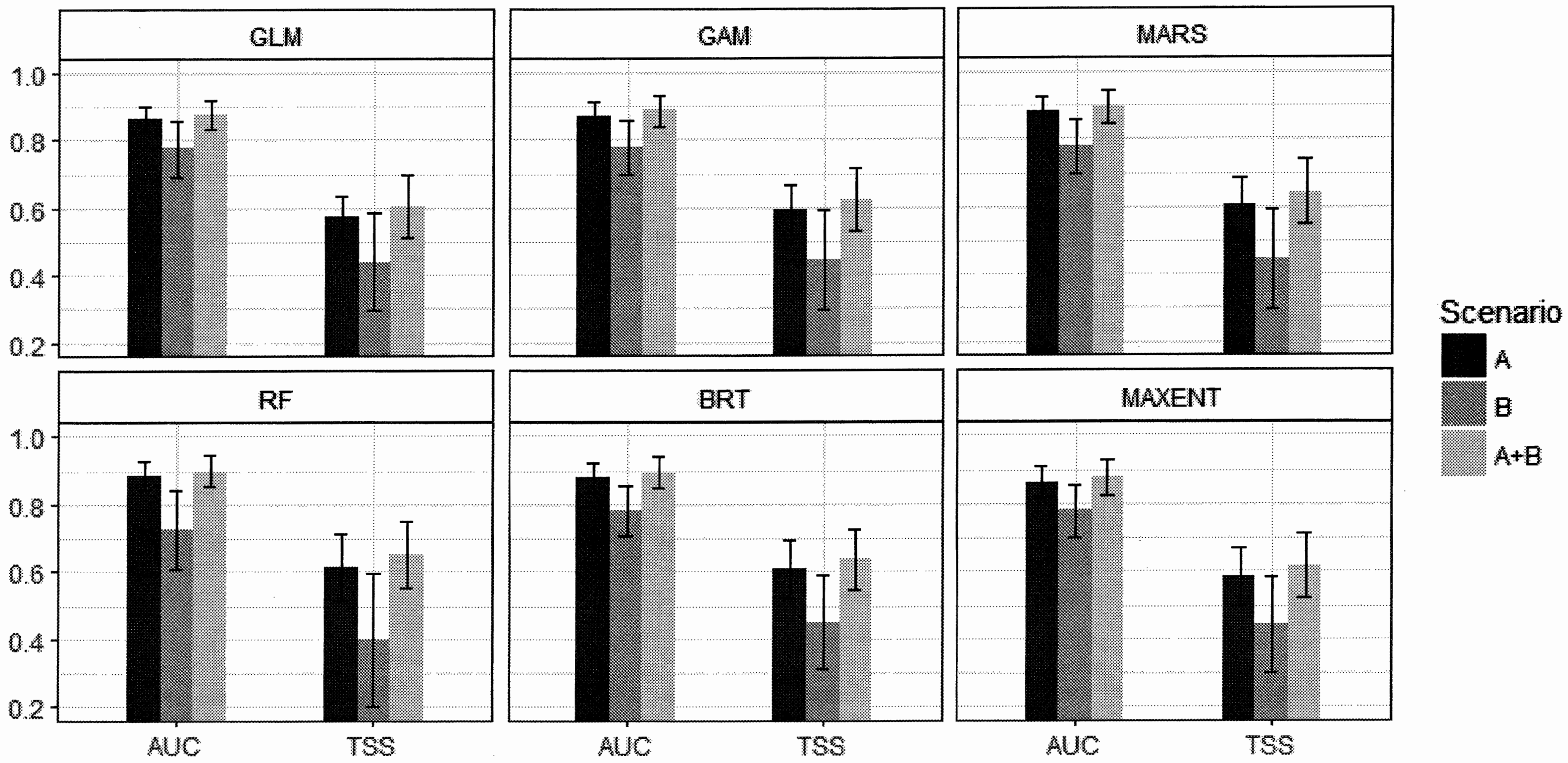

Figure 2.5: Model evaluation assessment through AUC and TSS across all species' models. Computations were done for (A) models including only abiotic variables, (B) models using only biotic information, and (A+B) models incorporating both abiotic and biotic variables. Models' labels provided in Figure 2.4.

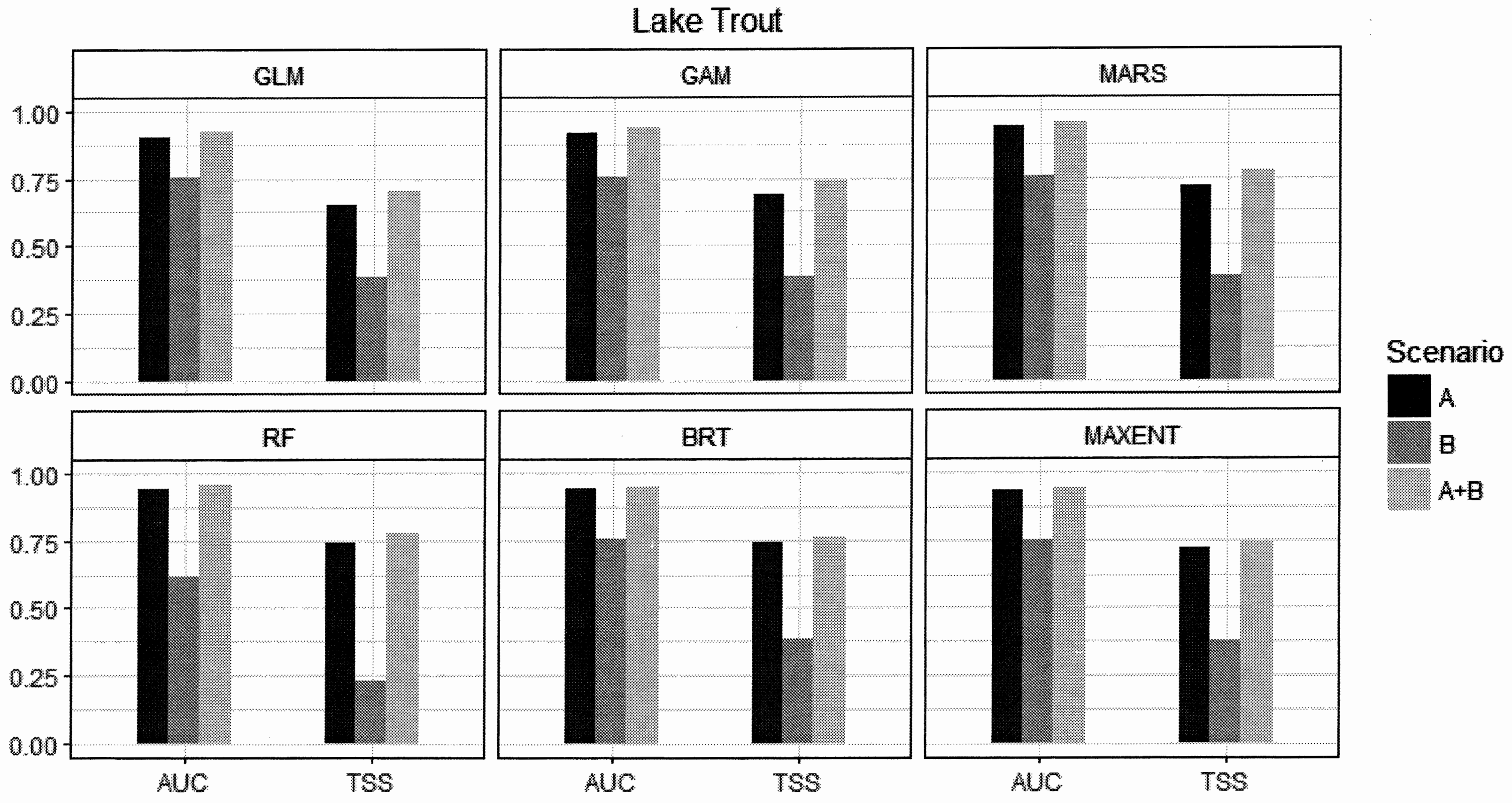

Figure 2.6: Model evaluation assessment through AUC and TSS for lake trout models. Computations were done for (A) models including only abiotic variables, (B) models using only biotic information, and (A+B) models incorporating both abiotic and biotic variables. Models' labels provided in Figure 2.4.

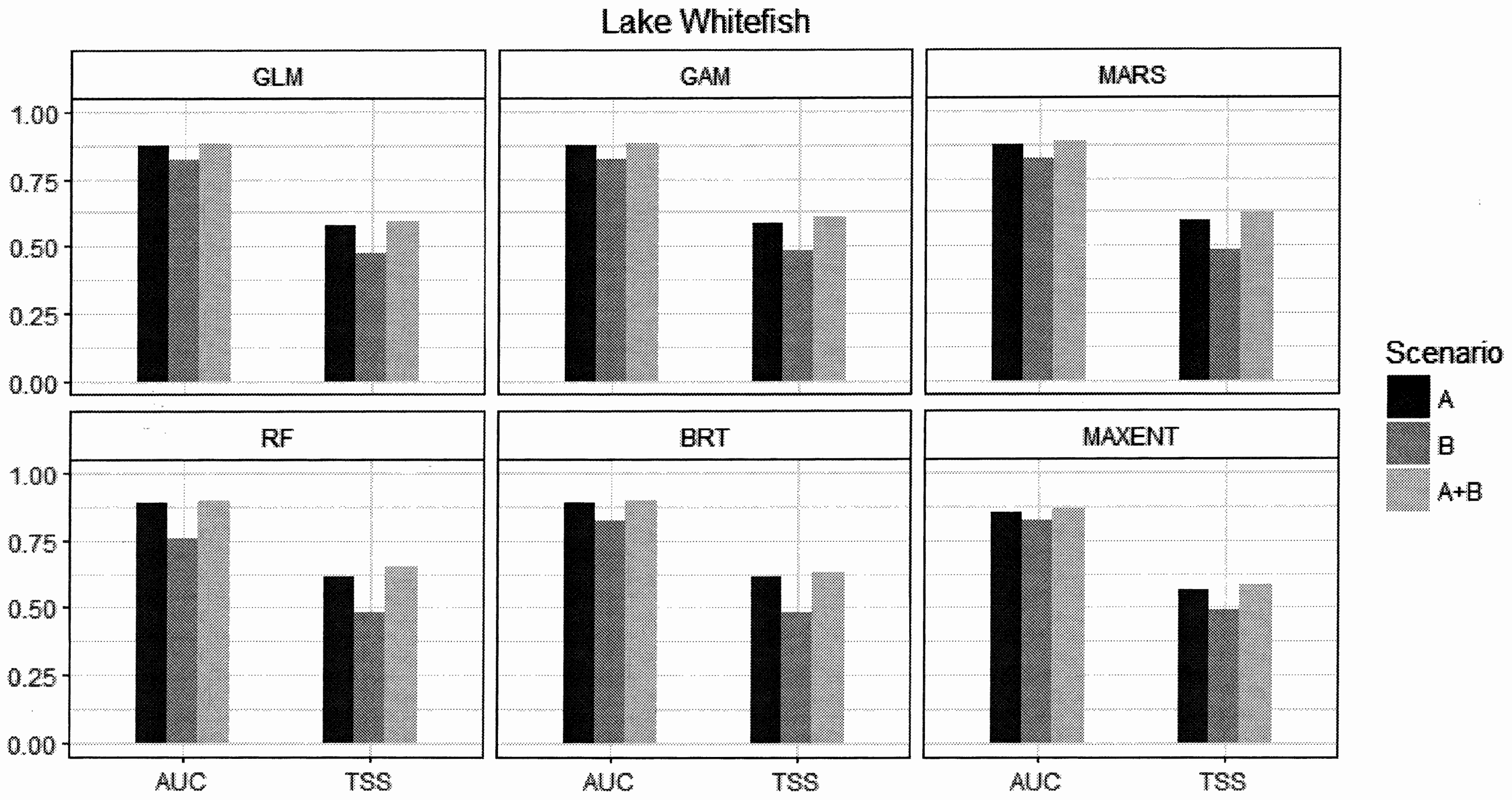

Figure 2.7: Model evaluation assessment through AUC and TSS for lake whitefish models. Computations were done for (A) models including only abiotic variables, (B) models using only biotic information, and (A+B) models incorporating both abiotic and biotic variables. Models' labels provided in Figure 2.4.

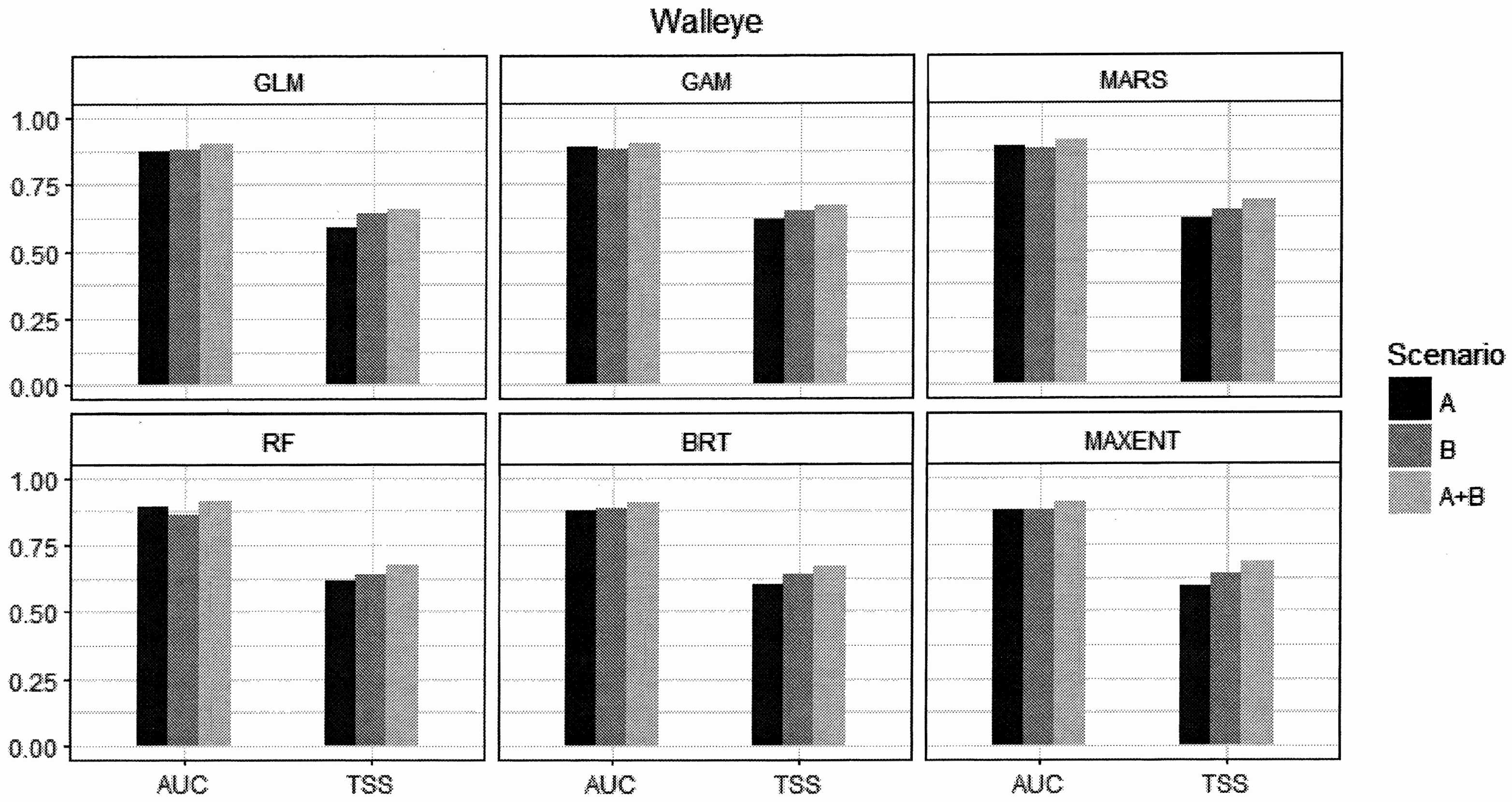

Figure 2.8: Model evaluation assessment through AUC and TSS for walleye models. Computations were done for (A) models including only abiotic variables, (B) models using only biotic information, and (A+B) models incorporating both abiotic and biotic variables. Models' labels provided in Figure 2.4.

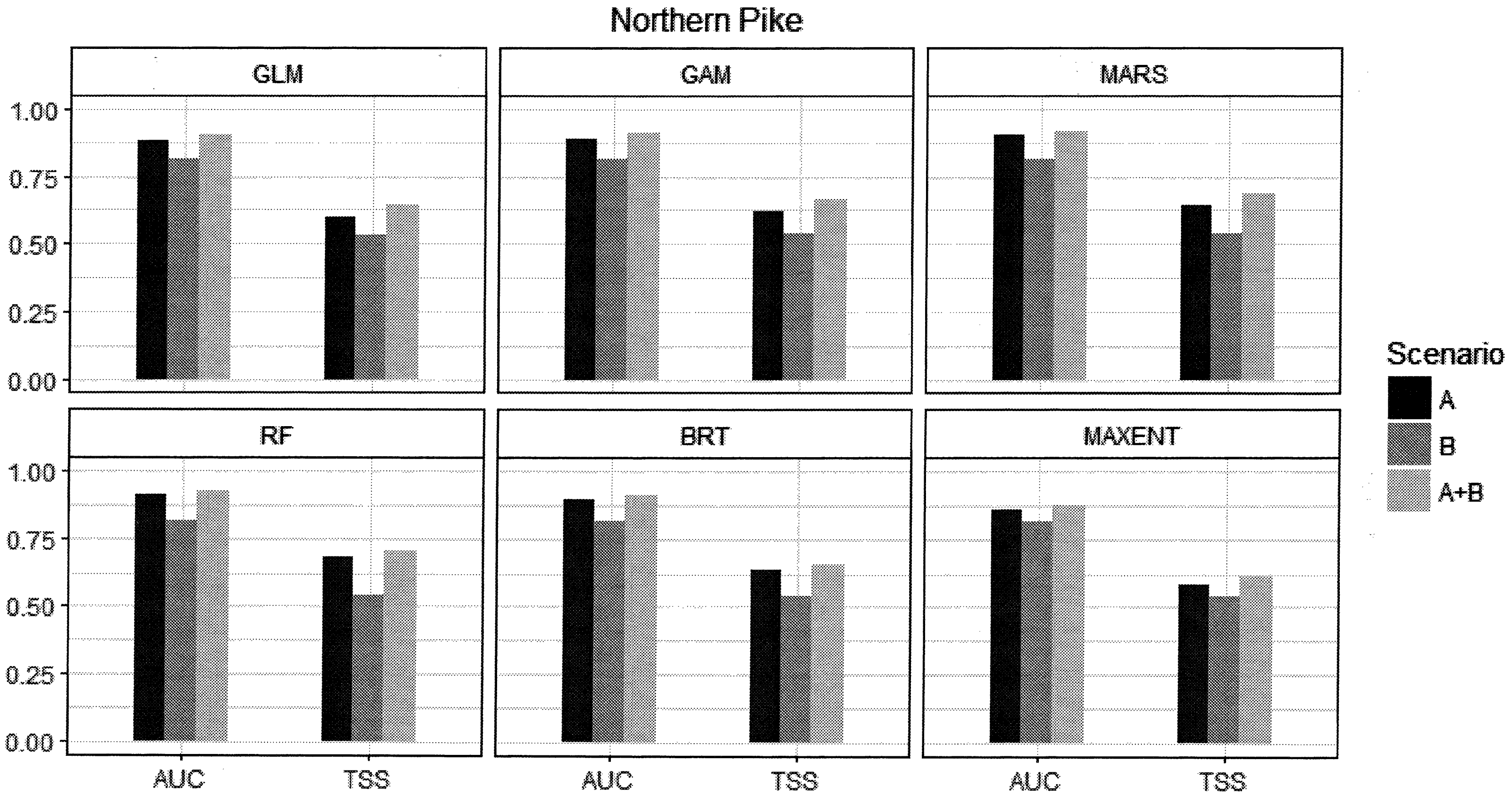

Figure 2.9: Model evaluation assessment through AUC and TSS for northern pike models. Computations were done for (A) models including only abiotic variables, (B) models using only biotic information, and (A+B) models incorporating both abiotic and biotic variables. Models' labels provided in Figure 2.4.

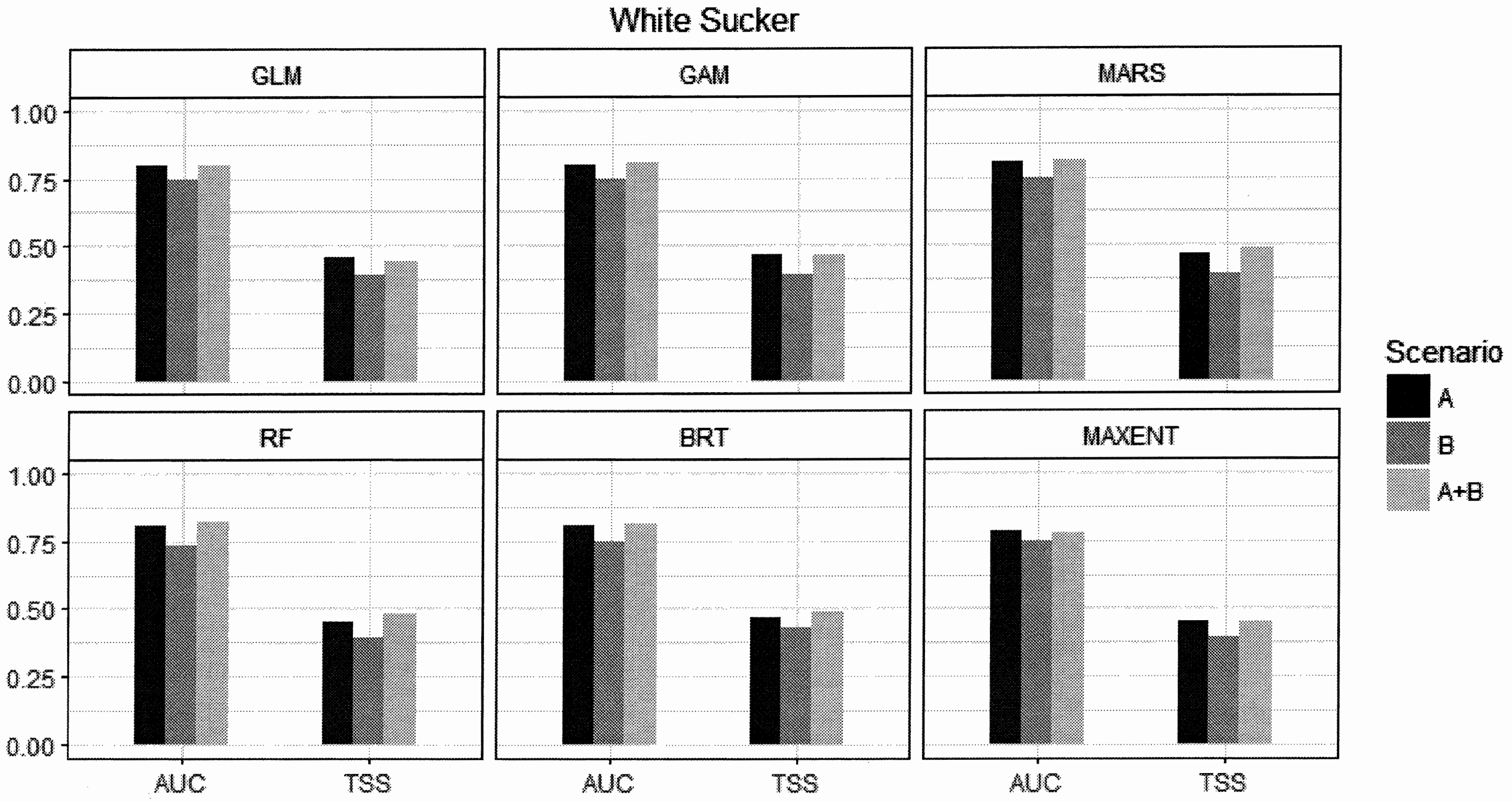

Figure 2.10: Model evaluation assessment through AUC and TSS for white sucker models. Computations were done for (A) models including only abiotic variables, (B) models using only biotic information, and (A+B) models incorporating both abiotic and biotic variables. Models' labels provided in Figure 2.4.

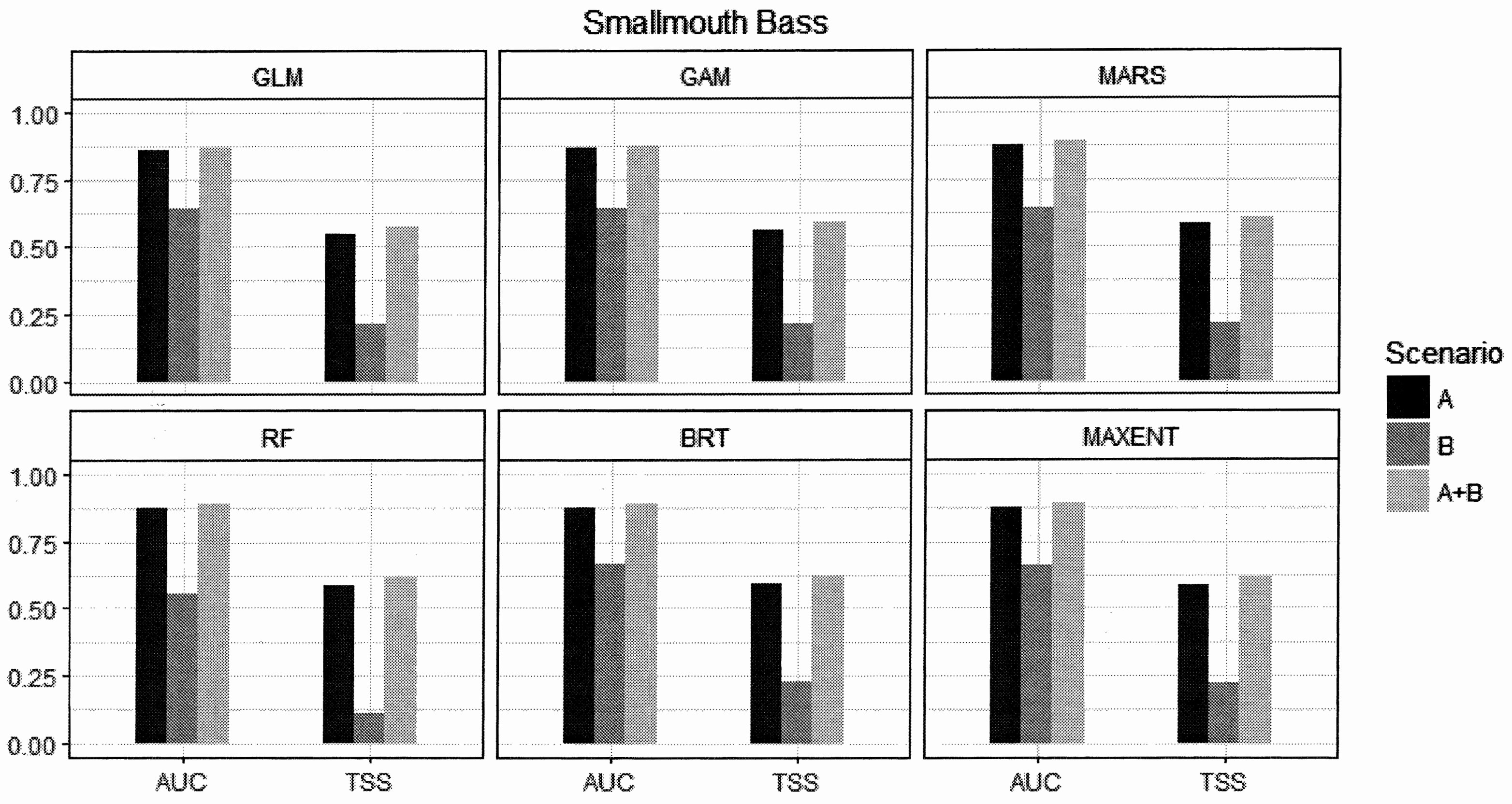

Figure 2.11: Model evaluation assessment through AUC and TSS for smallmouth bass models. Computations were done for (A) models including only abiotic variables, (B) models using only biotic information, and (A+B) models incorporating both abiotic and biotic variables. Models' labels provided in Figure 2.4.

2.4.3 Patterns of co-occurrence in freshwater fish species, abiotic parameters and community composition

Tables 2.4-2.6 summarise the relationships among the six freshwater fish species and the variables used as predictors, across all models and considering the three classes of models. The relationships were drawn from all species distribution models' results by verifying the type of influence (positive or negative) the different variables had on fish occurrences. (A) models showed that mean annual air temperature has a negative influence on cold and cool water species presence, while the opposite is found for smallmouth bass which represents a warm water species (Table 2.4 and 2.6). DD5 was negatively correlated with cold water species' presence, but positive on warm water species and cool water species such as walleye and northern pike. Warmer temperatures clearly has a positive impact on smallmouth bass since both climate variables showed to influence the presence of this species.

Among the lake morphometry predictors, lake size (Log10Area) was the only variable influencing positively all species' presences, and this result can be seen for both (A) and (A+B) models. Additionally, these two classes of models also suggested that mean and max depth have a positive influence on lake trout presence, while shoreline development factor (SDF) seems to have a negative influence on all species. The effect of water chemistry parameters on species' presence showed slightly different results from (A) models to (A+B) models, more specifically, in (A) models (Table 2.4), secchi influences the presence of lake trout and smallmouth bass in a positive way, while in (A+B) models (Table 2.6) it also influences lake whitefish presence. Those result suggests that the incorporation of biotic information through community composition (e.g. other fish species' presences) reveals trends not seen when modelling using abiotic variables alone. Finally, (A) and (A+B) model results suggested that water chemistry parameters such as TDS influences positively walleye and northern pike presence, while pH has a negative impact on all freshwater fish species.

Table 2.5 results suggest that a given freshwater fish species' presence influences positively other species within the same thermal guild. For instance, lake trout and lake whitefish have a positive effect on each other's presence. The same trend can be seen regarding walleye, northern pike and white sucker. These results still hold true when the information about the community composition is incorporated together with abiotic predictors to estimate species' presence in (A+B) models (Table 2.6).

Table 2.4: Relationships' directionality among the six freshwater fish species and the variables used as predictors in (A) models. "+" represents a positive influence of a given variable and "-" represents a negative influence.

Species	(A) Model Predictors								
	DD5	MAT	L10Area	SDF	Depth_Max	Depth_Mn	TDS	Secchi	pH
Lake trout	-	-	+	-	+	+	-	+	-
Lake whitefish	-	-	+	-	-	-	-	-	-
Walleye	+	-	+	-	-	-	+	-	-
Northern pike	+	-	+	-	-	-	+	-	-
White sucker	-	-	+	-	-	-	-	-	-
Smallmouth bass	+	+	+	-	-	-	-	+	-

Table 2.5: Relationships' directionality among the six freshwater fish species and the variables used as predictors in (B) models. "+" represents a positive influence of a given variable and "-" represents a negative influence. P_LaTro: presence of lake trout; P_LaWhi: presence of lake whitefish; P_Walle: presence of walleye; P_NoPik: presence of northern pike; P_WhSuc: presence of white sucker; P_SmBas: presence of smallmouth bass.

Species	(B) Model Predictors					
	P_Latro	P_LaWhi	P_Walle	P_NoPik	P_WhSuc	P_SmBas
Lake trout		+	-	-	+	+
Lake whitefish	+		+	+	+	-
Walleye	-	+		+	+	+
Northern pike	-	+	+		+	-
White sucker	+	+	+	+		+
Smallmouth bass	+	-	+	-	+	

Table 2.6: Relationships' directionality among the six freshwater fish species and the variables used as predictors in (A+B) models. "+" represents a positive influence of a given variable and "-" represents a negative influence. P_LaTro: presence of lake trout; P_LaWhi: presence of lake whitefish; P_Walle: presence of walleye; P_NoPik: presence of northern pike; P_WhSuc: presence of white sucker; P_SmBas: presence of smallmouth bass.

Species	(A+B) Model Predictors														
	DD5	MAT	L10Area	SDF	Depth_ Max	Depth_ Mn	TDS	Secchi	pH	P_Latro	P_LaWhi	P_Walle	P_NoPik	P_WhSuc	P_SmBas
Lake trout	-	-	+	-	+	+	-	+	-		+	-	-	+	+
Lake whitefish	-	-	+	-	-	-	-	+	-	+		+	+	+	+
Walleye	+	-	+	-	-	-	+	-	-	-	+		+	+	+
Northern pike	+	-	+	-	-	-	+	-	-	-	+	+		+	+
White sucker	-	-	+	-	-	-	-	-	-	+	-	+	+		+
Smallmouth bass	+	+	+	-	-	-	-	+	-	+	-	+	+	+	

2.4.4 Models' agreement and variable importance assessment

Figures 2.12-2.17 show results related to how the six species distribution models agree in terms of their ranked variable importance, for each class of models (A, B, A+B) and freshwater fish species. Results reveal that the way different SDM approaches rank variables across species and with increased complexity (A and B models are less complex than A+B ones) is generally quite consistent, which makes them reliable modelling choices in terms of stability and reduced uncertainty. Overall, (B) models showed a higher degree of agreement, with the walleye (B) models displaying more similarity among them, which means that the predictor variables were ranked in importance in a similar way across all (B) models for this fish species. Among all species, northern pike models were considerably consistent in terms of variable importance ranking across all models and classes of models.

The chosen SDM methods can be classified broadly as "regression methods" (encompassing GLM, GAM and MARS) and "machine learning methods" (RF, BRT, MAXENT). It was expected to see a higher degree of agreement among methods in each of these two classifications, but here results were quite mixed across all freshwater fish species and classes of models. In addition, it was also expected to obtain a decreased level of agreement with increasing model complexity (e.g. A+B models), however results across species regarding this aspect are mixed as well. For instance, lake trout, walleye, white sucker and smallmouth bass models showed an improvement in agreement level from (A) models to (A+B) models, while lake whitefish and northern pike models showed a slight decrease in agreement.

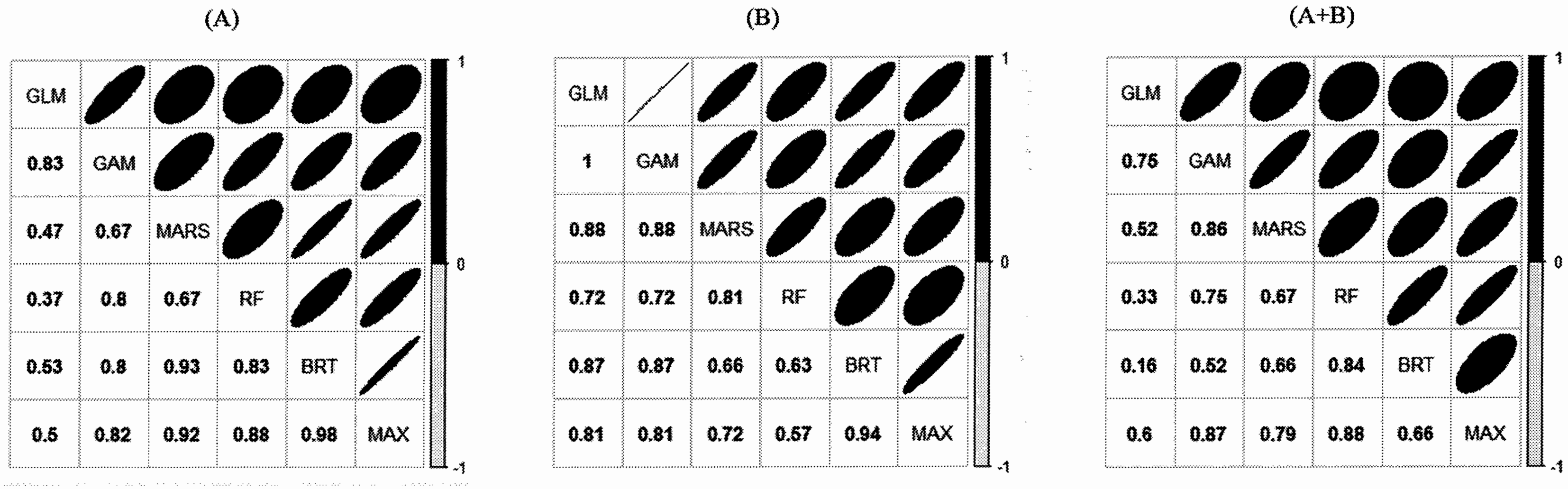

Figure 2.12: Model agreement assessment for lake trout, obtained for (A) models including only abiotic variables, (B) models including only biotic information, and (A+B) models incorporating both abiotic and biotic parameters. GLM: generalized linear model based on Poisson distribution; GAM: generalized additive model; MARS: multivariate adaptive regression splines; RF: random forests; BRT: boosted regression trees; MAX: MAXENT.

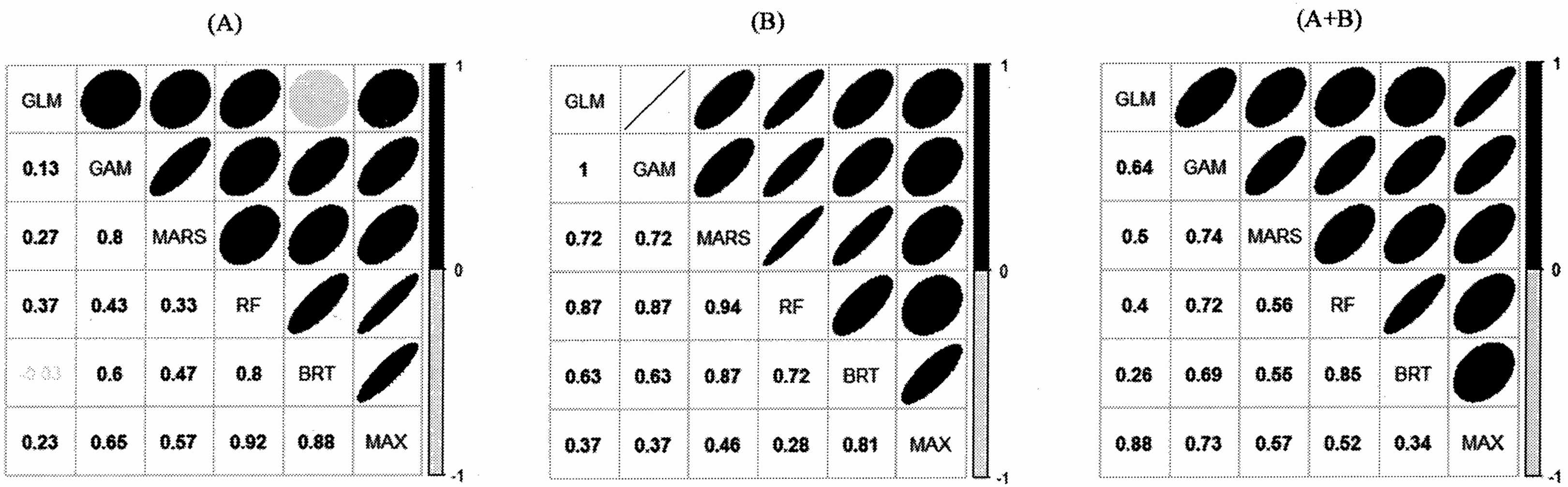

Figure 2.13: Model agreement assessment for lake whitefish, obtained for (A) models including only abiotic variables, (B) models including only biotic information, and (A+B) models incorporating both abiotic and biotic parameters. Models' labels provided in Figure 2.12.

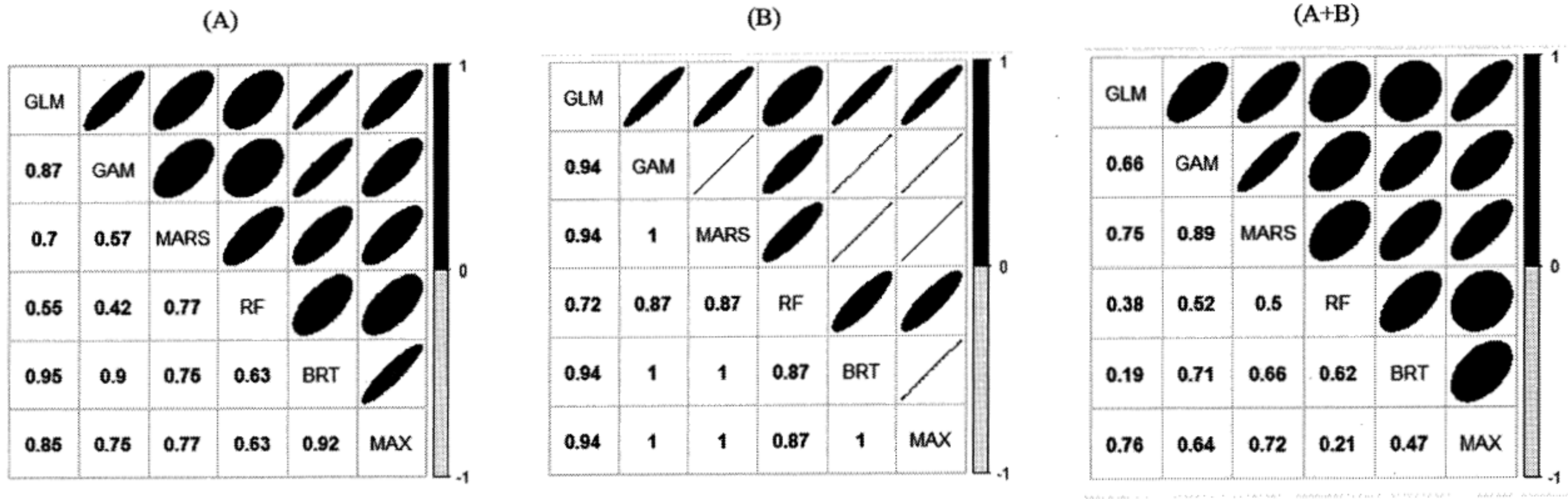

Figure 2.14: Model agreement assessment for walleye, obtained for (A) models including only abiotic variables, (B) models including only biotic information, and (A+B) models incorporating both abiotic and biotic parameters. Models' labels provided in Figure 2.12.

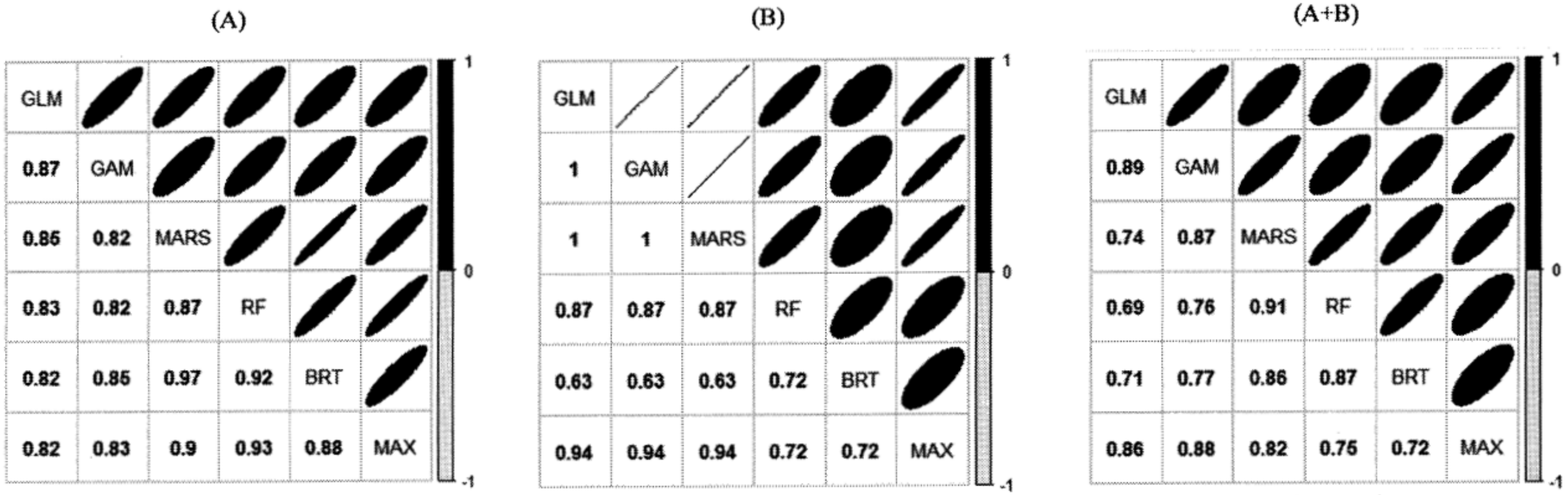

Figure 2.15: Model agreement assessment for northern pike, obtained for (A) models including only abiotic variables, (B) models including only biotic information, and (A+B) models incorporating both abiotic and biotic parameters. Models' labels provided in Figure 2.12.

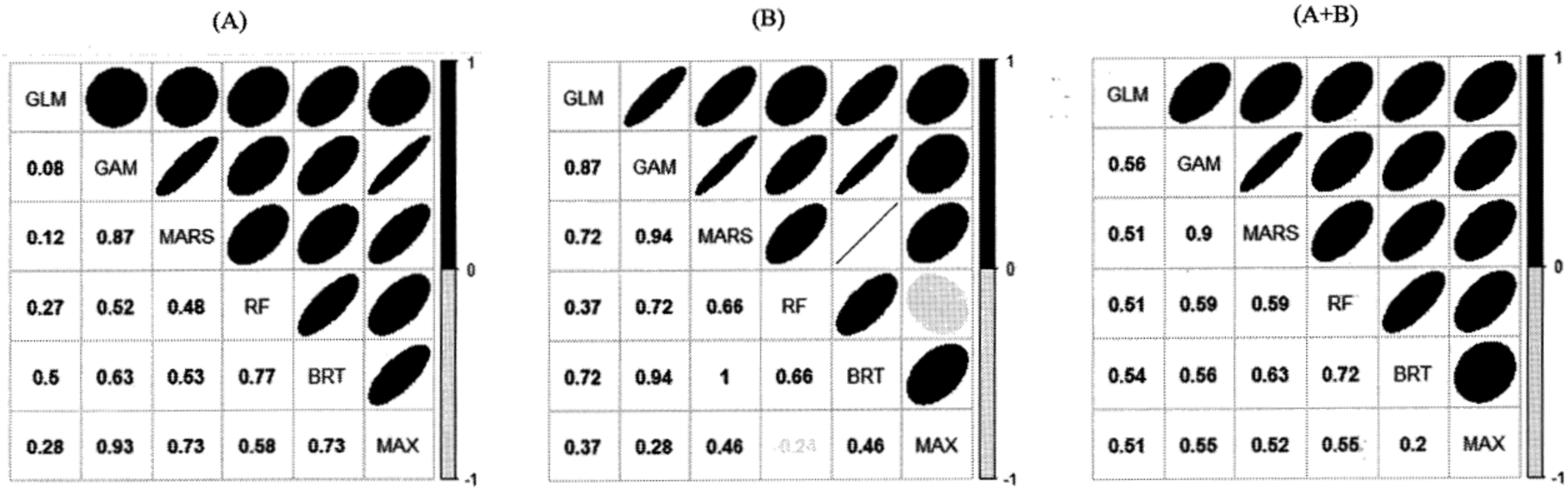

Figure 2.16: Model agreement assessment for white sucker, obtained for (A) models including only abiotic variables, (B) models including only biotic information, and (A+B) models incorporating both abiotic and biotic parameters. Models' labels provided in Figure 2.12.

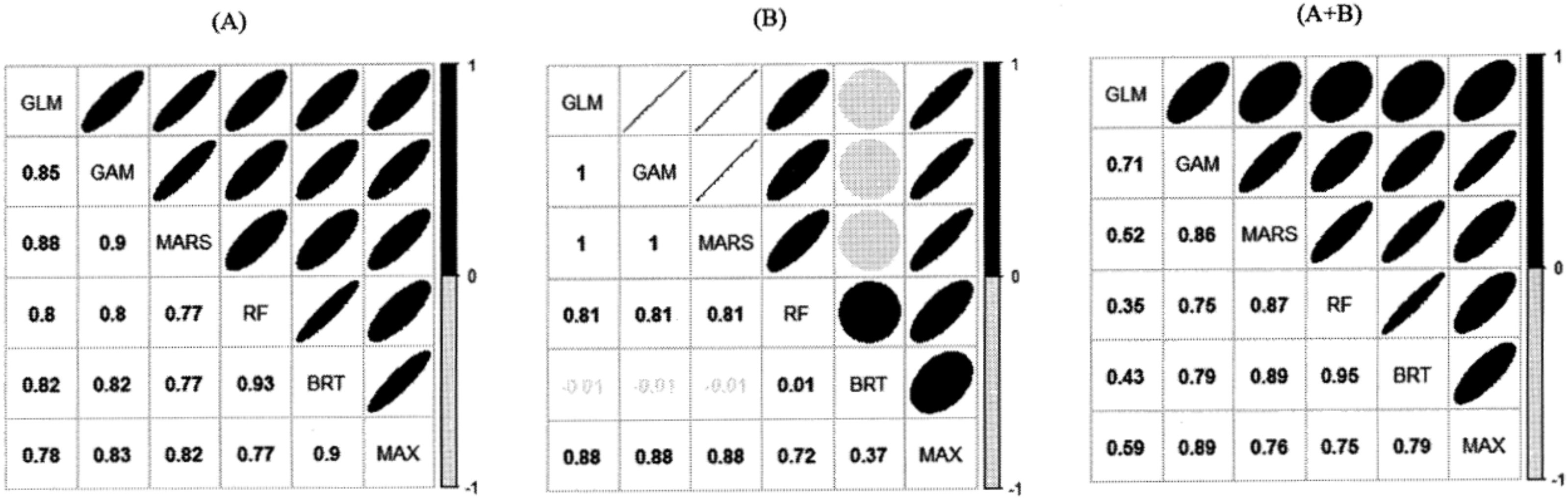

Figure 2.17: Model agreement assessment for smallmouth bass, obtained for (A) models including only abiotic variables, (B) models including only biotic information, and (A+B) models incorporating both abiotic and biotic parameters. Models' labels provided in Figure 2.12.

Figures 2.18 and 2.19 present the ranked variable importance estimated for (A) and (B) models built to estimate freshwater fish occurrences, computed by averaging across all six freshwater fish species. Results related to (A+B) models were not shown for the sake of simplicity, since the objective here is to assess how predictors were ranked according to different SDM approaches. There were a number of similarities and differences in relative predictive importance of the abiotic and biotic variables among modelling approaches. The results based on all species for (A) models showed that the modelling approaches showed close agreement in the importance of L10Area, but disagreed in the importance of others for predicting species occurrence. Overall differences in the mean ranked importance of abiotic variables existed among all approaches, but were most notable between regression methods and machine-learning methods. For instance, the importance of MAT and DD5 were higher in machine-learning methods (RF, BRT and MAXENT) compared to regression methods (GLM, GAM and MARS). Similar observations can be done regarding the results based on all species for (B) models (Figure 2.19).

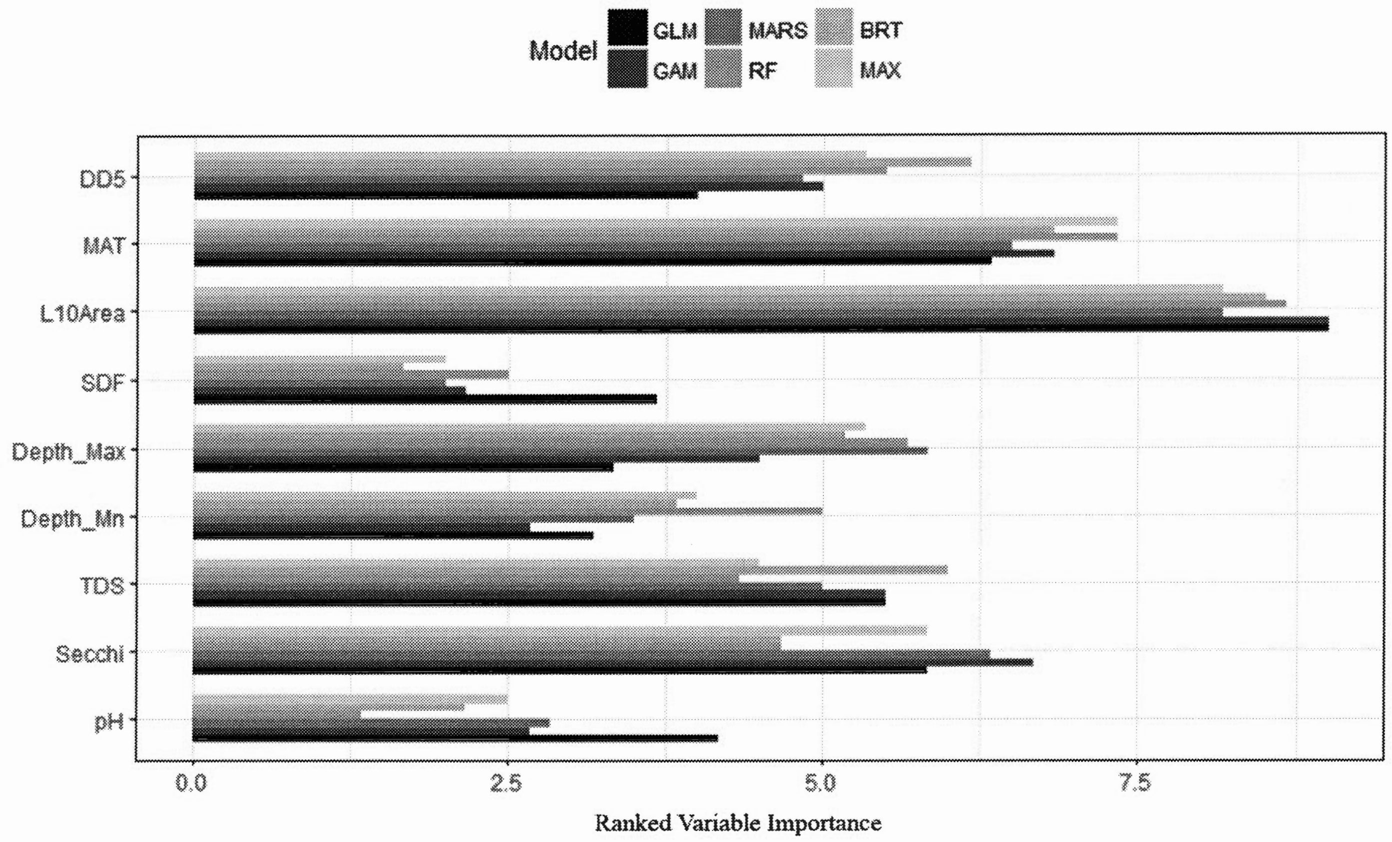

Figure 2.18: Ranked variable importance obtained for (A) models including only abiotic variables. Variable importance was computed as an average across all species. GLM: generalized linear model based on Poisson distribution; GAM: generalized additive model; MARS: multivariate adaptive regression splines; RF: random forests; BRT: boosted regression trees; MAX: MAXENT.

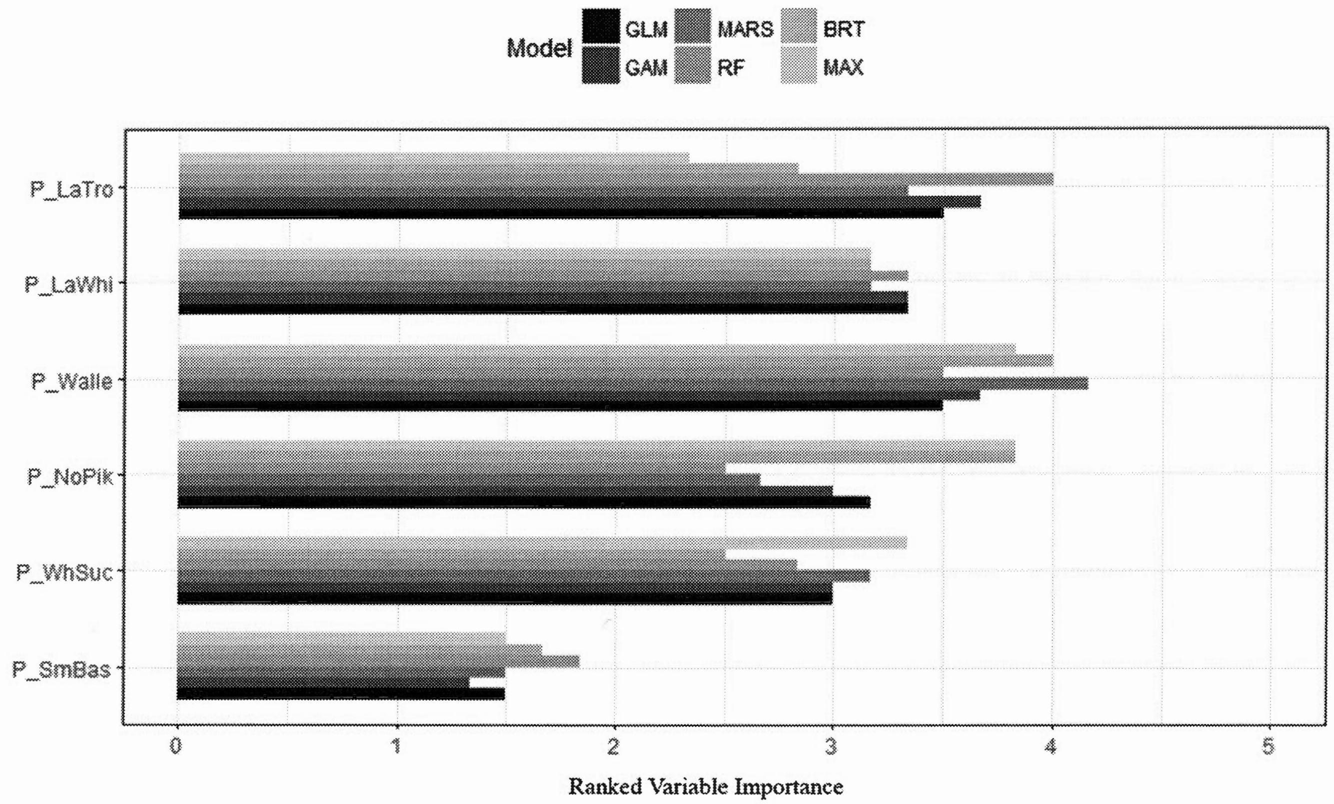

Figure 2.19: Ranked variable importance obtained for (B) models including only biotic variables. Variable importance was computed as an average across all species. Models' labels provided in Figure 2.18. P_LaTro: presence of lake trout; P_LaWhi: presence of lake whitefish; P_Walle: presence of walleye; P_NoPik: presence of northern pike; P_WhSuc: presence of white sucker; P_SmBas: presence of smallmouth bass.

Figures. 2.20-2.25 present the ranked variable importance estimated for each class of models built to estimate freshwater fish occurrences, computed by averaging across all SDM approaches. Models containing only abiotic variables (A) suggest that air temperature (MAT) and surface area of the lake (L10Area) have a major influence on the establishment of smallmouth bass, walleye, northern pike, white sucker and lake whitefish, while maximum depth (Depth_Max) plays an important role on the establishment of lake trout, followed by surface area of the lake.

Results from models containing only biotic information (B) suggest that presence of lake whitefish (P_LaWhi) in lakes plays a role in the distribution of lake trout (Figure 2.20). In addition, the B models also suggested that presence of walleye (P_Walle) influences the establishment of lake whitefish (Figure 2.21), while presence of northern pike (P_NoPik) affected the distribution of walleye (Figure 2.22). Finally, biotic (B) models built for northern pike, white sucker and smallmouth bass suggest that presence of walleye (P_Walle) influences the establishment of northern pike and white sucker (Figures 2.23 and 2.24 respectively), while presence of lake trout (P_LaTro) influences the establishment of smallmouth bass (Figure 2.25).

Models considering both abiotic and biotic variables demonstrate the importance of surface area of the lake (L10Area) on the establishment of all fish species, but also climatic variables and presence of other fish species play a role as seen in the results for these models. According to the A+B models developed for lake trout, its establishment is influenced mainly by lake morphometry parameters such as the aforementioned surface area, maximum and mean depth of the lake, denoting the preference of this species for deep, stratified lakes. For lake whitefish, its A+B models suggest that surface area of the lake is the most important factor affecting its establishment, followed by presence of northern pike, growing degree days above 5°C (DD5) and presence of walleye. Walleye A+B models indicate that its establishment is influenced primarily by presence of northern pike, surface area of the lake, then air temperature (MAT) and presence of lake trout, while northern pike A+B models show that its establishment is mostly affected by air temperature, followed by surface area of the lake, presence of walleye and growing degree days above 5°C. In addition, white sucker A+B models suggest that, besides surface area of the lake, DD5, maximum depth of the lake and presence of lake trout are the main ones influencing its establishment. Finally, smallmouth bass A+B models suggest that MAT, surface area of the lake and presence of walleye are the main ones influencing its distribution.

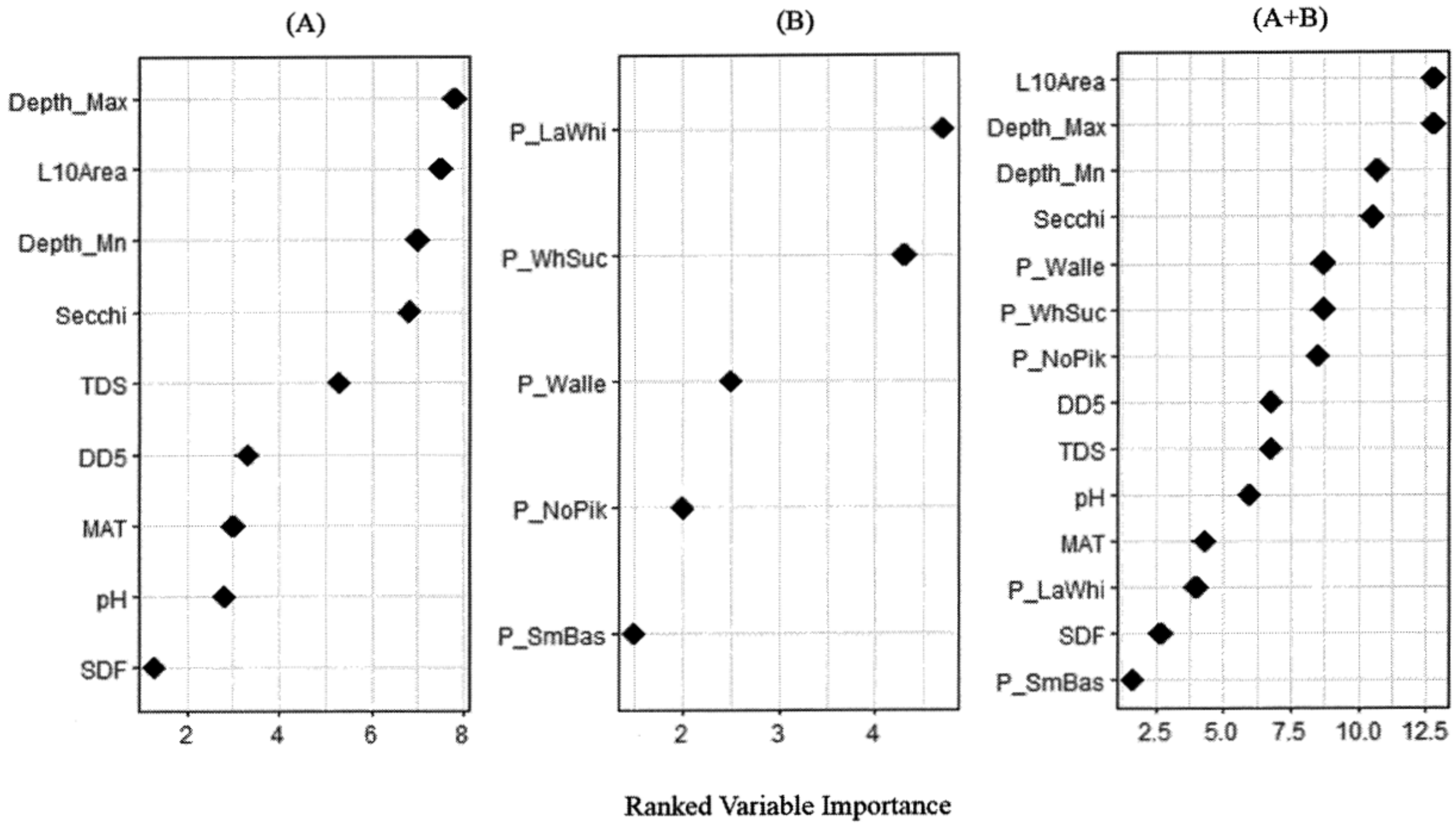

Figure 2.20: Ranked variable importance for lake trout, obtained for (A) models including only abiotic variables, (B) models including only biotic information, and (A+B) models incorporating both abiotic and biotic parameters. Variable importance was computed as an average across all models.

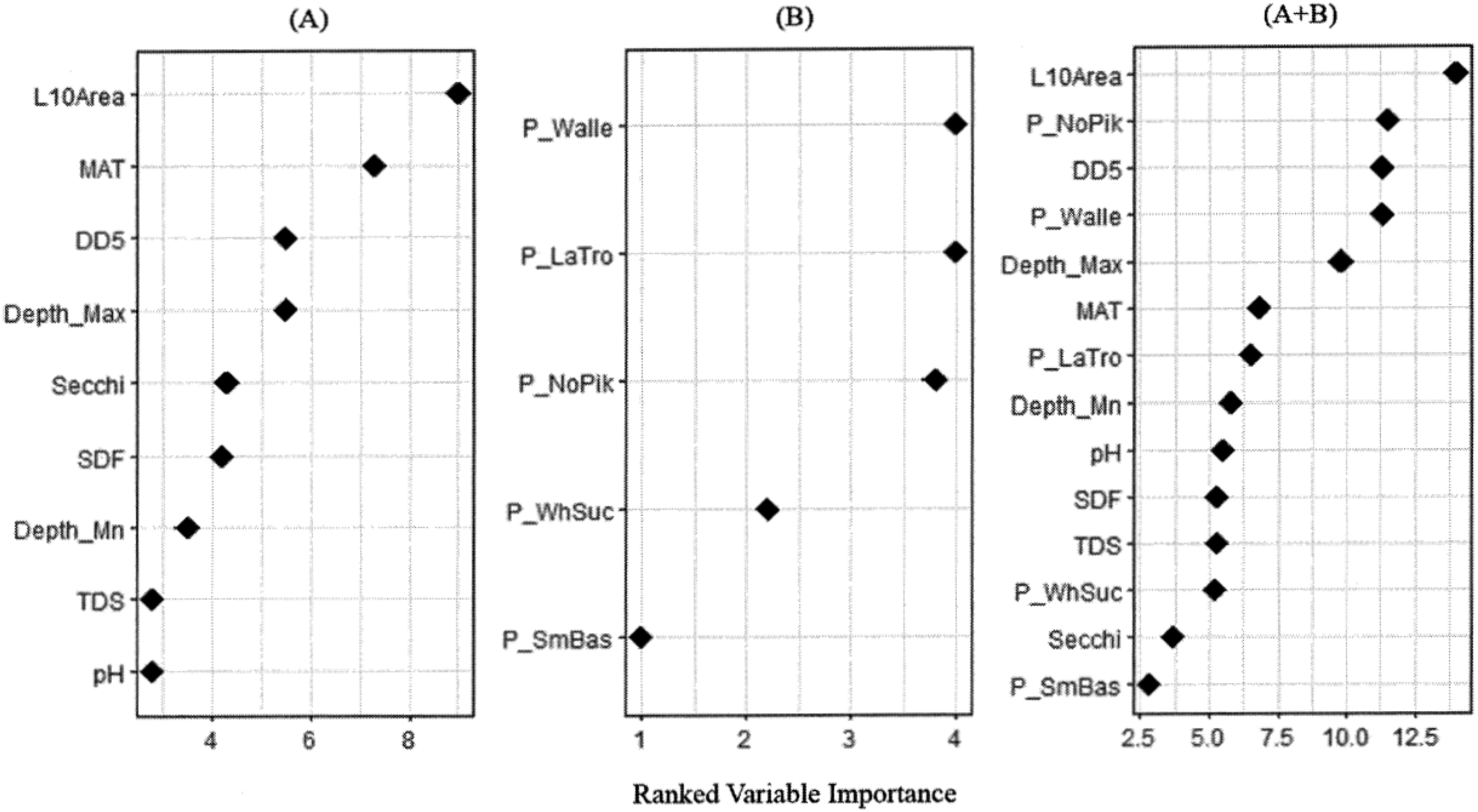

Figure 2.21: Ranked variable importance for lake whitefish, obtained for (A) models including only abiotic variables, (B) models including only biotic information, and (A+B) models incorporating both abiotic and biotic parameters. Variable importance was computed as an average across all models.

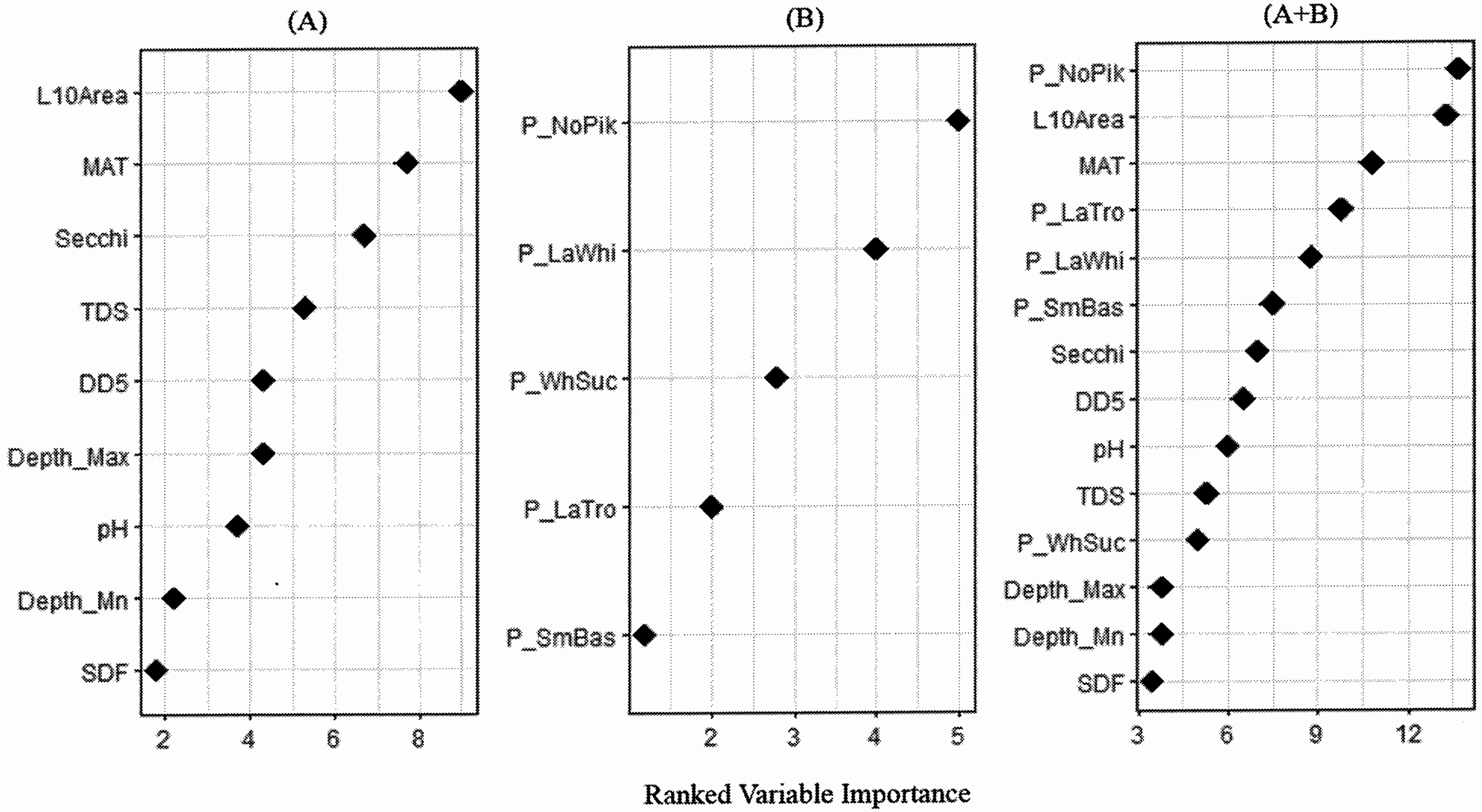

Figure 2.22: Ranked variable importance for walleye, obtained for (A) models including only abiotic variables, (B) models including only biotic information, and (A+B) models incorporating both abiotic and biotic parameters. Variable importance was computed as an average across all models.

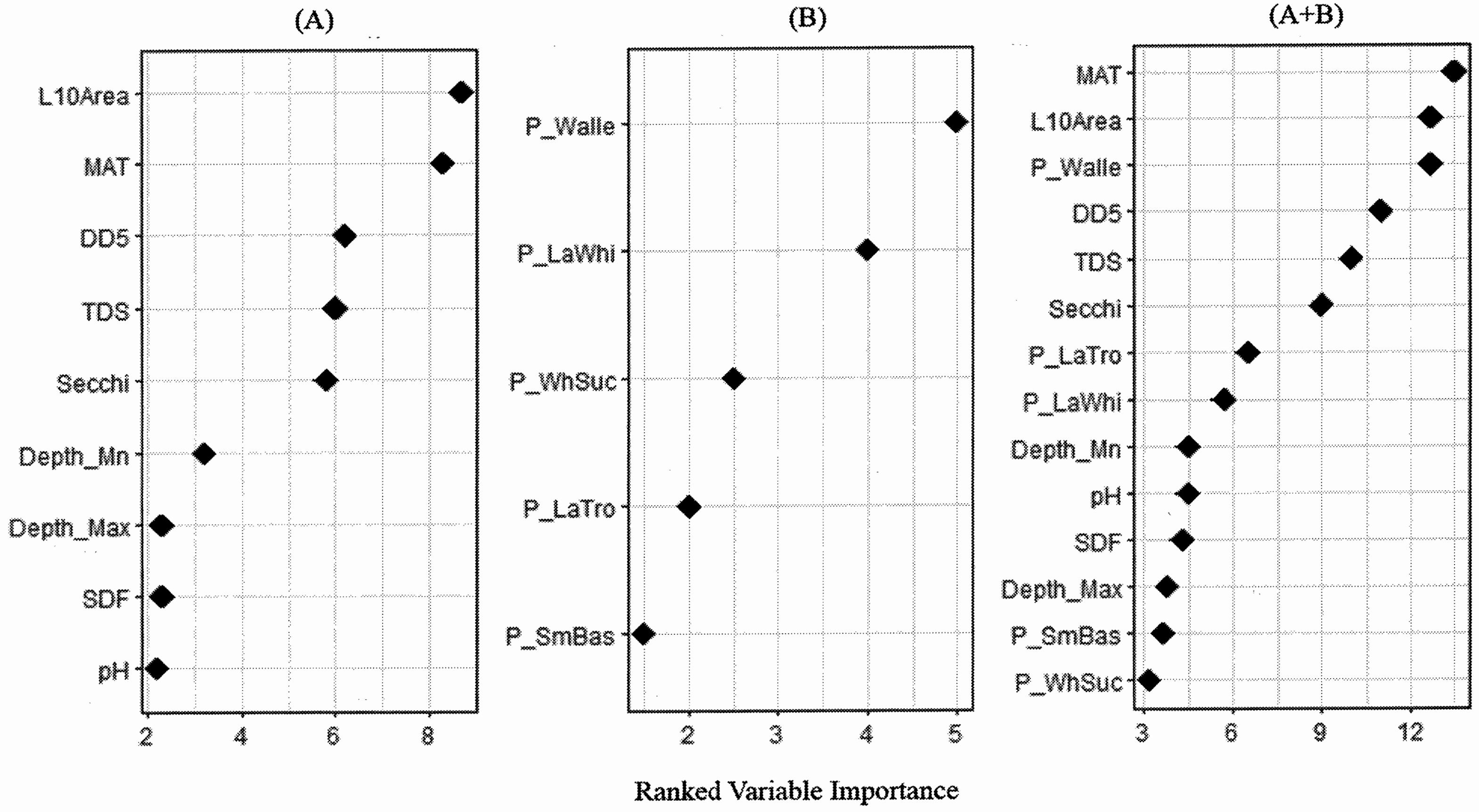

Figure 2.23: Ranked variable importance for northern pike, obtained for (A) models including only abiotic variables, (B) models including only biotic information, and (A+B) models incorporating both abiotic and biotic parameters. Variable importance was computed as an average across all models.

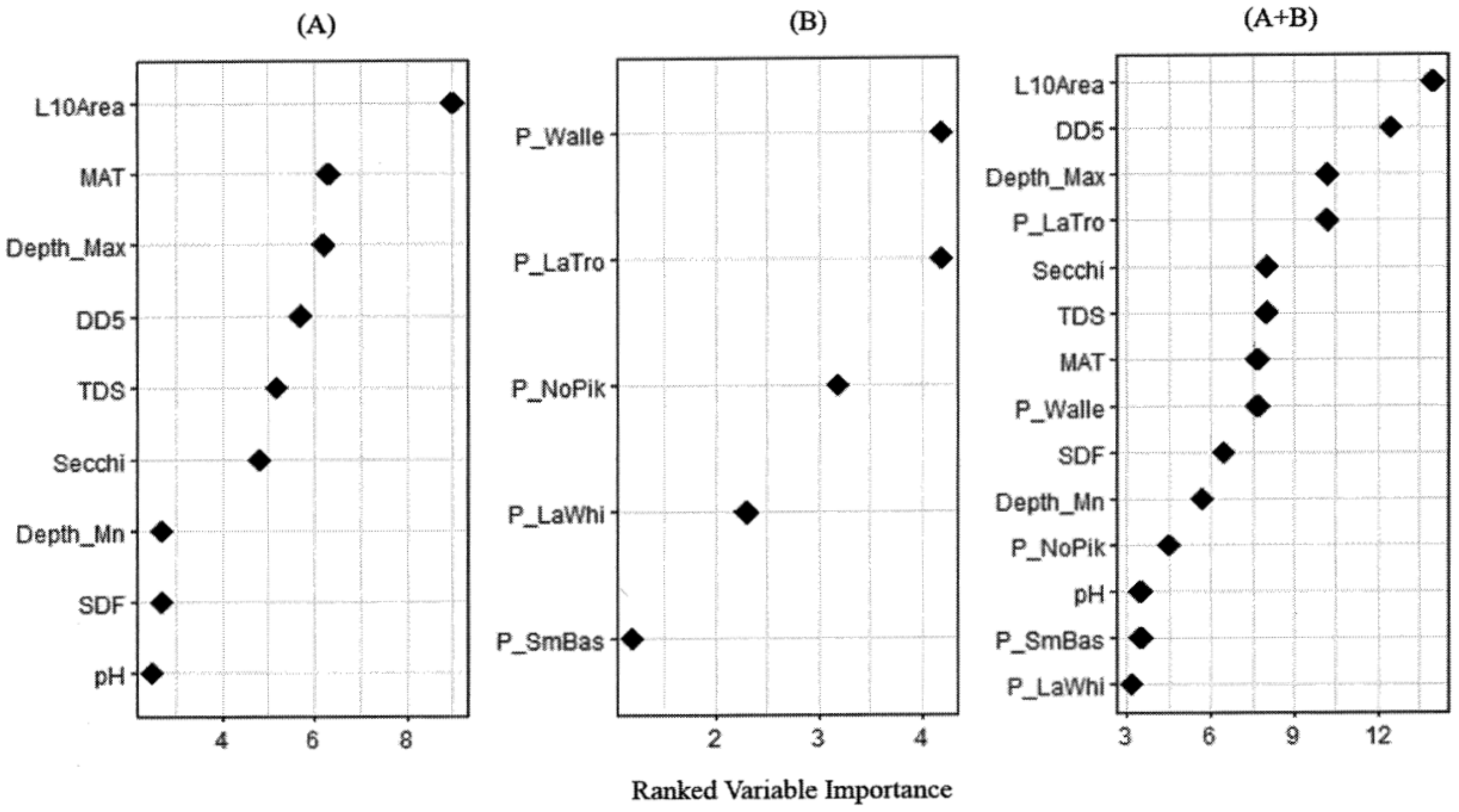

Figure 2.24: Ranked variable importance for white sucker, obtained for (A) models including only abiotic variables, (B) models including only biotic information, and (A+B) models incorporating both abiotic and biotic parameters. Variable importance was computed as an average across all models.

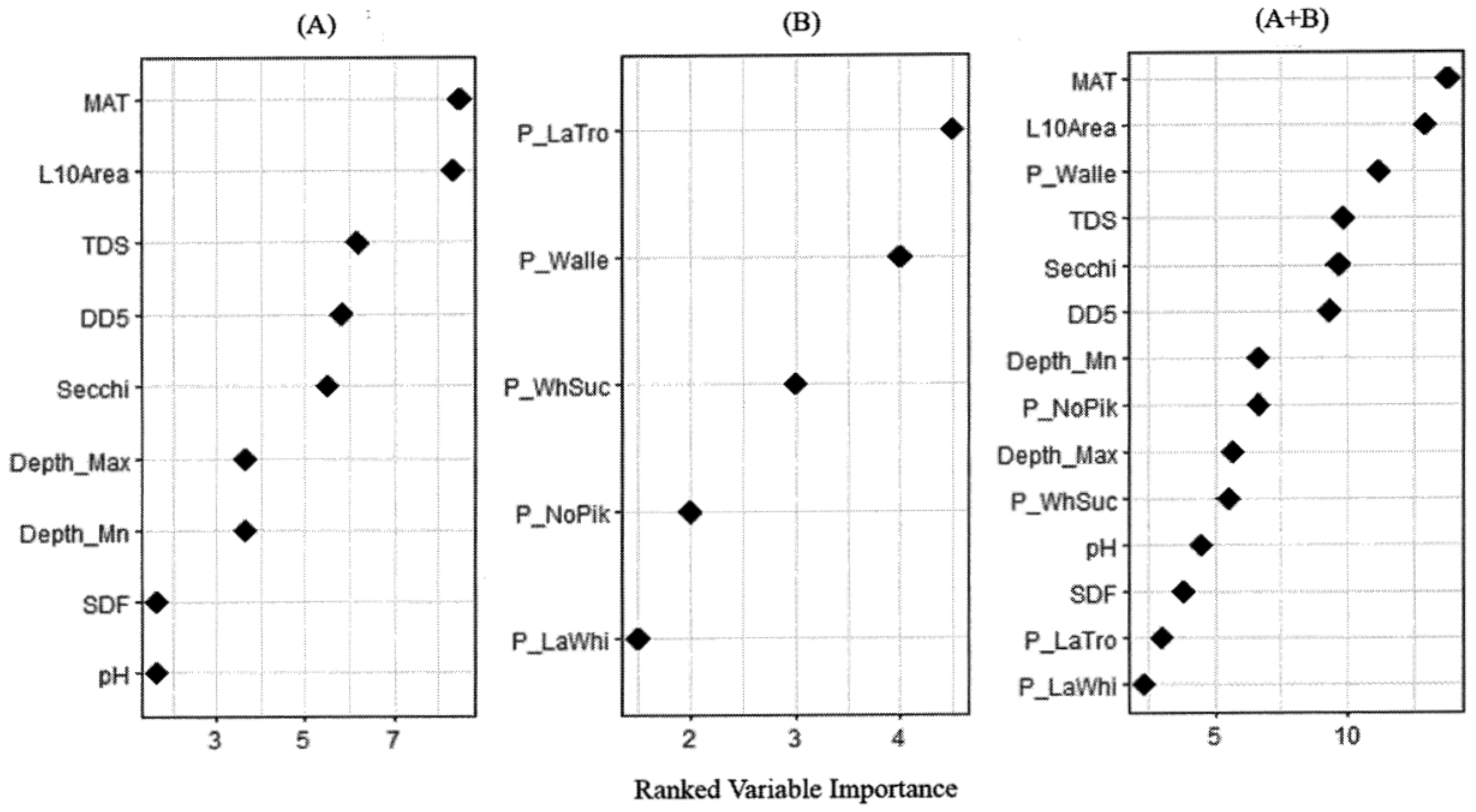

Figure 2.25: Ranked variable importance for smallmouth bass, obtained for (A) models including only abiotic variables, (B) models including only biotic information, and (A+B) models incorporating both abiotic and biotic parameters. Variable importance was computed as an average across all models.

2.5 Discussion

A better understanding of different freshwater fishes' habitat requirements consists of a solid ground for developing models that better describe fish-habitat relationships, given that spurious relationships can also be seen in species distribution models. The lake trout model results seen in Tables 2.4 and 2.6, and Figure 2.20 suggest that lake size, mean depth, max depth and secchi are the most important environmental predictors and influence positively this species' presence. Lake trout occurs only in relatively deep lakes through the southern part of the Canadian range, but in the northern half, specially in the Territories, it occurs also in shallow lakes and rivers (Scott and Crossman 1998). In inland lakes in southern Canada the depth-distribution of lake trout varies with the seasons. In autumn, usually early October, they move into rocky shallows in preparation for spawning. Spawning most often occurs over a large boulder or rubble bottom in inland lakes at depths of less than 12 m and sometimes as shallow as 30 cm. Temperatures at time of spawning have been reported at 8°-12°C (which also defines their fundamental thermal niche according to Magnuson et al. 1990), but spawning is not initiated by temperature alone, but also by light, since the spawning act takes place after dark (Scott and Crossman 1998). After spawning is completed they disperse freely throughout the lake at various depths and remain dispersed throughout the winter months. Investigations by O'Connor et al. (1981), as well as results of a literature search completed by Christie (1986), suggest that lake trout growth is optimal within its fundamental thermal niche, confirming the importance of temperature to the distribution of this species. In spring they often occur in surface waters immediately after breakup of ice. As surface waters warm with the advance of spring, lake trout retire to the cooler waters, eventually retreating to the hypolimnion, below the thermocline during the warmer summer months. The depth of the thermocline in a particular lake depends on many factors such as latitude, size of lake and height of surrounding land, so these are factors that are also important in determining lake trout distribution. Water clarity (here represented by secchi) is a proxy for overall lake productivity as well as visibility conditions, which is important because fish species vary in terms of their visual light sensitivities and reliance on sight for detection of food resources (Scott and Crossman 1998). Olden and Jackson (2001) proposed lake trout occurrence models based on artificial neural networks, logistic regression, linear discriminant analysis and classification trees, and their results indicated that surface area of the lake, total shoreline perimeter, lake volume and maximum depth were the main predictors of lake trout presence. A second study from the same authors using solely artificial neural networks as modelling approach suggested that elevation is also an important predictor. Some degree of agreement can be seen between the relationships presented here and the ones showed by Olden and Jackson, however, secchi was not considered in their models and this abiotic parameter has a considerable relevance in describing lake trout distributions, since it has a direct relationship with productivity levels at the benthic zone (Finstad et al 2014), i.e. food availability for lake trout.

Lake whitefish (A) and (A+B) models results also suggest that lake size is an important predictor of this species, having a positive influence on its occurrence (Tables 2.4 and 2.6, Figure 2.21). Figure 2.21 also indicate that climate variables (DD5 and MAT) and max depth are important predictors, however their influence (Tables 2.4 and 2.6) are negative ones. The negative influence of climate on cool water species habitat is a known issue already reported by the literature (Mackenzie-Grieve and Post 2006; Sharma et al. 2009; Callaghan et al. 2016), since the increasing temperatures cause a reduction in optimal habitat availability for this thermal guild. Since this species is classified as a cool water species, it was expected a positive influence of mean and maximum depth of the lake with this species occurrence. Lake whitefish descends into the cooler waters of the hypolimnion during summer months, over most of the southern part of its range, and only large, stratified lakes can provide ideal habitat for this species. Throughout most of the Great Lakes, these fish move from deep to shoal waters in early spring and back to deeper water as warming occurs. In the fall, lake whitefish move into shallow water to spawn, looking for hard or stone bottom but sometimes sand to lay its eggs (Sawatzky et al. 2007). Since lake whitefish are bottom feeders, benthic production is essential for this species due to food availability (like mentioned before for lake trout). The negative influences of mean and maximum depth on lake whitefish occurrences suggest that most lakes where this species was present were shallow ones, and this is likely due to: (i) greater food availability for this species in these lakes; (ii) the species mobility (Scott and Crossman 1998). Previous work by Olden and Jackson (2002) showed lake morphometry predictors such as surface area of the lake, lake volume, total shoreline perimeter and altitude to be important for lake whitefish occurrence, however the authors did not point out the type of influences these predictors have on lake whitefish occurrences. Here results pointed out the importance of climate to this species. Edwards et al. (2016) in their work using logistic regression and the same historical dataset used here, showed that a climate index variable (composed by ice-free duration, mean July air temperature, minimum January air temperature, and total annual precipitation), mean to maximum depth ratio, secchi, maximum depth and lake area were influential predictors of lake whitefish occurrence. Even though the climate variables used in their work are not similar to the ones chosen here, the influence of climate on occurrence was also seen to be negative. On the other hand, the influence of maximum depth presented by their lake whitefish occurrence model was positive, which does not agree with the results seen here. This lack of agreement suggests that the modelling method chosen by the authors had an impact on the relationships showcased by the model, since all six methods presented here agreed on the negative relationship of max depth with lake whitefish occurrence.

For walleye, (A) and (A+B) models results (Tables 2.4 and 2.6, Figure 2.22) revealed that lake size is also important, together with climate (especially MAT) and water chemistry parameters such as secchi and TDS. Lake size influence on walleye occurrence is positive (like seen for cool water species), so is the influence of TDS, while MAT and secchi showed to influence negatively the occurrence of this species. Walleye are tolerant of a great range of environmental situations, but appear to reach greatest abundance in large, shallow, turbid lakes (Scott and Crossman 1998). Optimum transparency in a shallow lake, which will allow walleye to feed intermittently throughout the day, is in the order of 1-2 m secchi, however in clear lakes, where they often lie in contact with the bottom apparently "sleeping", adult and subadult walleye are primarily crepuscular or nocturnal feeders, feeding from top to bottom (Ali et al. 1977). Movements involve a spring spawning run to shallow coarse-gravel shoals or to tributary rivers (where they seek for rocky areas), daily movements up and down in response to light intensity, and daily or seasonal movements in response to temperature or food availability (Scott and Crossman 1998). Model results regarding important abiotic features influencing walleye occurrence are in agreement with the main variables affecting walleye production reported by Lester et al. (2004), where they show that thermal-optical habitat area (TOHA, the benthic area of the lake that supplies optimum light and temperature conditions for walleye during an annual cycle), TDS and water clarity influence walleye habitat availability. In their work, walleye harvest increased in proportion to TOHA times the square root of TDS, and that increases in water clarity seen in the Great Lakes basin have reduced the supply of thermal-optical walleye habitat, thus having negative effects on walleye production. Chu et al. (2005) showed mean annual dew point, growing degree days, total annual precipitation and mean annual hourly wind speed to be significant variables in their walleye occurrence model, with mean annual dew point showing an expressive negative influence. Nate et al. (2000) and Hansen et al. (2015) also reported lake size to be an important abiotic feature in their walleye model. Spring water temperatures, which is a surrogate for regional climate variability (Beard et al. 2003) and/or water levels (Chevalier 1977; Quist et al. 2004) was reported as influential in walleye occurrences (Serns 1982; Hansen et al. 1998; Quist et al. 2003). The inclusion of abiotic variables not considered in this study but shown to be important in predicting walleye occurrence by previous works would consist of a future work perspective, with the objective of seeking a greater improvement of the models built here.

Model results for northern pike evidenced the importance of lake size, climate variables and water chemistry to its occurrence (Figure 2.23), with a positive influence of lake size, mixed influences provided by climate (positive by DD5 and negative by MAT), and water chemistry (positive by TDS and negative by secchi - Tables 2.4 and 2.6). These abiotic features priorities are very similar to the ones seen for walleye, which suggest habitat affinities between these two cool water species, like already reported by the literature (Scott and Crossman 1998; McMahon and Bennett 1996). In Canada, its habitat is usually clear, warm, slow, meandering, heavily vegetated rivers or warm, weedy bays of lakes. They do, however, occur in a wide range of habitat over the whole of their extensive distribution. They generally occur in shallower water in spring and fall but move to deeper cooler water at the height of summer temperatures. In general, northern pike are fairly sedentary, establishing a vague territory where cover and food are adequate. Spawning occurs during daylight hours on the heavily vegetated floodplains of rivers, marshes, and bays of larger lakes (Scott and Crossman 1998; Raat 1988; Casselman and Lewis 1996). Relationships between abiotic variables and northern pike occurrence presented earlier in the literature showed MAT to be an important predictor, e.g.

Schlesinger and Regier (1982) in their work present how this abiotic feature has a positive effect until reaching a certain temperature threshold, which past that starts to influence negatively. Minns and Moore (1992) also showcased the importance of MAT and lake size on northern pike. Neither of these two works showed the importance of water chemistry in determining northern pike distributions, as climate and lake morphometry are commonly shown to be more influential variables in northern pike models. The results presented here represent a contribution to the literature regarding the importance of this environmental feature to northern pike occurrences, and since this species has quite similar habitat requirements to walleye, the mechanisms through which TDS and secchi regulate the northern pike distribution can be considered the same as walleye.

White sucker (A) and (A+B) model results indicated that lake size, climate and maximum depth are influential predictors of its occurrence (Figure 2.24). Among these, lake size is the only predictor showing a positive relationship with occurrence (Tables 2.4 and 2.6). White sucker are usually fish of warmer, shallow lakes or warm, shallow bays, and tributary rivers of larger lakes. In addition to spawning migrations, movements, other than a general tendency to move offshore with increase in age, are random, probably in response to temperature (Scott and Crossman 1998). White suckers are moderately active during the daytime but active feeding is usually restricted to near sunrise and sunset when they move into shallower water. This species spawn in the spring, usually from early May to early June. Adults usually migrate from lakes into gravelly streams when stream temperatures first reach 10°C, but they are also known to spawn on lake margins, or quiet areas in the mouth of blocked streams. They are bottom feeders, so benthic production is very important for them. Previous white sucker occurrence models showed surface area, total shoreline perimeter and max depth to be important lake morphometry predictors of white sucker occurrence, which agrees with the model results showed here (Olden and Jackson 2002). Climate plays an important role on this species distribution, therefore the models presented in this study represent an improvement over the model presented by these authors. The negative influence of both DD5 and MAT on the white sucker distribution reflect the preference of this species for cool waters, since temperature increases will most likely represent loss of habitat for this species, which needs cool water for spawning (Scott and Crossman 1998).

For smallmouth bass, (A) and (A+B) models' results (Figure 2.25) show that lake size, climate, TDS and secchi are important in determining this species' distribution, with lake size, climate and secchi having positive influences, and TDS having a negative influence (Tables 2.4 and 2.6). Smallmouth bass habitat varies with size and time of year. In the spring, adult fish congregate on the spawning grounds, where males build nests on a sandy gravel, or rocky bottom usually near the protection of rocks, logs or, more rarely, dense vegetation (Scott and Crossman 1998). Late spring they are usually found in rocky and sandy areas of lakes and rivers, in moderately shallow water. In the heat of the summer they usually retreat to greater depth. The distribution of smallmouth bass has been historically limited to the south and south-central regions of Ontario where July air temperatures exceed 18°C (Shuter et al. 1980). Their current range is expanding throughout North America via natural and human-mediated dispersal (Sharma and Jackson 2008), and under scenarios of climate change, this species have been predicted to expand their range northward (Chu et al. 2005; Sharma et al. 2007). Previous walleye occurrence models have showed the importance of lake morphometry variables (Olden and Jackson 2002; Vander Zanden et al. 2004; Van Zuiden et al. 2016) and climate (Olden and Jackson 2002; Vander Zanden et al. 2004; Chu et al. 2005; Alofs and Jackson 2015; Van Zuiden et al. 2016). In addition, Van Zuiden et al. (2016) model results agree with the results presented here related to the positive influence of secchi on smallmouth bass distributions, indicating the preference of this species for clean, transparent waters. On the other hand, a lack of agreement was noted between the smallmouth bass occurrence model presented by Olden and Jackson (2001), where they show TDS to influence positively this species' distribution, and the smallmouth bass models obtained in this study, which showed TDS to have a negative relationship with occurrence.

The results presented by this study corroborates the idea of model improvement through the addition of biotic variables, in this particular case, including occurrence of other fishes from the same community as predictors. For instance, the result seen on Figure 2.22 and Table 2.6 for walleye (A+B) models indicate that presence of northern pike was the most important variable explaining its distribution. Previous studies on Ontario lakes reported a high frequency of occurrence of walleye and northern pike together in big lakes, sharing the same habitat requirements. Scott and Crossman (1998) reported that northern pike are probably the most important predators of the walleye over much of its range, or even an important competitor as it is the only other major, shallow-water predator in the north. Figure 2.22 also shows that presence of lake trout, lake whitefish and smallmouth bass are important in predicting walleye distributions, while Table 2.6 indicated that the influence of lake trout is negative on walleye occurrences, contrary to the influences of lake whitefish and smallmouth bass. Vander Zanden et al. (2004) found that the number of piscivores in a lake was positively related to smallmouth bass occurrence. However, fish species richness and food chain length are known to increase with lake size (Vander Zander et al. 1999, Post et al. 2000). Lake size and related increases in habitat heterogeneity may simultaneously favour predator richness and establishment of the fish species target of this study. Moreover, the positive relationships among species seen on Tables 2.5 and 2.6 can indicate biotic acceptance rather than biotic resistance explaining patterns in the establishment of all six freshwater fishes, and negative relationships between species could also indicate distinct habitat requirements, a "habitat checkerboard" (Gilpin and Diamond 1982; Jackson et al. 1992). Shared food and habitat resources can also be an explanation for these the high variable importance attributed to walleye when predicting smallmouth bass (Figure 2.25), northern pike (Figure 2.23) and lake whitefish (Figure 2.21). Several studies pointed out the importance of considering biotic variables when conceptualizing SDMs (Guisan and Zimmermann 2000; Guisan and Thuiller 2005; Ritz 2005; Schröder 2008; Elith and Leathwick 2009; Zimmermann et al. 2010; Wisz et al. 2013). Gilman et al. (2010) reported that interactions among species can strongly influence how climate change affects species at every scale and that failure to incorporate these interactions limits the predictive ability of models. Biotic interactions can affect species responses to abiotic environmental changes differently along environmental gradients, and abiotic environmental changes can likewise influence the nature of biotic interactions (Brooker 2006; Meier et al. 2011). In a recent study, Wang and Jackson (2014) showed how the incorporation of biotic interactions improved model generality when predicting biological invasions. The poor results related to models including only biotic information reinforces the importance of environmental features while predicting fish occurrences, however, these models can provide an insight about which species facilitates the occurrence of other species, through competition, predation or facilitation. The low deviance results seen for B models of smallmouth bass and lake trout (Figure 2.4) could be linked to limited distribution of these two species, since they represent extremes on the temperature preference gradient. Smallmouth bass is only found in the south where its distribution overlaps with every species (Figure 2.3), while lake trout distribution is linked to lake depth requirement, since deep lakes provide optimum habitat for this species.

The structural properties of a data set require the use of statistical approaches that best capture the response in the data set while providing information on the important environmental determinants structuring species distributions and communities. The models developed in this study using either abiotic variables (here called A models), biotic information (B models) or both of them (A+B models) provided some valuable insights into the relationships between species occurrences, habitat and how the fish community determines the establishment of a given fish species. Linear methods are traditionally the most popular approaches used (Sharma et al. 2008). However, results revealed that machine learning methods, specifically random forests and boosted regression trees, explained the species occurrence variation better (resulting in high deviance values), besides the good performance seen during the calibration and validation assessment. Machine-learning methods can have high predictive abilities for data sets that exhibit linear and nonlinear relationships between predictor variables, high-order interactions and multicollinearity (Breiman et al. 1984, De'ath 2002). Machine learning methods produce discontinuous changes at certain points along the predictor variables and identify high-order local interactions that the linear-based approaches used in this study do no appear to accommodate. Wang and Jackson (2014) reported good results obtained from calibrating and validating random forest models, with stable overall performance. However, a potential drawback of machine-learning methods can be the possibility of over-fitting a model (Sharma et al. 2008). How to choose appropriate modelling approaches, predictor variables, model validation methods, and performance metrics are important steps if we want to get more reliable inferences or predictions rather than data-specifics results or statistical artifacts. Although some of the models presented here are correlative, and thus we cannot directly determine, but only imply causation, the results are consistent with findings from many studies of north-temperate fish populations (e.g. Olden and Jackson 2002; Wang and Jackson 2014; Alofs and Jackson 2015). Each modelling approach possesses fine tuning parameters and exploring all their possibilities for each method is beyond the scope of this study, which focused on: (i) the development of fish-habitat models; (ii) comparison of variables of importance among species and (iii) compare predictive performance among models. Moreover, the results shown here provide an indication of the candidate models that are worth considering further. Finally, it is important to keep in mind that the fundamental goal of ecological research is to *understand* how observed ecological patterns are generated by specific processes or constraints, thus allowing for valid generalizations (Wiens 1984). Therefore, understanding patterns of distribution of species requires that we evaluate predictions about functional relationships between species and their environments, in other words, understanding and prediction are both important aspects in ecological modelling. As long as the chosen modelling approaches are constructed logically from mechanistic principles, new data and knowledge can be used to refine the conceptual framework over time, thereby increasing ecological understanding and, potentially, predictive capability (Wiens 1992).

The variable importance assessment indicated that surface area of the lake was an important morphometric feature influencing the establishment of each fish species. The importance of lake area on species diversity has been broadly recognized (Barbour and Brown 1974; Eadie and Keast 1984; Nate et al. 2000; Olden and Jackson 2002; Hansen et al. 2015; Edwards et al. 2016). Large lakes allow the development of abundant vegetation in the littoral zone, which facilitates the spawning of species such as smallmouth bass. Furthermore, surface area of the lake serves as an indirect measure of the diversity of habitats available in lakes, which may be important to support small-bodied, forage fish upon which these species feed.

Large lakes promote resource (habitat and prey) heterogeneity (Barbour and Brown 1974; Eadie and Keast 1984).

Together with lake size, climate is broadly recognized as a strong predictor of fish distributions (O'Connor et al. 1981; Schlesinger and Regier 1982; Christie 1986; Magnuson et al. 1990; Jackson et al. 2001; Poesch et al. 2016), and the results seen here for all species confirm that fact. Air temperature is especially important for smallmouth bass because a northern climatic zone was studied, and the restriction of this species' distribution to the southern part of Ontario shows a strong temperature dependency. Temperature is the controlling factor pacing the metabolic requirements for food and governing the rate processes involved in food processing (Brett 1979). Almost all species in the young stages show a typical rapid increase in growth rate as temperature rises passing through a peak (optimum temperature) and frequently falling precipitously as high temperatures become adverse. Since temperature is the most important abiotic factor controlling fish physiology (Kao et al. 2015), projected increases in water temperature resulting from climate change are expected to affect the growth of freshwater fishes, their population dynamics, and the ecology at community and ecosystem levels (Jeppesen et al. 2010; Portner and Farell 2008; Sheridan and Bickford 2011). These increases in water temperature may also change geographical distributions of freshwater fishes (Chu et al. 2005; Comte et al. 2013) and the production of freshwater fisheries (Ficke et al. 2007; Portner and Peck 2010). In this study, results revealed different influences of DD5 and MAT on different fish species (e.g. DD5 and MAT influence negatively lake trout occurrences, while DD5 and MAT have mixed influences on walleye and last, these two climate variables have a positive influence on smallmouth bass – see Table 2.4). One hypothesis that for this is what was stated in 2.3.1: species use different environmental cues to maximize different life-history stages (e.g. young versus old individuals) and needs (e.g. reproduction versus growth) and so different temperatures may influence species differently across these stages of life and needs (Lee et al. 2016).

Future studies should investigate how consider more elaborate indexes of community composition, which would incorporate information about other fish species belonging to the same lake communities. Biotic interactions are dynamic within changing environment, but the interactions and species assemblages under investigation should remain relatively static if it is of interest to include them into SDMs, and this is particularly important for studies on species under climate or land-use change (Davis et al. 1998). Some environmental conditions may be an indirect reflection of biotic interactions, falsely suggesting a direct dependence on abiotic factors where in fact a biologically mediated dependence may be the case. On the other hand, it is also possible that some biotic interactions might be just caused by abiotic factors (Guisan and Thuiller 2005). Adding biotic information into SDMs may increase the complexity of modelling, therefore the most efficient way to incorporate only the necessary biotic factors should be explored.

More effective conservation of aquatic biodiversity will require new approaches that recognize the protection of key local- and regional-scale processes that shape fish distributions (Angermeier and Winston 1999). Developments in these areas require an increased reliance on probabilistic models and will represent an important advancement in both population and community ecology. This study shows that statistical modelling approaches exhibit considerable promise in providing testable, predictive models for fish ecology. Predictive models can play an important role by forecasting the likelihood of local establishment and spread of non-native species and thus help set proactive conservation priorities for preserving vulnerable populations.

The results related to the importance of lake size to all species serve as a basis for conservation measures that incorporate large lakes as part of the conservation strategy, given their importance in providing adequate habitat for several species.

CHAPTER III

EFFECTS OF ABIOTIC FACTORS ON THE BIOMASS OF SIX FISH SPECIES IN ONTARIO LAKES

W.S.C. Moreira, N. Lester and P.R. Peres-Neto

3.1 Summary

The ecological and economical importance of freshwater fish populations is unquestionable, being valuable for many reasons. They regulate ecosystem structure and function through the processes of selective predation, nutrient cycling, bioturbation, play a major ecological role in structuring the benthic and zooplanktonic invertebrate communities, they are key indicators of ecosystem health and environmental disturbance and possess a central spiritual value to many native cultures. However, high aquatic habitat degradation from overexploitation, water pollution, flow modification and invasion of exotic species are causing freshwater fish to become one of the most threatened groups of vertebrates worldwide. It is therefore vital that fisheries managers, governments and the outdoor recreation industry prioritize their actions on ensuring the sustainability of the actual fish stocks under exploitation and habitat loss, via productivity assessments. It is of interest to fisheries ecologists to know how fish production and biomass vary among ecosystems and populations or due to changes in environmental conditions. A first step towards this goal is to determine which characteristics of ecosystems have the greatest impact on them. The objective of this chapter is to develop biomass models for six freshwater fish species present in Ontario lakes (lake trout, lake whitefish, walleye, northern pike, white sucker and smallmouth bass), using environmental predictors based on principal environmental determinants (light, heat, nutrients), together with fishing pressure. The species were chosen in order to disentangle the effects of climate and exploitation. The database used to develop the biomass models was provided by a broad-scale monitoring (BsM) program from the Ontario Ministry of Natural Resources and Forestry and consisted of surveys done on 722 lakes. Biomass was expressed through biomass-per-unit effort (BPUE, kg/gang-night) and modelled using boosted regression trees. We hypothesized that angling score would be an important predictor of walleye, northern pike, lake trout and smallmouth bass, but results showed that it does not represent an important variable for these species, just moderately for northern pike. Our study provides a better understanding of the different factors driving biomass of the different species; how they interact, while bringing to light the importance of lake morphometry and climate to predict both occurrence and biomass of freshwater fish species from Ontario lakes. From the management point of view, these two environmental determinants should be primarily taken into account, not only when selecting new conservation areas but also when seeking an optimization of lake resources contributing to the sustainable practice of sport fishing.

3.2 Introduction

The ecological and economical importance of freshwater fish populations is unquestionable. They regulate ecosystem structure and function through the processes of selective (non-random) predation (Carpenter et al. 1985), nutrient cycling (Schindler 1992) and bioturbation (Gelwick et al. 1997); they play a major ecological role in structuring the benthic and zooplanktonic invertebrate communities (Benndorf et al. 1984; Mazumder et al. 1988; McQueen et al. 1990); they are key indicators of ecosystem health and environmental disturbance (Karr 1981); and they possess a central spiritual value to many native cultures (Swezey and Heizer 1977). Economically, sport freshwater fisheries are increasingly replacing commercial fisheries landings, consisting of a driving force generating recreational values of $CDN 8.8 billion in Canada in 2010 (Fisheries and Oceans Canada 2012), and sportfishing's social value lies in its role as a recreational experience.

However, high aquatic habitat degradation from overexploitation, water pollution, flow modification and invasion of exotic species are causing freshwater fish to become one of the most threatened groups of vertebrates worldwide (Allan and Flecker 1993; Naiman et al. 1995; Naiman and Turner 2000; Jackson et al. 2001; Malmqvist and Rundle 2002; Rahel 2002; Postel and Richter 2003; Revenga et al. 2005). It is therefore vital that the fishing industry and governments prioritize their actions on ensuring the sustainability of the actual fish stocks under exploitation and habitat loss. Understanding the effects of such stresses on the abundance of various fish species depends on our knowledge of how natural factors affect fish species.

At equilibrium, unexploited fish stocks produce exactly enough biomass to balance natural mortality (Downing and Plante 1993). The renewal of the stock is characterized by production, defined by Ricker (1975) as amount of tissue generated per unit time per unit area, regardless of whether or not fish survive during the time interval. By definition, production (P) = biomass (B) * mortality (M), with production and biomass being commonly standardized by lake area (Mertz and Myers 1998). It is thus of interest to fisheries ecologists to know how biomass and fish production varies: (1) among ecosystems and populations; (2) over time; (3) due to changes in environmental conditions (Bradford and Irving 2000) and/or (4) across different life history strategies (Musick 1999). A first step towards this goal is to determine which characteristics of ecosystems have the greatest influence on biomass and consequently, production. Investigations that have as a main scope to identify the main factors driving biomass of a certain species also assist in understanding which factors drive the production of the same species. In the previous chapter, we showed how the occurrence of six fish species in Ontario lakes is affected by various abiotic factors and the community composition of these lakes; in this chapter we explore the effect of abiotic factors on the biomass of the same fish species.

An exhaustive number of models to account for freshwater fish biomass and production is available in the literature, however, these models disagree about the characteristics of lakes that have the greatest influence on them. For instance, previous fish biomass models showed that total phosphorus (Quiros 1990; Randall et al. 1995), carbonate content (Carlander 1955) , total nitrogen (Moyle 1956) and total dissolved solids (Rawson 1960; Jenkins 1967) were good predictors; for fish production models, mean depth or lake area (Rounsefell 1946; Rawson 1955), fish density (Randall et al. 1995), primary production of the water column (Downing et al. 1990), secchi depth (Lester et al. 2004), thermal-optical habitat (Lester et al. 2004), morphoedaphic index (Ryder 1965), annual average fish standing stock (Downing et al. (1990), total phosphorus (Hanson and Legget 1982), algal biomass (Jones and Hoyer 1982), macrobenthos (Matuszek 1978), climate (Schlesinger and Regier 1982) and lake morphometry (Schlesinger and McCombie 1983) were shown to be influential parameters. Although these models disagree in terms of the most important predictors of biomass and production, all the factors identified as influential fit within a conceptual model proposed by Ryder and Kerr (1989).

Biomass of different species is mainly driven by, according to Ryder and Kerr (1989), the four fundamental ecosystem properties: light, heat, dissolved oxygen and nutrients, which are critical environmental factors for fish survival and subsequent production. These four fundamental properties do not constitute an exhaustive list of all factors critical to the survival and production of fishes, but rather are a subset of high priority. Heat and light are the principal energetic inputs affecting the survival of fishes while oxygen and nutrients constitute the two fundamental and essential material items (see Figure 3.1). As a generalization, survival depends on fish satisfying their most critical needs first (dissolved oxygen, preferred temperature), at which point behaviour becomes dependent on the controlling factors, nutrients and light. These four factors are interrelated to varying degrees, for example temperature affects dissolved oxygen saturation levels; light controls photosynthesis, and hence, nutrient availability; photosynthesis, in turn, contributes to the concentrations of dissolved oxygen.

Temperature is one of the most important abiotic factors influencing fish survival and performance (Christie and Regier 1988). As fish are obligate poikilothermic ectotherms, their body temperatures are equal to or within a few fractions of a degree of the surrounding water temperature (Beitinger et al. 2000). Therefore, they are highly dependent on water temperature to maintain important biochemical, physiological and life history processes (Beitinger et al. 2000). Previous studies have shown that physiological performance of fish is maximized within a narrow temperature range and that, depending on the species, optimal temperatures for many processes centre around a specific value (Jobling 1994). Christie and Regier (1988) show strong relationships between thermal habitat measures and sustained yield of lake trout, lake whitefish, walleye and northern pike. Shuter et al. (1983) have shown that differences in the supply of thermal habitat for fish are predicted largely from lake morphometry and air temperatures, indicating that these variables are also important when assessing fish production.

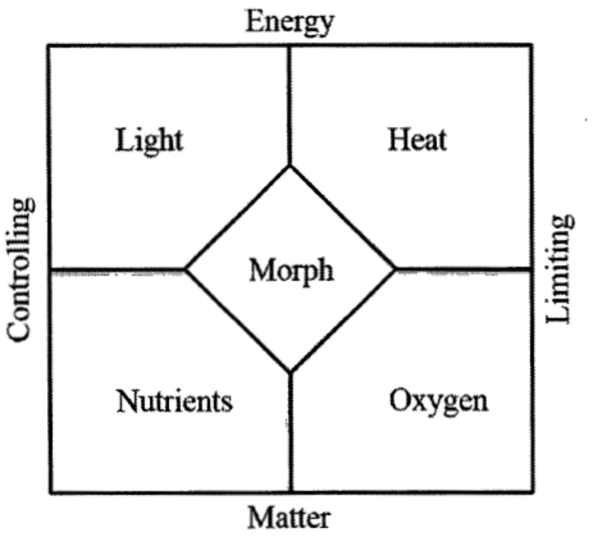

Figure 3.1: Schema of four principal environmental determinants as configured and entrained by the morphology of an ecosystem, in this case, habitat. The four determinants are comprised of two energetic factors (light and heat) and two material factors (nutrients and oxygen). Light and nutrients are controlling in the sense that they influence behavior and metabolism, but are rarely lethal at any level found in nature. Heat and oxygen are controlling in their intermediate ranges, but may be limiting at the extremities, and therefore are often lethal. Extract from Ryder and Kerr (1989).

Light is another important environmental determinant controlling fish behavior and metabolism. Subsurface illumination, because of its environmental pervasiveness at various levels of intensity, is rarely considered to be as critical an environmental determinant as dissolved oxygen and temperature. The retinal response to light is not the same in all fish species, despite their broad tolerances of subsurface illumination.

Some species are able to tolerate these wide ranges because they possess specialized retinal structures, or have the ability to occlude light with moving pigments (Ali et al. 1977). Other species utilize the turbidity of the water column to shield sensitive retinae from the refracted sunlight of the water surface, while still others use physical structures for the same purpose (Ryder 1977). Some fishes, such as the walleye (*Sander vitreus*), have scotopic (dim-light) vision and are eminently well adapted to crepuscular and nocturnal feeding and spawning (Scott and Crossman 1998). On the other hand, a sympatric piscivorous species, the northern pike (*Esox Lucius*), normally feeds and spawns only during daylight hours (Sawatzky et al. 2007). Walleyes may actually be temporarily stunned at high daylight intensities and are unlikely to be found feeding at the surface of a clearwater lake on a sunny day. Northern pike which have photopic (daylight) vision, are ineffectual feeders at night despite their extensive use of the lateral lines in feeding (Nikolsky 1963). Based on these observations, Ryder and Kerr (1989) regarded light as an ecological cleaver that determines the relative dominance or subordinance of species. Lester et al. (2004) stressed the importance of water clarity in predicting potential yield of walleye, integrating this variable with lake bathymetry, thermocline depth and climate.

The other controlling factor in the Ryder and Kerr model is nutrients. Nutrients reach fish through the intermediacy of lower trophic levels, being first absorbed by autotrophic plant life and passed up the trophic ladder through the agencies of grazing and ultimately, predation. As in the case of light, nutrients are rarely a lethal factor by themselves, unless augmented to extremely high levels through eutrophication and associated with anoxic levels of dissolved oxygen. Low levels of nutrients are not likely to be lethal as fish communities adapt by maintaining low levels of standing stocks through compensatory feedback (Ryder and Kerr 1989). The morphometry of the lake is especially important in respect to the way nutrients are channeled in the system. The first published fish production models were a function of simple morphometric variables like area or lake mean depth (Rounsefell 1946; Rawson 1952). Later on, Ryder (1965) introduced the morphoedaphic index (MEI = Total Dissolved Solids/ Mean Depth), which brought popularity to nutrients as an important environmental determinant to predict fish production. MEI's simplicity has made it a landmark of lake fisheries management, being extensively used together with other physical, chemical and biological indices. Since its first appearance, MEI has been the subject of more than 100 publications in primary scientific literature (e.g. Jenkins 1967; Henderson and Welcomme 1974; Oglesby 1977; Ryan and Harvey 1977; Matuszek 1978) . Later on, Ryder (1982) reviewed MEI and acknowledged that fish production should be addressed at three hierarchic levels: global (where temperature and area are of prime importance and are interrelated through total caloric content), regional (where mean air temperature, lake morphometry and nutrients are important) and infra-regional (where mean depth would be influential). Schlesinger and Regier (1982) proposed an adaptation to MEI by incorporating climate via mean annual air temperature, showing good model results while accounting for regional differences in climate. On the other hand, Downing and Plante (1993) criticized the performance of MEI in their work, showing that it was not significantly correlated with fish production when biomass and body mass effects were considered. Models linking fish production to phytoplankton (Oglesby 1977) and benthos (Matuszek 1978) productivity became popular in the 70s, showing themselves to constitute improved indicators over the morphoedaphic index. Recent works have used total phosphorus (TP) as a nutrient indicator, since it is a proxy for nutrient control of phytoplankton biomass, having a direct link to primary production and consequently, production on higher trophic levels of the lake (Prairie et al. 1989; Vadenbocoeur et al. 2008; Vander Zanden et al. 2011). In addition, in studies considering the effect of total phosphorus on fish production, it was reported its relationship with dissolved organic carbon (DOC) and how the latter serves as a proxy of water clarity, impacting benthic production (Finstad et al. 2014).

Oxygen (i.e., dissolved oxygen, DO) is another limiting fundamental environmental determinant of aquatic life. The proportion of oxygen in air that dissolves in water is about 35%. This solubility is governed by atmospheric and hydrostatic pressure, turbulence, temperature, salinity, currents, upwellings, ice cover, and biological processes (Wetzel 1983). Major sources of DO in water are the atmosphere and the photosynthesis by aquatic vegetation. The amount of oxygen available for aquatic life, however, depends on the factors that affect its solubility. In large and deep freshwater systems, oxygenation depends on circulation by winds, currents, and inflows to move aerated water away from the surface. Due to human activities, thermally stratified lakes are affected by oxygen depletion via increase of phosphorus loadings, that causes algal blooms and therefore, depletion of late-summer dissolved oxygen levels in the hypolimnion (cold lower water layer) of lakes (Lienesch et al. 2005). In Ontario, surface water temperature in the summer becomes too warm for cold water species. Consequently, these three factors (morphometry of the lake, nutrients load and DO) interact to determine habitat availability for cold water species (Evans 2007).

Overall, the research on fish production has shown that the main parameters related to it are alkalinity, algal biomass (chlorophyll a), air temperature, area, benthos standing crop, body size, fishing effort, mean depth, phytoplankton productivity, total dissolved solids, total nitrogen concentration and total phosphorus concentration. Understanding which factors drive the dynamics of different species and better describe their habitat preferences is key to better manage impacts of climate change on fish production.

The objective of Chapter 3 is to develop biomass models for the six fish species for which fish-habitat models were developed in Chapter 2 (lake trout, lake whitefish, walleye, northern pike, white sucker and smallmouth bass). The same abiotic variables were chosen – all of which are based on the principal environmental determinants identified by Ryder and Kerr (1989). In addition, we included fishing pressure because several of these species are harvested by anglers (mainly lake trout, walleye and northern pike; smallmouth bass is also a popular sport fish target, but harvested less because anglers practice more catch and release for this species). Also, fishing pressure included to assist in understanding the effects of abiotic factors, although it is of interest to know how biomass of target species have been impacted by fishing pressure. The species were chosen in order to disentangle the effects of climate and exploitation (Table 3.1), since angling pressure tends to correlate with temperature. Lake trout, walleye, northern pike and smallmouth bass were selected because they are popular targets of recreational fishing (Fisheries and Oceans Canada 2012), and this group of species includes one cold water species (lake trout), two cool water species (walleye and northern pike) and one warm water species (smallmouth bass). Angling pressure on these species varies across the province of Ontario, tending to be higher in the South which has higher human population. The other two species consist of one cool water (white sucker) and one cold water (lake whitefish) species, both receiving very little stress from angling.

Table 3.1: Species chosen according to the exploitation level and thermal guild

Thermal Guild	Exploitation	
	Low	High
Cold water	Lake Whitefish	Lake Trout
Cool water	White Sucker	Walleye Northern Pike
Warm water		Smallmouth Bass

Four models with different sets of predictors were built for each species: (i) a full model including all the environmental features to be used in this study, together with angling pressure; (ii) a model that included exactly the same environmental variables used for the development of fish occurrence models in Chapter 2, in order to verify the level of agreement between biomass and presence-absence models for the six fish species; (iii) a model with secchi depth, total dissolved solids (TDS) and pH encompassing the pool of variables representing the water chemistry environmental determinant; and (iv) a model with total phosphorus (TP), dissolved organic carbon (DOC) and pH representing water chemistry. Models (iii) and (iv) allow comparisons in order to determine which set of water chemistry parameters provide a better model parsimony, consisting of an optimization of the full model. Fishing pressure (in the form of an angling score) was also part of the modelling routine in models (i), (iii) and (iv), and the degree of its influence on the biomass of the different species is evaluated. No study to date has carried out a thoroughly investigation on biomass models specific to each of the fish species targeted here, evaluating: (i) the environmental determinants that are important for each species; (ii) the extent to which biomass is affected by angling pressure; (iii) the level of agreement between the importance of variables to predict biomass and presence-absence; and (iv) how the different environmental features contribute to biomass prediction via BPUE.

3.3 Methods

3.3.1 Fish-environment data source

The database used to develop the species biomass in this study was provided by the Inland Lakes Broad-scale Monitoring program of Ontario (BsM). This program is part of Ontario's Ecological Framework for Fisheries Management which aims to: (i) describe of the geographic distribution, extent and characteristics of aquatic resources in Ontario; (ii) estimate, with known confidence, the current status and trends in selected indicators of fishery resources; (iii) identify natural and anthropogenic stresses affecting the condition of aquatic resources and (iv) provide periodic reports on the state of aquatic resources.

3.3.2 Lake Selection

The target population for the BsM program is lakes larger than 50 ha, of which there are approximately 11000 in Ontario. The data used in this study is based on surveys from 721 lakes which were conducted during the first cycle of BsM (2008-2012). The dataset is a stratified random sample of lakes (see Figure 3.2). The province of Ontario is divided into 14 inland management units (i.e. excluding the Great Lakes and Ottawa River); within each of these zones, lakes were divided into size classes based on lake area (50-500 ha, 500-1500 ha, 1500-5000 ha, and > 5000 ha) and an equal number of lakes within each size class were randomly selected. This selection procedure was adjusted to ensure that 10% of known lake trout lakes and 20% of known walleye lakes were sampled in each zone. Standardized surveys were conducted on each lake to sample lake morphometry, water quality, fish assemblages, and angling pressure in each lake.

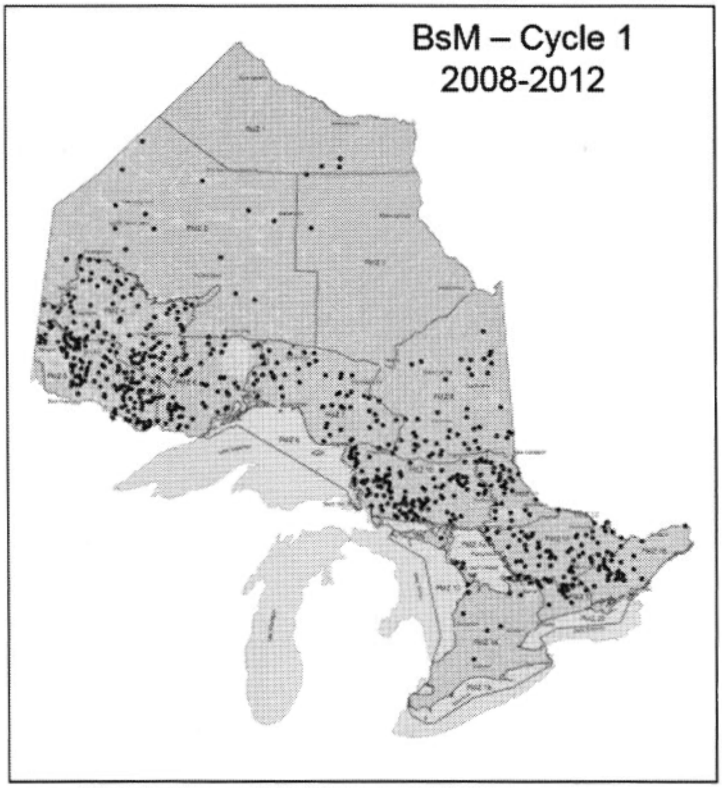

Figure 3.2: Distribution of the sampled lakes across Ontario during the first cycle of the BsM program.

3.3.3 Fish community sampling

Fish assemblages were sampled using small and large mesh gillnets. Data used in this study is based on the large mesh gillnets which conforms with the proposed North

American standard (Bonar et al. 2009): 24.8 m in length with eight panels of mesh ranging from 38 to 127 mm (38, 51, 64, 76, 89, 102, 114, and 127 mm). Nets were set in the afternoon and lifted in the morning with a target soak time of 18 h. A double gang strap was typically used (i.e., two nets strung together). Sampling details are described in Sandstrom et al. (2011).

3.3.3.1 BPUE

Sampling was depth stratified, and gillnets gear was fished in all depth strata: 0-3, 3-6, 6-12, 12-20, 20-35, 35-50, 50-75, 75+ m. Sampling sites were assigned randomly within each stratum. The number of sites per stratum ranged from two to nine depending on lake size (see table 4 in Sandstrom et al. 2011). Species and fork length were recorded for all caught fishes. Lake-wide BPUE (biomass-per-unit effort) was calculated as area-weighted estimates of BPUE. For each lake, we calculated the BPUE and benthic area in each sampled depth stratum and then calculated the area-weighted mean. The resulting BPUE was based on the full range of body sizes caught in each lake, measured as kg/gang-night, where a gang is 24.8 m long and 1.8 m high, equivalent to 45 m2. BPUE is used as an index of biomass density (kg/hectare), using as an assumption the relationship of BPUE to stock size and consequently, biomass density (Hanchet et al. 2005). Table 3.2 shows descriptive statistics of BPUE for the six freshwater fishes in this study.

Table 3.2: Descriptive statistics of BPUE (kg/gang-night) for the six sport fish species

Thermal guild	Fish species	Number of lakes with species present	Mean	Standard Deviation	Range
Cold	Lake trout	288	0.699	0.729	6.53
	Lake whitefish	362	1.004	1.196	10.5
Cool	Walleye	474	1.827	1.974	12.08
	Northern pike	499	0.882	0.921	6.16
	White sucker	643	1.303	1.362	9.18
Warm	Smallmouth bass	310	0.350	0.320	1.89

3.3.4 Environmental predictors and angling score

Table 3.3 summarises the descriptive statistics of all environmental predictors that were used to predict the biomass of all six freshwater fish species, along with fishing pressure: growing degree days above 5°C for 1981-2010 (DD5), mean air temperature for 1981-2010 (MAT), surface area of the lake transformed using log10 (L10Area), shoreline development factor (SDF, see Chapter II section 2.3.1 for details), maximum and mean depth of the lake (Depth_Max and Depth_Mn respectively), total dissolved solids (TDS), total phosphorus (TP), dissolved organic carbon (DOC), secchi depth (Secchi), pH and angling score, which was computed based on angling pressure (angler-hours/hectare). Angling pressure was measured as part of the BsM protocol. Aerial surveys of angler activities in summer (eight weekday and eight weekend flights) and winter (six weekday and six weekend flights) provided the data used to estimate annual angling pressure. The summer count included the number of vessels and shore anglers, whereas the winter count included the number of ice huts and anglers on open ice. Weekend and weekday angler counts were used to calculate mean angling activity for each season in each lake i:

$$Activity_{i,s,m} = \frac{5/7\,\overline{Count}_{weekday,i,s,m} + 2/7\overline{Count}_{weekend,i,s,m}}{Area_i}$$

$Activity_{i,s,m}$ is then multiplied by the number of anglers in vessels, shore anglers, ice huts, open-ice anglers, season length, and hours fished per day and then summed for all seasons and methods to produce annual estimates of angling pressure.

$$Ang_i = \sum_{s,m=1}^{s,m} Activity_{i,s,m} \times angler_m \times season \times hours$$

See Chu et al. (2016) for more details on these expressions. Angling pressure was then converted to angling score as a form of standardization given that angling pressure has a very skewed distribution, with big gaps among some values of its range. The BsM survey made available information about extra water chemistry variables such as TP and DOC, which were not collected in the AHI survey used to build species distribution models in Chapter 2. Table 3.4 provides information on the correlations among all predictors to be used in the biomass modelling process. The same comments done regarding the correlations seen in Chapter 2 hold true for this chapter. Additionally, the correlation between angling score and climate reflects how angling is dependent on climate, with increased angling in the south of Ontario. The negative correlation seen between secchi and DOC is related to the impact of DOC on water clarity, while the correlation between pH and TDS is explained by conductivity, which is positively related to both of these water chemistry parameters.

Table 3.3: Descriptive statistics of all environmental predictors and fishing pressure. The ranges for angling score were: 1= 0 – 1 hr.ha^{-1}.yr^{-1}; 2= 1.1-5 hr.ha^{-1}.yr^{-1}; 3= 5.1-10 hr.ha^{-1}.yr^{-1}; 4= 10.1-20 hr.ha^{-1}.yr^{-1}; 5= 20.1-40 hr.ha^{-1}.yr^{-1} and 6= 40.1+ hr.ha^{-1}.yr^{-1}.

Predictors		n	Mean	Standard Deviation	Range
Climate	DD5 (growing degree-days)	683	1611.7	212.2	1187
	MAT (°C)	683	3.229	1.762	10
	L10Area (ha)	683	2.782	0.667	3.64
Lake Morphometry	SDF	681	4.69	3.12	19.78
	Depth_Max (m)	683	29.67	22.93	185.1
	Depth_Mn (m)	683	8.74	6.64	39.5
	TDS (mg/L)	683	49.76	49.02	385.6
	TP (ug/L)	682	9.93	6.45	50.8
Water Chemistry	DOC (mg/L)	682	7.33	3.51	18.4
	Secchi (m)	682	3.60	1.79	11.7
	pH	673	7.16	0.53	3.1
Angling Score		663	2.53	1.28	5

Table 3.4: Pearson correlations among BPUE predictors

	DD5	MAT	L10Area	SDF	Depth_ Max	Depth_ Mn	TDS	TP	DOC	Secchi	pH	Angling_ Score
DD5	1.00											
MAT	0.91	1.00										
L10Area	-0.11	-0.24	1.00									
SDF	0.00	-0.09	0.63	1.00								
Depth_Max	0.06	0.07	0.39	0.29	1.00							
Depth_Mn	0.11	0.12	0.28	0.06	0.90	1.00						
TDS	0.32	0.33	0.07	-0.08	-0.14	-0.16	1.00					
TP	0.05	-0.04	0.09	-0.05	-0.36	-0.35	0.38	1.00				
DOC	-0.42	-0.50	0.02	0.01	-0.38	-0.45	-0.03	0.46	1.00			
Secchi	0.23	0.29	0.02	-0.01	0.49	0.54	0.03	-0.53	-0.68	1.00		
pH	0.07	0.01	0.20	-0.01	-0.18	-0.19	0.75	0.32	0.09	-0.04	1.00	
Angling_Score	0.60	0.65	-0.03	0.00	-0.09	-0.08	0.43	0.16	-0.27	0.10	0.22	1.00

3.3.5 Statistical analyses

Boosted regression tree analyses (BRT) were used to identify which variables influenced BPUE for each freshwater fish species. The choice of modelling approach comes from the good results shown by this method in Chapter 2, added to its capability to create partial dependence plots, which allows the assessment of the influence of each independent variable on the response variable. BRTs combine regression trees (generated with recursive binary splits) with boosting (combining trees to improve predictive performance) to produce a model built on multiple instead of single trees (Elith et al. 2008). The general approach is to iteratively compute a sequence of trees in which each successive tree is built from the prediction residuals of the preceding tree. The optimal number of trees is determined when the mean predictive deviance of the 10-fold cross-validation process is minimized given the learning rate, tree complexity, and bag fraction settings. Deviance measures how much the predicted values differ from the observations and represents the loss in predictive performance due to a suboptimal model. Percentage deviance explained, or pseudo R^2, was calculated as 1- (residual deviance/total deviance) and, similar to traditional regression techniques, should approach one when predictive performance is maximized (Elith et al. 2008). The 10-fold cross-validation process evaluates predictions of the model using 10 subsets of randomly selected data (Emmrich et al. 2014).

Relative influences of the explanatory predictors are determined based on the number of times that they are included in the trees, and the improvement in the models with their inclusion is averaged across all trees (Friedman and Meulman 2003; Elith et al. 2008). They are scaled to 100 with higher numbers indicating stronger influences on the response variables (Elith et al. 2008). The BRT models were developed with Gaussian error distributions, learning rate of 0.005, tree complexity of 5 and bag fraction of 0.5. All computations were done using R 3.3.3 (R Development Core Team 2008), and the BRT models were developed using the "dismo" and "gbm" packages as outlined by Elith et al. (2008).

For each fish species, four different sets of variables were used to estimate BPUE. The first set consisted of the full set of predictors shown in Table 3.3; The second set (set 2) excluded TP, DOC and angling score in order to keep the same set of variables used to predict presence-absence in Chapter 2, with the objective to verify if the main variables used to predict BPUE were also indicated as important ones to predict presence-absence in that chapter; The third set (set 3) excluded TP and DOC, having pH, TDS and secchi depth as the predictors representing water chemistry; And finally the fourth set (set 4) excluded TDS and secchi depth, having pH, DOC and TP as predictors representing water chemistry. Sets 3 and 4 had as main objective to allow verification of whether predictive performance is improved by keeping a more parsimonious set of water chemistry variables. The relative influence of each predictor, the mean squared error and the percentage deviance of BPUE explained by the BRT models were also computed to serve as assessment of the modelling quality.

3.4 Results

Table 3.5a and 3.5b summarises the BRT results for the four sets of predictors. These models explained deviance in BPUE ranging from 19.71% to 77.86%, a fairly good explanation with a considerably broad range. Figure 3.3 shows the explained deviance variation across different sets and freshwater fish species. Overall: (i) lower explained deviances were obtained for the smallmouth bass models, with the minimum deviance value seen for the full model; (ii) higher deviances were obtained from walleye models, with the maximum value seen for walleye BPUE being predicted by the set 4 of predictors (when water chemistry was represented by DOC, pH and total phosphorus). Mixed results regarding which set of predictors better predicted BPUE are seen when deviance results were compared within species. For lake trout and white sucker, the full model showed the best results; for lake whitefish and smallmouth bass, the model using the set 3 of predictors was the best performing, while for walleye and northern pike, it was the model using the set 4 of predictors that showed optimal results. However, among the subsets of predictors from the full model used to predict BPUE (represented by sets 2-4), in average, the set 4 was the one presenting the higher deviance values (50.5%, against 46.9% for set 2 and 48.2% for set 3).

Graphical calibration results were expressed in terms of MSE/Mean BPUE for each freshwater fish species. Results (Figure 3.4) indicated that smallmouth bass models were the most precise (MSE/Mean BPUE 0.217-0.242), and lake whitefish models were the least precise (MSE/Mean BPUE 0.629-0.711). On average, the BPUE models that showed the best precision were the full models (mean MSE/Mean BPUE = 0.4998) and the ones using the set 4 of predictors (mean MSE/Mean BPUE = 0.5061), while the models using the set 2 of predictors performed the worst (mean MSE/Mean BPUE = 0.5502).

Table 3.5a: Variable relative importance, mean squared error and adjusted deviance obtained from boosted regression trees developed to predict BPUE, for each freshwater fish species

Independent Variables		Cold Water Species				Cool Water Species						Warm Water Species	
		Lake Whitefish		Lake Trout		White Sucker		Walleye		Northern Pike		Smallmouth Bass	
		Full Model	Set 2*	Full Model	Set 2*	Full Model	Set 2*	Full Model	Set 2*	Full Model	Set 2*	Full Model	Set 2*
Climate	MAT 8110	10.32	11.81	8.74	14.49	13.50	14.89	34.48	38.35	19.24	26.72	8.36	10.58
	DD5 8110	16.93	16.63	7.11	9.41	16.90	20.69	6.49	7.29	7.42	10.77	9.54	10.81
Lake Morphometry	log10 Area	14.39	16.49	4.23	5.84	6.02	7.14	8.43	10.11	5.97	9.98	10.48	14.37
	Depth Max	5.37	7.01	4.19	7.27	14.30	15.31	7.52	9.20	12.92	17.34	5.81	7.75
	Depth Mean	4.29	6.57	8.46	12.93	11.00	11.70	8.54	9.61	8.65	11.45	16.67	15.78
	SDF	19.13	21.47	4.76	8.87	6.12	7.98	4.95	6.04	7.23	9.42	8.36	10.6
Water Chemistry	Total Phosphorus	3.53	E	8.84	E	3.65	E	5.98	E	12.10	E	9	E
	Total Dissolved Solids	6.34	6.60	3.61	4.20	4.07	4.36	3.38	4.08	3.10	3.80	11.48	12.56
	Secchi	5.25	9.25	3.41	10.69	3.75	5.60	4.71	7.04	1.45	3.20	5.2	10.05
	DOC	9.61	E	14.39	E	6.04	E	6.07	E	8.20	E	6.68	E
	pH	3.41	4.16	31.06	26.31	10.63	12.31	6.47	8.27	4.66	7.31	5.75	7.5
Angling Score		1.42	E	1.18	E	4.02	E	2.98	E	9.06	E	2.67	E
	MSE	0.700	0.683	0.231	0.354	0.856	0.872	1.056	1.242	0.437	0.482	0.085	0.077
	Mean BPUE	1.004	1.004	0.699	0.699	1.303	1.303	1.827	1.827	0.882	0.882	0.350	0.350
	Deviance Explained	52.52	53.30	58.62	34.13	54.70	53.57	74.56	69.19	49.48	43.78	19.71	27.35

E: Variable excluded for given set of variables

* This set of variables correspond to the same used to develop presence-absence models in Chapter 2.

Table 3.5b: Variable relative importance, mean squared error and adjusted deviance obtained from boosted regression trees developed to predict BPUE, for each freshwater fish species

Independent Variables		Cold Water Species				Cool Water Species						Warm Water Species	
		Lake Whitefish		Lake Trout		White Sucker		Walleye		Northern Pike		Smallmouth Bass	
		Set 3**	Set 4***	Set 3**	Set 4***	Set 3**	Set 4***	Set 3**	Set 4***	Set 3**	Set 4***	Set 3**	Set 4***
Climate	MAT 8110	10.54	11.11	12.33	12.38	13.34	12.92	36.94	33.03	24.61	19.37	10.05	8.21
	DD5 8110	17.10	17.40	10.16	7.03	19.99	18.06	6.86	7.81	9.44	7.81	10.17	13.05
Lake Morphometry	log10 Area	15.46	17.45	5.40	3.70	6.97	6.24	9.77	9.12	9.32	7.54	12.03	11.04
	Depth Max	7.02	5.54	7.03	4.10	16.04	15.43	9.21	8.17	16.06	13.16	7.75	7.62
	Depth Mean	6.93	5.78	14.48	9.58	10.69	10.75	9.50	8.59	10.84	9.63	16.18	18.27
	SDF	21.59	20.45	8.64	4.75	6.93	7.08	5.83	5.93	8.40	7.12	10.8	12.18
Water Chemistry	Total Phosphorus	E	4.30	E	8.57	E	4.67	E	7.21	E	11.98	E	10.22
	Total Dissolved Solids	6.79	E	5.09	E	4.51	E	3.71	E	4.02	E	12.62	E
	Secchi (Spring)	9.41	E	9.59	E	6.10	E	6.86	E	3.22	E	9.64	E
	DOC	E	11.85	E	19.29	E	7.30	E	8.06	E	8.40	E	8.76
	pH	3.62	4.49	25.85	29.06	11.06	12.90	7.48	8.32	6.54	5.97	7.33	7.8
Angling Score		1.54	1.62	1.42	1.53	4.37	4.66	3.82	3.76	7.56	9.01	3.44	2.84
	MSE	0.632	0.714	0.340	0.313	0.873	0.869	1.222	0.918	0.485	0.429	0.076	0.078
	Mean BPUE	1.004	1.004	0.699	0.699	1.303	1.303	1.827	1.827	0.882	0.882	0.350	0.350
	Deviance Explained	57.11	51.22	36.98	42.24	53.60	53.82	69.89	77.86	43.51	50.24	28.42	27.47

E: Variable excluded for given set of variables

**Water chemistry variables consisted of Secchi (Spring), Total Dissolved Solids and pH.

***Water chemistry variables consisted of DOC, Total Phosphorus and pH.

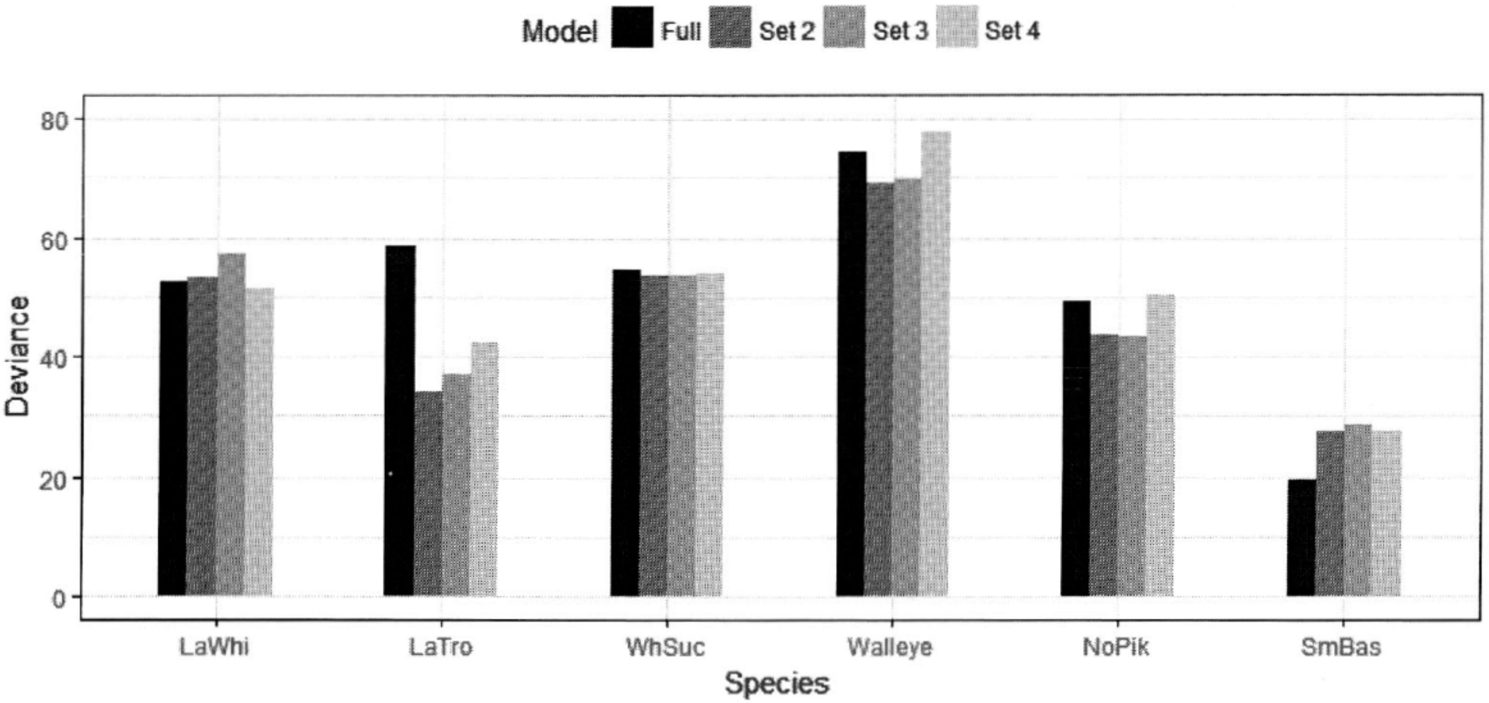

Figure 3.3: Explained deviance from boosted regression trees analyses to predict BPUE of six freshwater fish species.

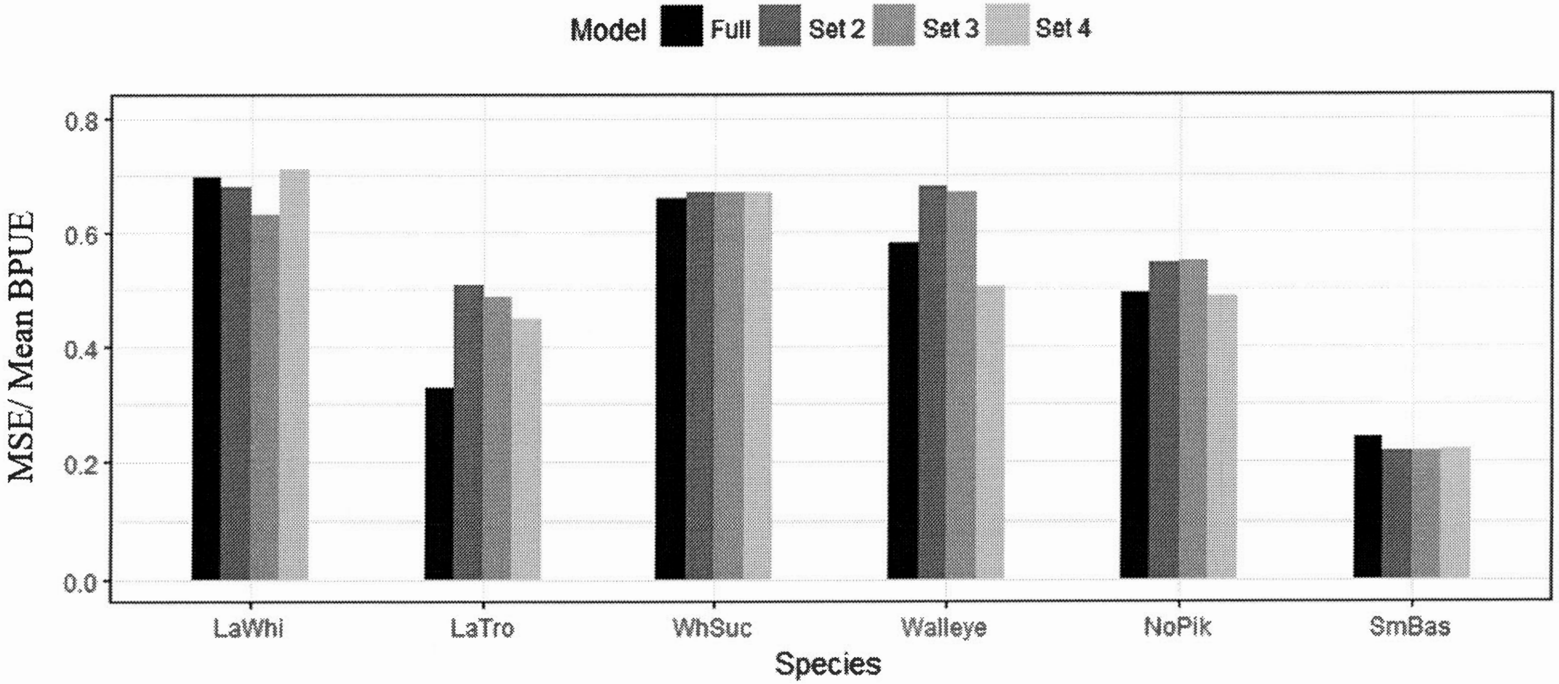

Figure 3.4: Mean squared error (MSE)/ Mean BPUE estimated from boosted regression trees analyses to predict BPUE of six freshwater fish species.

3.4.1 Environmental determinants of species biomass

Figure 3.5 presents the relative importance of variables in each species biomass model, based on the full set of predictors. Models developed for cold water species showed differences in the most important predictors. For lake trout, water chemistry was the most important environmental feature (61.31%; contributions of each predictor showed in Table 3.5a), followed by lake morphometry (21.64%) and climate (15.85%); pH was the most important variable (31.06%). For lake whitefish, lake morphometry is most important (43.18%), followed by roughly equal contributions from water chemistry (28.14%) and climate (27.25%); SDF was the most important variable (19.13%).

Mixed results were also observed for cool water species. Lake morphometry was the most important predictor for BPUE of white suckers (37.44%) and northern pike (34.77%), while climate was the primary predictor for walleye BPUE (40.97). However, the single most important variable for all cool water species was a climate variable: DD5 for white sucker (16.90%), and MAT for northern pike (19.24%) and walleye (34.48%).

Smallmouth bass, the single representative of warm water species, was most influenced by lake morphometry (41.32%), closely followed by water chemistry (38.11%) then climate with 17.9%; Mean depth was the single most important variable (16.67%).

Angling Score was universally low as a predictor of BPUE. Angling Score showed some influence on northern pike biomass, where it ranked as the fourth most important variable (9.06%). Biomass models for other species ranked Angling Score as last in all cases, except white sucker.

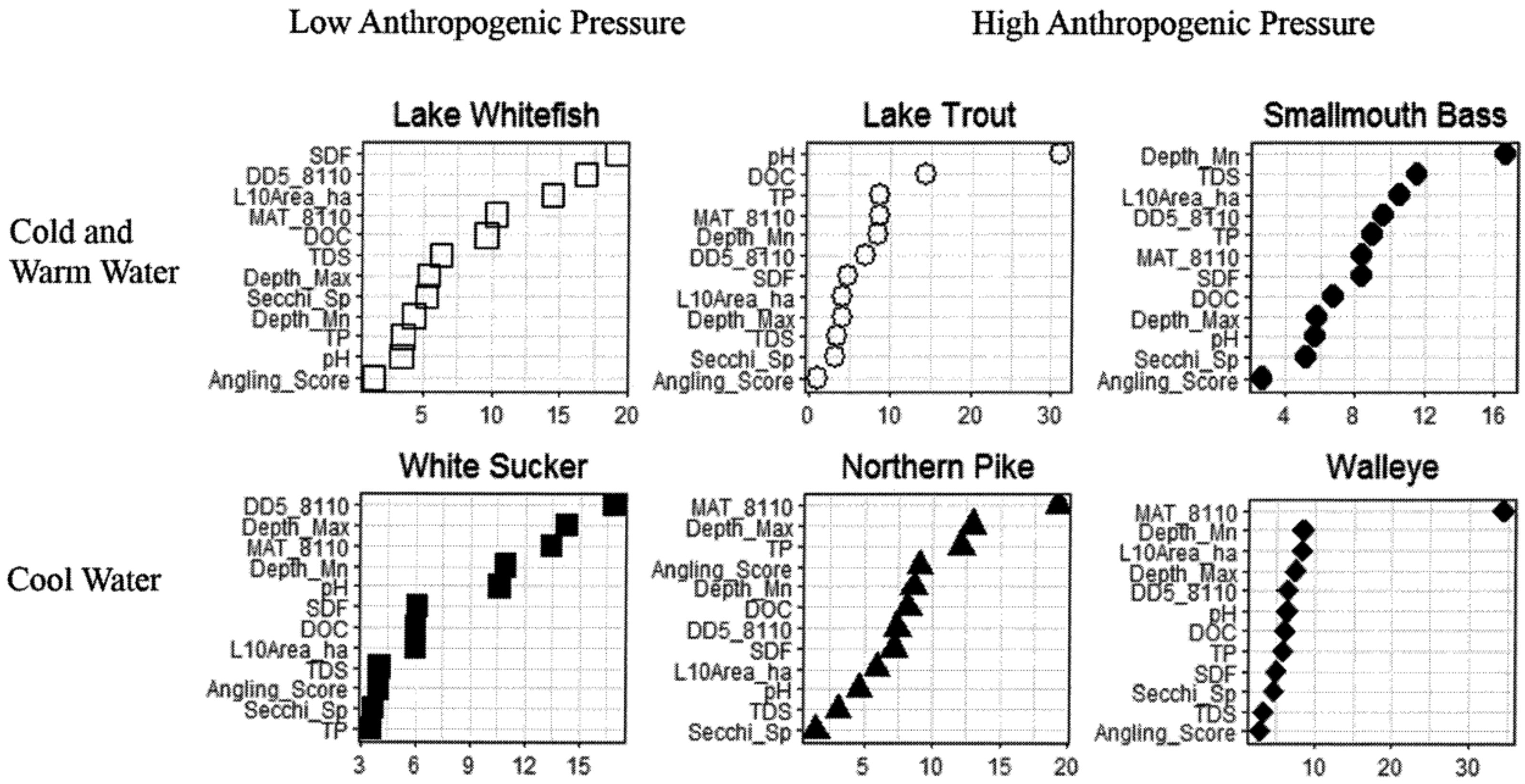

Figure 3.5: Variable relative importance plots generated from boosted regression trees analyses to predict BPUE of six freshwater fish species, using the full set of predictors.

3.4.2 Comparison of biomass and occurrence models.

Figure 3.6 shows the variable relative importance for the six freshwater fish species' models using the set 2 of predictors, which is the same used to estimate presence-absence of the same six fish species in Chapter 2, and the purpose here is to both verify (1) in which way variable relative importance changes in comparison to the full model and (2) make a parallel with the results shown in Chapter 2, identifying the existence of any similarities between the environmental determinants which are important to predict BPUE of the six fish species, and the ones which are important to predict presence-absence of the same species.

Cold water species' biomass models continued to display differences in respect to the most important environmental features, but in the same way as seen on the full models' results. Water chemistry was still the most important environmental feature for lake trout (41.20%, Table 3.5a), followed by lake morphometry (34.91%) and climate (23.90%); pH remained the most important variable (31.06%). For lake whitefish, lake morphometry was still the most important environmental feature (51.54%), then climate (28.44%) and water chemistry (20.01%). In addition, SDF remained as the most important variable in the prediction of lake whitefish biomass (21.47%).

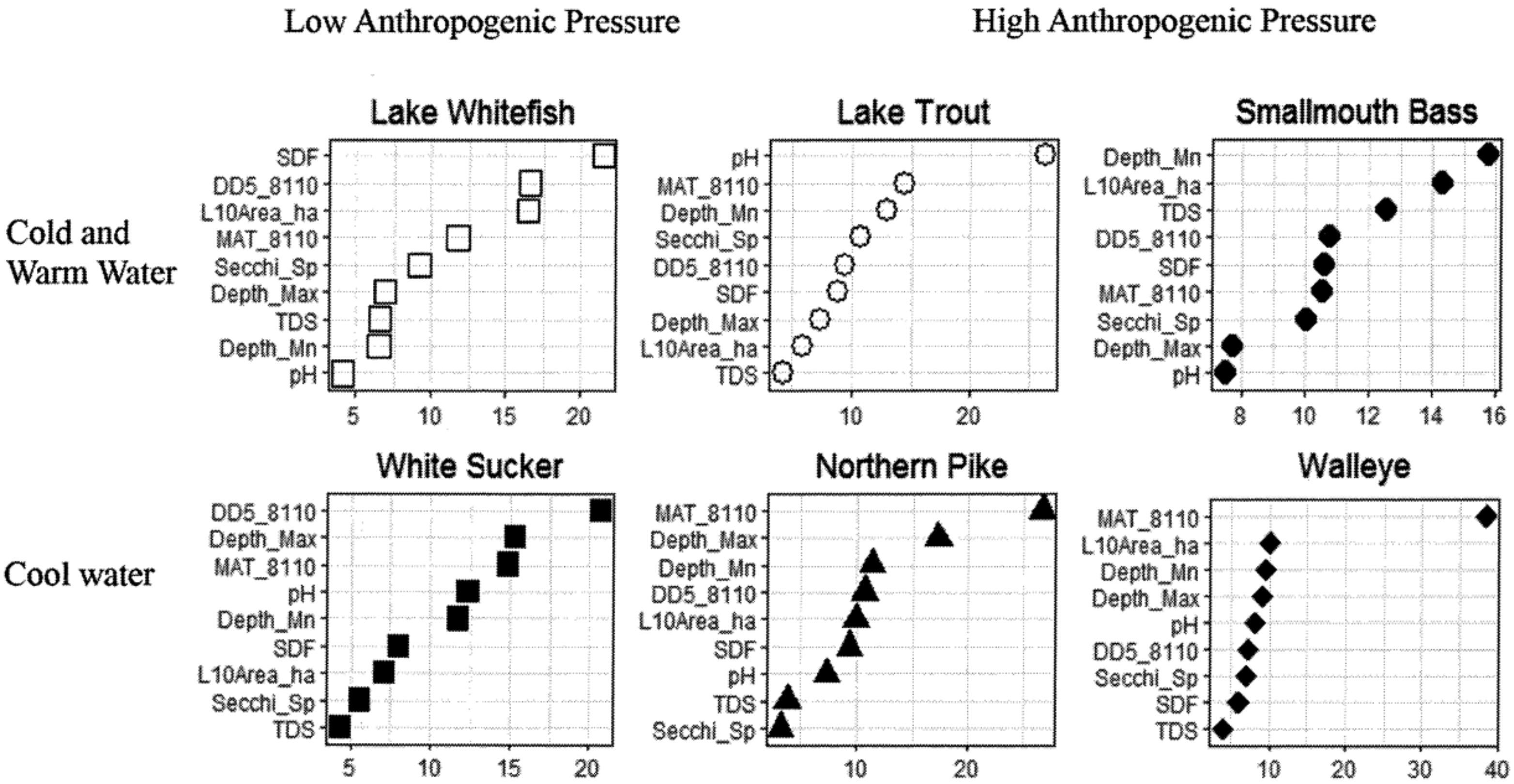

Figure 3.6: Variable relative importance plots generated from boosted regression trees analyses to predict BPUE of six freshwater fish species, using the set 2 of predictors.

Results from Chapter 2 indicated lake morphometry as the most important determinant in predicting the occurrence of lake trout and lake whitefish, with Max Depth and Log10Area, respectively, being the most important variables. In the case of lake whitefish, climate also played an important role. For lake whitefish, the biomass model (Figure 3.6) roughly agrees with the occurrence model: lake morphometry and climate were the most important determinants. For lake trout, however, the two models did not agree: water chemistry was the primary predictor of biomass.

For cool water species, lake morphometry continued to be the most important environmental feature for white sucker and northern pike biomass (42.13% and 48.19% respectively, Table 3.5a); climate continued to be the most important for walleye (45.64%). Like seen in the full models, the most important variables for the three cool water species' biomass were climate ones: DD5 for white sucker contributing 20.69%, and MAT for northern pike and walleye contributing 26.72% and 38.35%, respectively.

The models built in Chapter 2 to predict presence-absence of the three cool water species showed lake morphometry through log10 area to be the most important variable, followed by climate through MAT. It is clear from these results that these two environmental determinants are fundamental in determining both occurrence and biomass of cool water species, however the results suggest that lake size is the most important variable shaping occurrence of cool water species, while climate variables are the most important for biomass (Figure 3.6).

Finally, smallmouth bass results were similar to the full model ones, following the same trend seen for cold and cool water species. The most influential environmental determinant for this species' biomass was still lake morphometry (48.50%; Table 3.5a), followed by water chemistry (30.11%) and climate (21.39%); Mean Depth was the most important variable (15.78%), followed by a roughly equal contribution of Log10Area (14.37%). Chapter 2 results indicated both climate through MAT and lake morphometry through Log10Area to be important predictors of smallmouth bass occurrence, which suggests that these two environmental features are important in shaping this species' presence-absence, while its biomass assessed via BPUE is primarily driven by lake morphometry.

3.4.3 Seeking model parsimony: a comparison between models using the set 3 and 4 of predictors.

Figure 3.7 presents the variable relative importance for the six freshwater fish species' models using the set 3 of predictors, which consists of water chemistry being represented by pH, Secchi Depth and TDS, and Figure 3.8 presents the results using the set 4 of predictors, consisting of water chemistry being represented by pH, Total Phosphorus and DOC. The results of the two models were compared to determine which set of water chemistry predictors provides a better model parsimony, consisting of a model optimization.

Comparison of results from sets 3 and 4 indicates that an increase in the water chemistry contribution occurred when TDS and Secchi Depth were replaced by DOC and TP. The greatest increase occurred for lake trout (from 40.53% to 56.92%). The sole exception was smallmouth bass, for which the water chemistry contribution decreased slightly (from 29.59% to 26.78%).

Figures 3.3 and 3.4 indicated that, in average, set 4 had the highest explained deviance and lowest MSE among sets 2-4. Therefore, these findings suggest that water chemistry represented by DOC and TP is a good parsimonious choice of variables for predicting biomass.

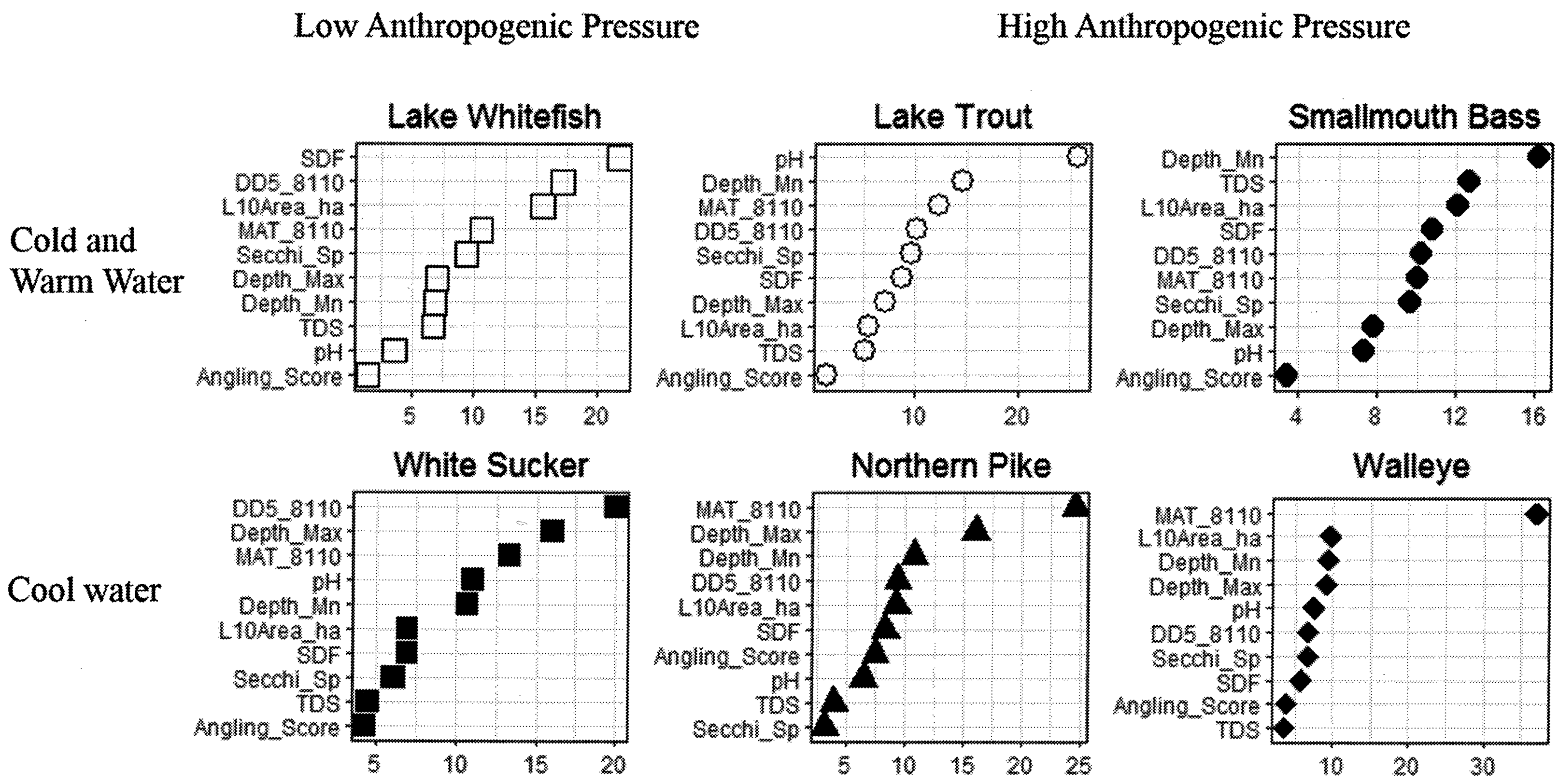

Figure 3.7: Variable relative importance plots generated from boosted regression trees analyses to predict BPUE of six freshwater fish species, using the set 3 of predictors.

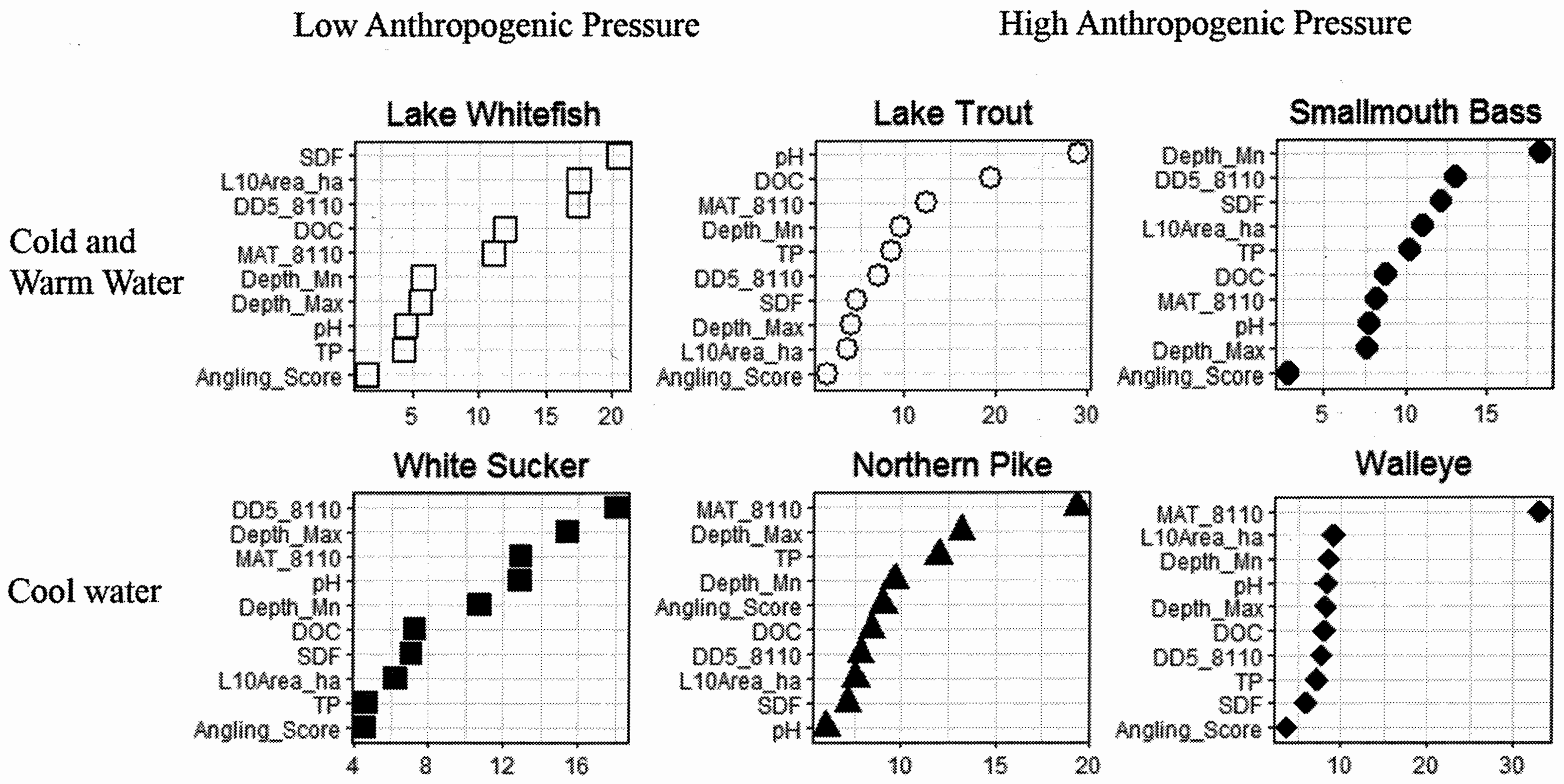

Figure 3.8: Variable relative importance plots generated from boosted regression trees analyses to predict BPUE of six freshwater fish species, using the set 4 of predictors.

3.4.4 Responses of BPUE from all six freshwater fish species according to the different environmental determinants

Partial dependency plots isolate the influence of individual predictor variables by fixing the variable of interest to a set value, predicting outcomes for all possible combinations of the other variables in the model, and averaging the predicted response. In this instance, the y-axis represents the probability of success in biomass production, which varies in response to the value of the predictor (represented on the x-axis). The shape of the relationship in the partial dependence plot therefore indicates how biomass patterns change as one moves along abiotic variables.

Partial dependence plots showing the responses of BPUE to the different environmental determinants are presented in Figure 3.9 (a to c). These results are based on the analysis of the predictor set 4 (Table 3.5b, Figure 3.9). Comparisons are made taking into consideration only the effect of each variable (positive or negative) on the biomass response accounted by the y-axis.

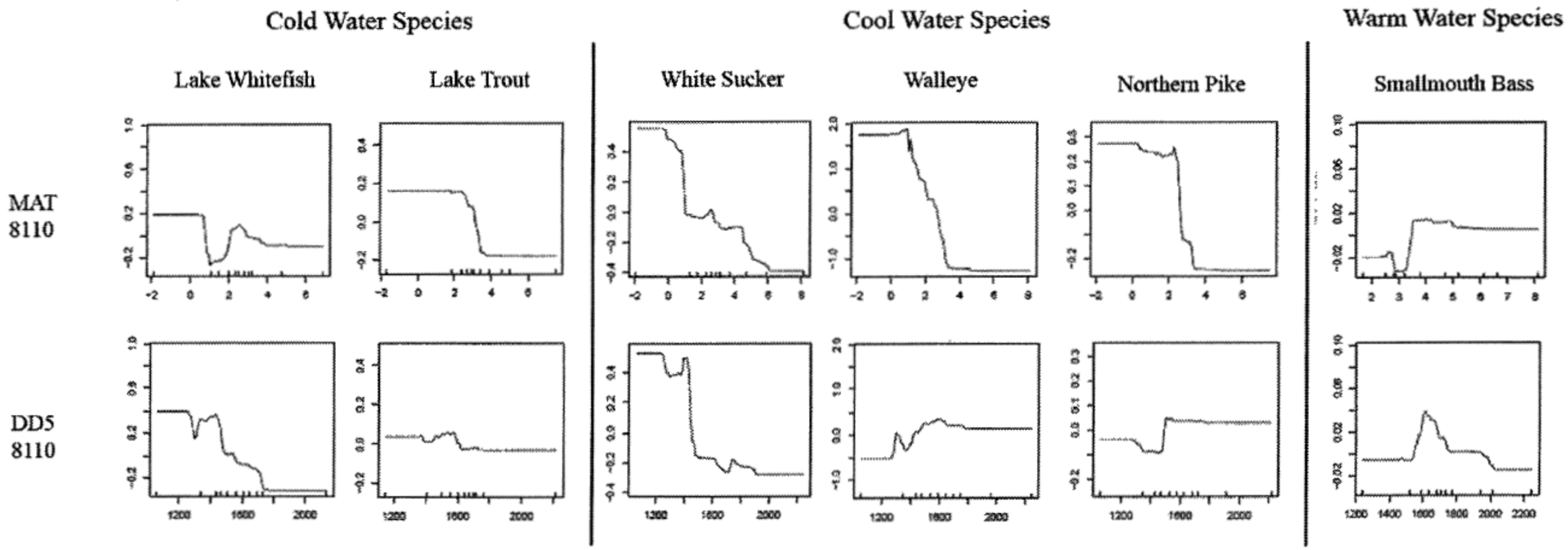

Figure 3.9a: Partial dependence plots generated from boosted regression tree analyses, showing the responses of BPUE to climate variables for cold water, cool water and warm water freshwater fish species.

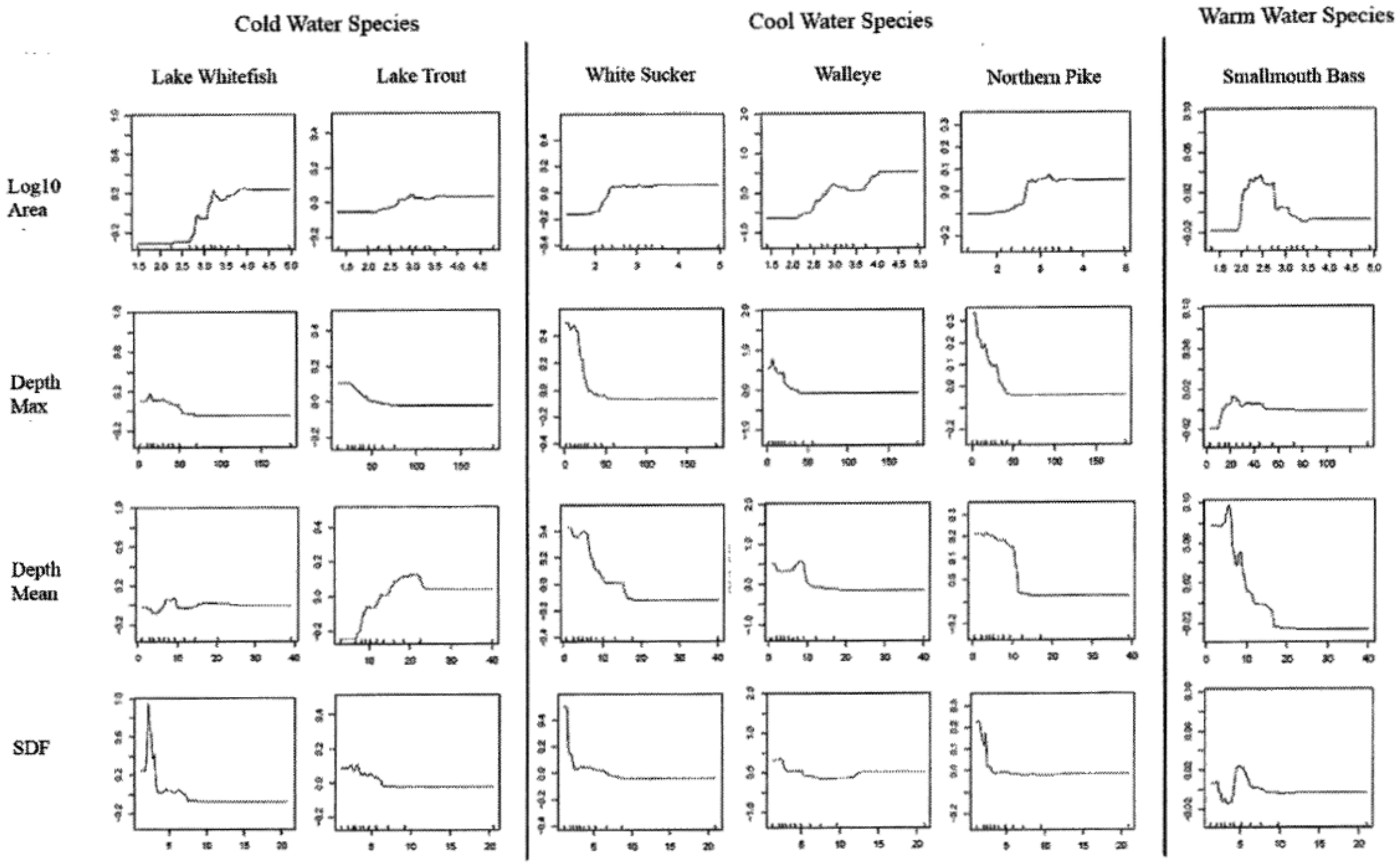

Figure 3.9b: Partial dependence plots generated from boosted regression tree analyses, showing the responses of BPUE to lake morphometry variables for cold water, cool water and warm water freshwater fish species.

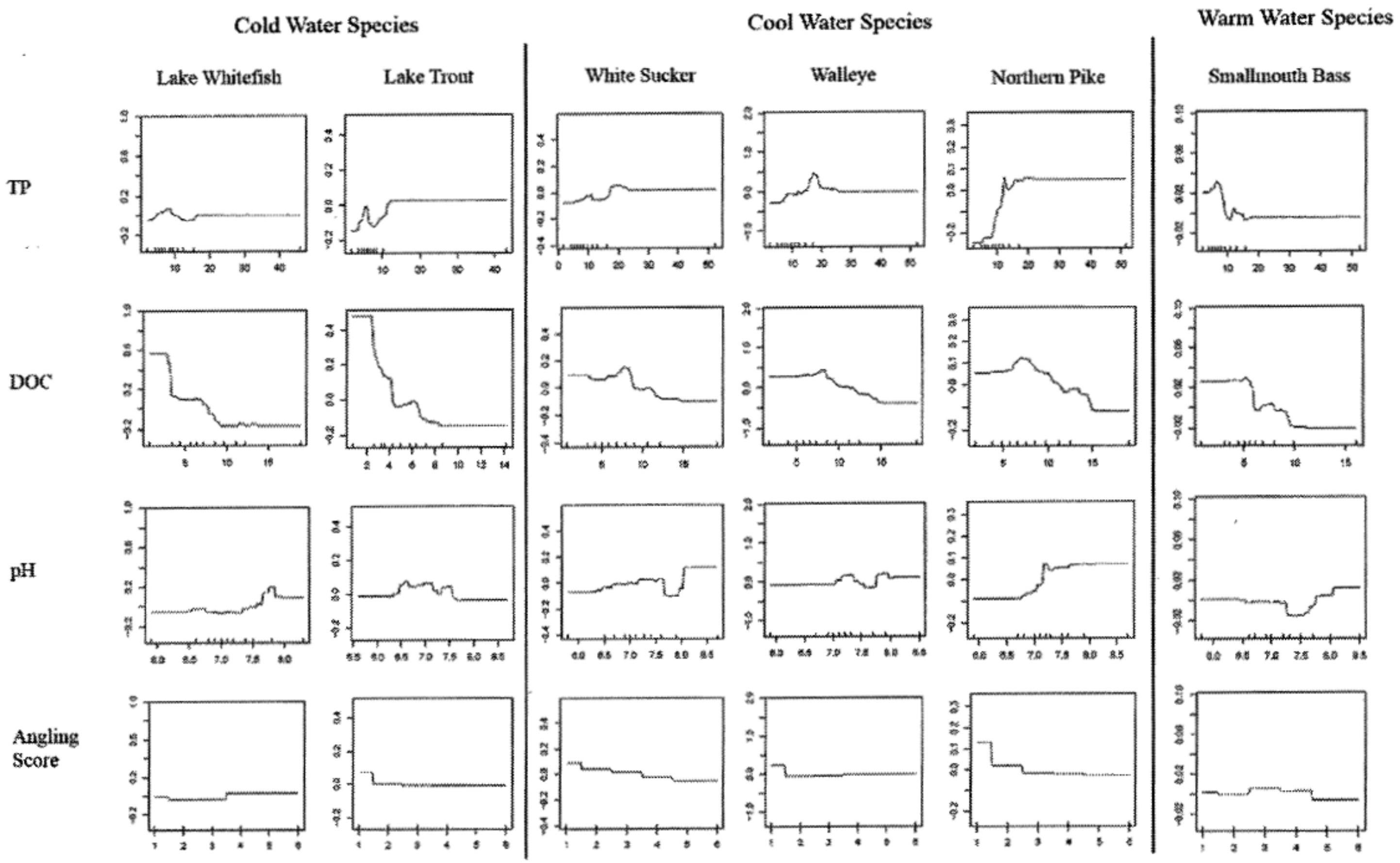

Figure 3.9c: Partial dependence plots generated from boosted regression tree analyses, showing the responses of BPUE to water chemistry variables and angling score for cold water, cool water and warm water freshwater fish species.

3.4.4.1 Climate

Climate clearly has different effects on the biomass of species in different thermal guilds (Figure 3.9a). As expected, the biomass of cold and cool water species declines as temperature increases, while the warm water species increases with temperature.

a) Cold water

MAT had a smaller effect on cold water species, but a decreasing trend was still evident. For lake trout, biomass declines rapidly when MAT exceeds 2°C. Lake whitefish show a less clear response to MAT, but a steady decline with DD5.

b) Cool water

MAT exerted a greater influence on cool water species, with BPUE showing a sharp decrease around 1°C for white sucker and walleye, and around 2°C for northern pike. DD5 showed a less consistent trend: white sucker biomass declined steeply when DD5 exceed 1400 growing-degree days; walleye and northern pike biomass increased with DD5, reaching a plateau when DD5 exceeded 1500 growing-degree days.

c) Warm water

For smallmouth bass, MAT showed an increasing influence around 3°C, that rapidly stabilizes. Biomass for values of MAT below 2°C were zero, which explains the lack of response (and consequent representation of the x-axis starting from 2°C). DD5 influence on biomass shows a moderate peak in response that can be seen around 1600 growing-degree days, rapidly declining for higher values of this variable.

3.4.4.2 Lake morphometry

Lake morphometry influence on BPUE response for the three thermal guilds is shown on Figure 3.9b.

a) Cold water

Results for this thermal guild showed that lake area influences the biomass of lake whitefish more than lake trout. Lake whitefish biomass declines rapidly when Log10Area falls below 3.2 (i.e. area = 1500 ha). Lake trout biomass also declines in smaller lakes, but to a much lesser extent. The response to maximum depth indicates an optimum around 30m for both cold water species, with a slight decrease beyond that value. This response to depth is less dramatic than in the case of cool water species. Mean depth had little influence on lake whitefish biomass, but a strong effect on lake trout with optimum values above 20 m. A rising trend could be noticed starting at about 5m, reaching its optimum at about 20m. For both species, biomass declined with increasing SDF (i.e. higher shoreline complexity). This effect was more pronounced in lake whitefish.

b) Cool water

All three cool water species in this study responded significantly to variations in lake morphometry parameters. BPUE showed an increasing trend with Log10Area, with walleye showing the most expressive response, which represents the importance of large lakes to the biomass of this species. Looking at the effect of depth max and mean depth, BPUE values showed a sharp decline with increasing depth, starting at around 10m. Northern pike was less affected by increasing depths but its optimum is clearly seen in non-stratified lakes. Regarding the effect of SDF on BPUE results of the three species, it affected white sucker and northern pike more, both showed decreasing trends with increased lake complexity. Walleye BPUE was quite unaffected by this last lake morphometry parameter.

c) Warm water

Smallmouth bass BPUE responded remarkably to variations in lake morphometry predictors, with a more expressive influence of mean depth on BPUE response. The influence of Log10Area on biomass initially showed a moderate increasing trend at 2.0 (100 ha) followed by a decrease at 2.7 (500 ha). Depth max did not show a notable effect on biomass, a low peak could be seen around 20m but quickly estabilized with increasing depths. The highest values of biomass could be seen in shallow lakes presenting mean depth of around 5m, followed by a sharp decrease with increasing mean depth. This result suggests a preference of this species for shallow lakes. Finally, the influence of SDF on smallmouth bass BPUE shows a drop around 3, with a moderate peak around 5, then decreasing again and estabilizing for more complex lakes. This result denotes a preference of this species for lakes with low complexity, but not necessarily lakes with shapes similar to a circle.

3.4.4.3 Water chemistry

The influence of DOC, TP and pH on the BPUE response of the three thermal guilds is shown on Figure 3.9c.

a) Cold water

The water chemistry parameter that had the most influence on cold water species was DOC; biomass exhibited a sharp decline with increasing values of this parameter. TP had little effect on lake whitefish, but a more pronounced effect on lake trout. Lake trout biomass initially increased with TP, reached an initial peak around TP = 5ug/L. It then declined slightly before rising to a plateau when TP exceeded 20 mg/L. The effect of pH on lake whitefish biomass shows a peak between 7.5-8, and some variation was displayed for lake trout biomass between 6.5-7.5.

b) Cool water

TP did not exhibit a large influence on the biomass of white sucker and walleye, but a biomass peak occurs around 20 ug/L. For northern pike, biomass increased more rapidly with TP and plateaued when TP exceeded 20 mg/L. High DOC values had a negative effect on the biomass of cool water species, but little effect when DOC was less than 7 mg/L. On average, pH had a slight positive effect on the biomass of cool water species, although less so for walleye.

c) Warm water

TP had a moderate effect on the biomass of smallmouth bass, with a sharp drop after 5 ug/L and stabilization after around 15 ug/L. DOC also has a moderate effect on smallmouth bass, declining when DOC exceeds 5 mg/L. pH exerted low influence on BPUE, with a slight dip around 7.5.

3.4.4.4 Angling score

The influence of angling score on the BPUE response of the three thermal guilds is shown on Figure 3.9c.

a) Cold water

Angling score did not display any remarkable impact on the BPUE of both cold water species.

b) Cool water

Fishing pressure inspected through angling score did not show any remarkable influence on the BPUE of white sucker and walleye, however, a moderate impact could be noticed on northern pike BPUE, with a biomass decrease starting at values above 1.

c) Warm water

Angling score did not exhibit any remarkable influence on the biomass of smallmouth bass.

3.5 Discussion

The main purpose of this study was to develop biomass models for six freshwater fish species present in Ontario lakes, based on climatic, morphometric and water chemistry variables. Climate and exploitation were the main factors dictating the choice of species. Lake trout, walleye, northern pike and smallmouth bass were selected because they are popular targets of recreational fishing, and white sucker and lake whitefish receives very little stress from angling. The inclusion of harvested and non-harvested species would help in describing the impact of fishing.

Four models were built using different sets of predictors: the first model was developed using 12 predictor variables representing climate, lake morphometry, water chemistry and fishing effort; a second model was based on a reduced set of predictor variables which matched the set of predictors used in Chapter 2 to develop species occurrence models. This model was used to compare the relative importance of variables used to predict species occurrence and biomass. Two additional models were built to compare alternative measures of water clarity (e.g. Secchi and DOC) and nutrients (TDS and TP). In Chapter 2, occurrence models were built using Secchi and TDS because detailed water chemistry data (including measures of DOC and TP) was not conducted when the data from 10,000 Ontario lakes was collected. More recent surveys, conducted by Ontario's Broad-scale Monitoring program, include detailed water chemistry making TP and DOC available for analysis. We tested whether use of TP and DOC (instead of TDS and Secchi) was more effective in predicting fish biomass by comparing a third model (based on TDS and Secchi) to a fourth model (based on DOC and TP). Both the third and fourth models included a common set of other variables. This comparison indicated TP and DOC was slightly better than TDS and Secchi, so the fourth model was chosen for exploring the effects of each environmental variable. These effects were described in a series of graphs (i.e., partial dependence plots), showing how biomass of each species responds to variation in environmental variables.

Overall, the biomass models presented in this study showed satisfactory results, being able to showcase the main environmental determinants influencing each freshwater fish species, with consistency across different model types. The ability of the different models in explaining the variation of biomass evaluated through the deviance results was quite variable (19.71 – 77.86%), with an average of 49.30%

across all models, which consists of a moderate degree of explanation. Boosted regression trees are becoming increasingly popular in the literature given their ability to: (i) cope with uncertainty in model structure; (ii) show strong predictive performance; (iii) capture non-linearity in the response to individual predictors and interactions among predictors (Elith et al. 2008; Bond et al. 2011). Additionally, it consists of a flexible modelling approach since it also showed satisfactory results in Chapter 2 when modelling species occurrences, which are represented by a binary response variable.

It is important to mention that in using BPUE as an index of biomass density, we are assuming that BPUE is proportional to biomass density (i.e., biomass/area) multiplied by a catchability coefficient (commonly addressed as *q*). Catchability varies across species because some species are more vulnerable to be captured by gillnets, but here we assume that *q* is constant across lakes for a given species. However, gillnetting is size selective, which can add a potential bias in using BPUE as an index of biomass. A future perspective of this work is to improve the biomass models built here by adjusting BPUE values considering correction factors for size bias.

One of the objectives of this chapter was to address the level of agreement between the biomass models developed in this chapter and the occurrence models developed in Chapter 2, in terms of important variables shaping biomass and distribution of fish species. Altogether, results from both chapters indicate that both climate and lake morphometry are important environmental determinants shaping occurrence and biomass, showing that occurrence and biomass models show a high level of agreement. The sole exception was noted for lake trout, a cold-water species, whose occurrence is influenced by lake morphometry while biomass is determined by water chemistry. To the best of our knowledge, no previous study assessed the degree of agreement between occurrence and biomass models by identifying which abiotic features are important in shaping occurrence and biomass, while using the same set of abiotic variables in both models. The literature possesses vast information about occurrence and biomass/production models for the species addressed here, but these models were built using varying numbers of predictors or different modelling approaches, which makes the comparison task difficult to be carried out.

The optimization suggested by removing Secchi and TDS from the full model had varying impacts for the different species, but in average it presented the best deviance among sets 2 – 4, coupled with the lowest MSE values. These are desirable modelling features. All the variables considered in this study consist of abiotic ones, which raise a question related to how improved these models would be if we had considered biotic variables in the modelling process, along the lines of the investigation presented in Chapter 2 for occurrence models. This extension of the modelling approach is a topic that could be further investigated.

Results indicated that biomass for each freshwater fish species is influenced by different environmental determinants, independent of the thermal guild which they make part. For instance, lake whitefish and lake trout are cold water species, however lake morphometry showed to be more influential on the biomass of this species, while water chemistry was more influential on the biomass of lake trout. In addition to the assessment of the main environmental determinants controlling the biomass of different species, some relationships can be established through inspection of interactions between variables which comprehend a given environmental determinant, allowing to assess their different contributions to the biomass response.

The final boosted regression trees models developed for all freshwater fish species demonstrate that climate, represented by MAT (mean air temperature) and DD5 (growing degree days above 5°C) are highly important in explaining the variation of biomass predicted via BPUE, with a higher effect seen on cool water species. Figure 3.9a shows for northern pike and walleye that while a decreasing trend in BPUE happens with increasing values of MAT, an increase in BPUE is seen for increasing values of DD5, followed by a plateau. Figure 3.10-11 allow a further investigation of MAT and DD5 interaction.

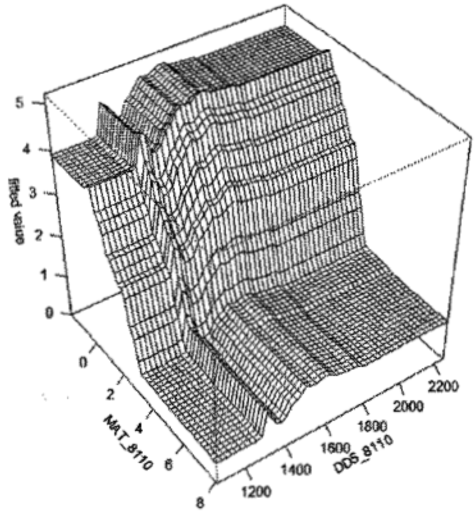

Figure 3.10: Perspective plot generated from the boosted regression tree model developed for walleye, showing the responses of BPUE to the interaction between MAT and DD5.

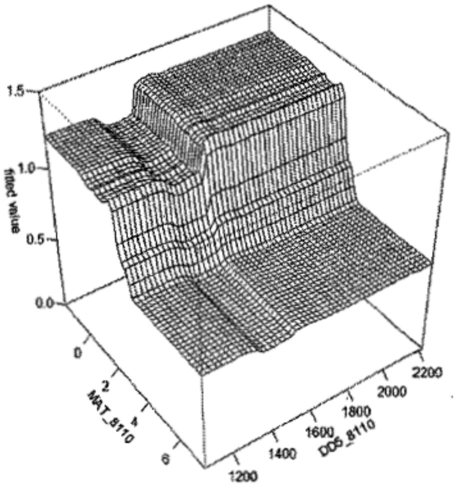

Figure 3.11: Perspective plot generated from the boosted regression tree model developed for northern pike, showing the responses of BPUE to the interaction between MAT and DD5.

According to Figures 3.10-11, the highest values of BPUE for these two cool water species occur when MAT < 2°C and DD5 > 1600 growing-degree days, with MAT showing a negative effect beyond 2°C. Many studies have shown the influence of temperature on production of fish (Magnuson et al. 1979; Wismer and Christie 1987; Downing and Plante 1993; Chezik et al. 2013; Hansen et al. 2017); growth rate typically increases as temperature approaches an optimal value and then declines as temperature exceeds the optimum (Hanson et al. 1997). However, none of these studies have investigated the interaction among climate variables and their joint effect. The perspective plots presented by Figures 3.10-11 suggest that MAT determines a thermal threshold, while DD5, as a cumulative metric of temperature, encompasses the thermal niche of a species, which includes multiple life stages with differing thermal requirements (Wismer and Christie 1987). Additionally, DD5 may also encompass a myriad temperature influences in a single value and provide a measure of the metabolically relevant temperature experience of a fish that is difficult to quantify based on temperature optima.

Coupled with climate, our results also support the importance of lake morphometry to the biomass of both cold water and cool water fishes object of this study. Lake morphometry affects ecosystem size, that is directly related to food-chain length, species diversity, habitat availability and habitat heterogeneity. Overall, results demonstrated that biomass assessed through BPUE increased with lake size, markedly decreased with lake depth (with the exception of lake trout) and decreased with lake complexity. Large species require large home ranges (Minns 1995; Woolnough et al. 2008) and tend to be predatory (Romanuk et al. 2010). Physically larger ecosystems support more (the species-area relation; Rosenzweig 1995) and larger species, for both energetic and population dynamic reasons, while food chain length has been suggested to increase with ecosystem size and/or ecosystem productivity (Schoener 1989; Vander Zanden et al. 1999; Post et al. 2000; Thompson and Townsend 2005; Baiser et al. 2012). The results related to the mean depth influence on the biomass of all three thermal guilds reflect well their habitat preferences. For instance, lakes with mean depth of above 10 m seem to not support biomass of warm and cool water species, whereas these lakes promote biomass of lake trout, a cold-water species that prefers large, thermally stratified lakes that provide cold water refuge in the summer. Mean depth is a surrogate for hydrological characteristics such as thermal stratification, nutrient circulation and dilution, all of which affect how energy is processed within the water column (Chow-Fraser 1991). Mean depth also governs the amount of oxygen stored in the hypolimnion during stratification (Walker 1979), which plays an important role in defining the living space of cold water and cool water species, together with temperature. Finally, the negative influence of SDF on cool water and cold-water species is not obvious, which indicates that this predictor interacts with other variables to affect BPUE.

Inspection of Figure 3.9c revealed some interesting relationships among the water chemistry predictors and their influence on the biomass for all three thermal guilds. Biomass increased with increasing TP values, but decreased when DOC exceeded 8 mg/L (with the exception of smallmouth bass biomass, which decreased with small values of DOC). Figures 3.12-14 show the interaction between DOC and TP on the biomass of all three cool water species. The point which determines the threshold between the positive and negative effect of DOC can be clearly seen (approximately DOC = 8 mg/L and TP = 20 ug/L).

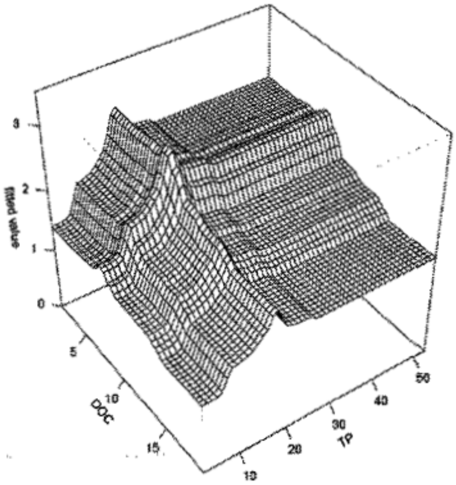

Figure 3.12: Perspective plot generated from the boosted regression tree model developed for walleye, showing the responses of BPUE to the interaction between DOC and TP.

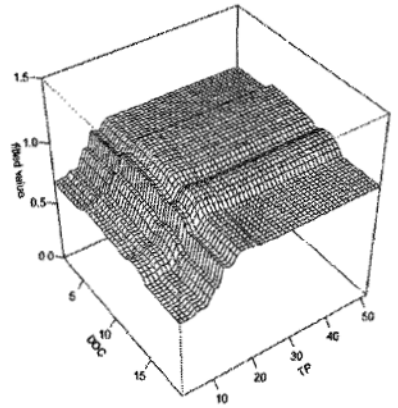

Figure 3.13: Perspective plot generated from the boosted regression tree model developed for northern pike, showing the responses of BPUE to the interaction between DOC and TP.

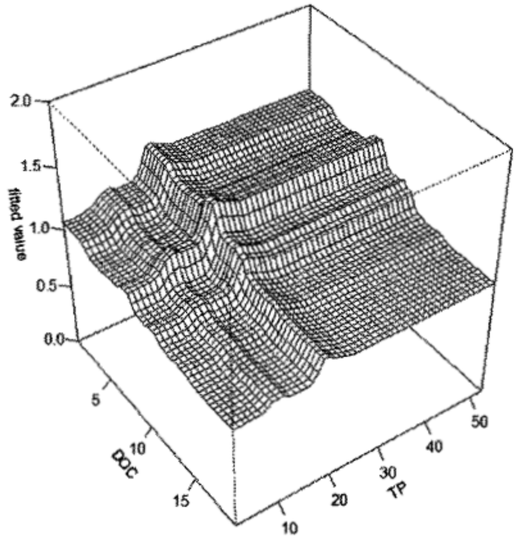

Figure 3.14: Perspective plot generated from the boosted regression tree model developed for northern pike, showing the responses of BPUE to the interaction between DOC and TP.

A plausible explanation for the existence of a threshold between the positive and negative effect of DOC is that increased input of nutrients increases production at the basic trophic level (phytoplankton), which also contributes to top consumer (fish) biomass (Karlsson 2009). On the other hand, benthic algae, which are responsible for benthic primary production, also play an important role in fish biomass (Karlsson 2009) and it is primarily dictated by light availability instead of nutrients, with light availability being regulated by DOC. The literature reports a direct relationship between TP and DOC (Tanentzap et al. 2014), so when DOC increases with TP, the biomass increase promoted by TP is offset by the increasing loads of DOC which directly interfere with water clarity, thus affecting light penetration and hindering benthic primary production. The initial positive response may depend on several factors, such as screening of harmful UV-radiation (Williamson et al. 1996), subsiding organic C to heterotrophic production (Cole et al. 2011) or organic P and N being associated with DOC. In boreal, oligotrophic lakes surrounded by pristine catchments DOC is also a major contributor to total P and N (Hessen et al. 2009). However, while N and P are mostly in organic form and thus of low bioavailability, some fractions of these pools will eventually mineralize and thus support primary production. In the case of cold water species, biomass declined rapidly as DOC increased (see Figure 3.9c). The reduction in light caused by higher DOC concentration is most likely explaining the strong negative influence that was observed for lake trout. In general, water clarity in lake trout lakes is controlled by DOC and thus, changes in optimal habitat resulting from a variation in TP would be much less significant than those resulting from DOC-induced modifications in water clarity (Clark et al. 2004), which may explain the higher impact of DOC compared to TP on biomass of lake trout. A model developed by Lester et al. (2004) shows the importance of light, together with temperature and bathymetry in controlling walleye production. Their model proposed the benthic area associated with optimum light and temperature conditions for walleye during an annual cycle (i.e. Thermal Optical Habitat Area, TOHA) as a predictor. When testing the applicability of their model in explaining known walleye yield, they found that little walleye habitat exists in really dark waters. The available TOHA increased rapidly as water clarity increased, but then declined exponentially after reaching what corresponds to the optimal water clarity. These observations also agree with the proposition of a threshold separating the positive and negative effects of DOC observed by Finstad et al. (2014).

Our water chemistry results revealed a high variable importance linked to pH in lake trout models, which was unexpected. The literature reports the sensitivity and past impact of acid waters to this species in a few Ontario lakes (Gunn and Keller 1990), as large lakes are more susceptible to the effects of acid rain promoted by industrial activities, which release acidifying substances to the atmosphere. Another reason that can explain the high importance of pH in predicting lake trout biomass is the interaction of pH and mean depth (see Figure 3.15. Depths of 20 m coupled with pH values of approximately 6.5 seem to offer optimal habitat for this species. The interaction plot demonstrates a peak in biomass occurs at these values of mean depth and pH.

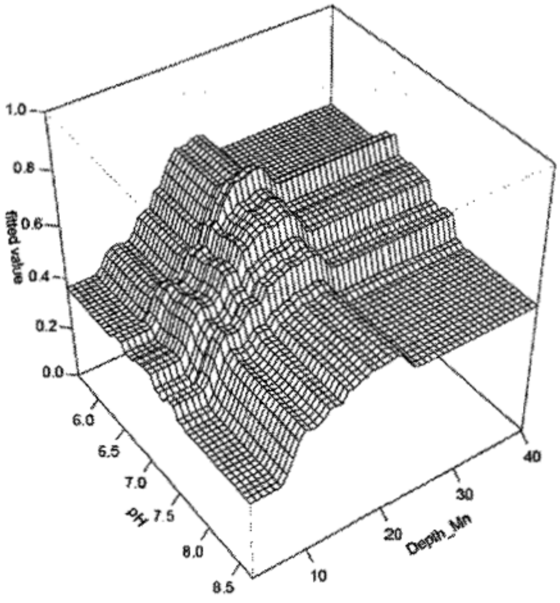

Figure 3.15: Perspective plot generated from the boosted regression tree model developed for lake trout, showing the responses of BPUE to the interaction between pH and mean depth.

We expected that angling score would have a larger variable importance in explaining the biomass of walleye, northern pike, lake trout and smallmouth bass, because they are popular species in recreational fishing. In addition, Table 3.5b shows that the contribution of angling score on white sucker biomass was higher than the one seen for walleye, even though white sucker is an unexploited species. Our results only support the moderate influence of this variable to northern pike biomass. The removal of angling score from the model using the set 4 of parameters (results not presented here) showed an increase on the other variables' contributions, but this increase did not happen only on climate variables as we hypothesized (since there is a correlation between angling score and climate variables). The quantification of angling score takes in consideration current stress, implying that current biomass of walleye, northern pike, lake trout and smallmouth bass will most likely be better explained by previous reports of angling pressure.

Even though the models' explained deviances for all six species were quite variable, our study provides a better understanding of the different factors driving biomass of the different species, while bringing to light the importance of lake morphometry and climate to predict both presence-absence and production of freshwater fish species from Ontario lakes. From the management point of view, these two environmental determinants should be primarily taken into account, not only when selecting new conservation areas but also when seeking an optimization of lake resources contributing to the sustainable practice of sport fishing. Since temperature is the most important abiotic factor controlling fish physiology (Brett 1979), projected increases in water temperature resulting from climate change are expected to affect the growth of freshwater fishes, their population dynamics, the ecology at community and ecosystem levels (Portner and Farrell 2008; Sheridan and Bickford 2011), change geographical distributions (Comte et al. 2013) and finally, production (Portner and Peck 2010). Increases in water temperature associated with climate change should benefit the growth of fishes with higher physiological thermal optima and stress the growth of those with lower (Graham and Harrod 2009). However, changes in temperature experienced by a fish may not be as great as changes in the mean water temperature of its ecosystem because the fish may behaviorally thermoregulate by selecting thermal habitats in which temperatures are favorable to growth (Coutant 1987). Hence, effects of climate change on fish growth may not be direct but may be mediated by availability of thermal habitat (King et al. 1999), which in turn largely depends on physical characteristics such as depth and trophic state of the ecosystem (Kling et al. 2003). Climate change may also indirectly affect fish growth through altering prey availability (Woodward at al. 2010). Regardless of the physiological thermal optimum, the growth of fish may decrease with increases in metabolic costs in a warming climate if prey consumption remains constant over time (Sheridan and Bickford, 2011), which would be an interesting aspect to be considered in future modelling studies focusing the same species.

CONCLUSION

Quantifying the relationships between the distributions of species and their abiotic and/or biotic environments has a long history in ecological research. While understanding where species occur is a fundamental ecological requirement, prediction of occurrence and quantification of its productivity are essential for much conservation and population management. With the increased availability of statistical packages, it is becoming increasingly easy to undertake species distribution modelling. This creates opportunities for the applied ecologists to develop management tools for conservation in a way that was unprecedented 15 years ago. Fish-habitat models can play an important role in prioritising surveys and monitoring programmes for fish populations because limitations to resources often preclude exhaustive and continual sampling of sites and that extensive sampling is needed to accurately sample lake fish communities (Jackson and Harvey 1997). Applications of predictive models in management include: (1) forecasting or measuring the effects of habitat alteration and changing land-use patterns (Oberdorff et al. 2001); (2) providing first-order estimates of habitat suitability to establish potential locations for re-introduction (Evans and Oliver 1995); (3) predicting the likelihood of local establishment and spread of exotic species (Peterson and Vieglais 2001); (4) predicting "hotspots" of species persistence for the conservation of biodiversity (Williams and Araujo 2000); and (5) revealing additional populations of threatened species, or alternatively revealing unexpected gaps in their range.

Under that context, the goal of this thesis was to conduct an investigation on improvement aspects in currently used modelling practices (more specifically species distribution models), while ultimately providing analytical tools to be applied in ecological assessments related to freshwater fish populations. Redundancy analysis coupled with variation partitioning, which have become core in the study of communities, were the focus of Chapter I. For years ecologists have been resorting to popular OLS approaches such as them, without taking into account the appropriateness of their usage, specially taking into account the nature of the species-environment relationship which frequently does not fulfill the required assumptions to conduct such analyses. To this end, the presented framework based on GLM solutions showed itself to be quite robust alternative, and it consists of a generalization of RDA, while widening the scope of applicability of variation partitioning, contributing to its popularity and importance when analyzing species data matrices.

The focus of Chapter II was to assess the importance of biotic interactions in SDMs, since the recognition of its importance just started recently with some publications addressing the impact of inclusion of this information in the modelling process, consequently helping to decrease uncertainty. To increase confidence in model projections, methodologies must acknowledge clearly the uncertainties involved and try to obtain conditional measurements of confidence in the forecasts made. Whereas confidence in model projections is likely to increase as the realism of key assumptions within a particular modelling approach are improved, tractability often decreases, due to more demanding parameterizations. Thus, a trade-off exists between complexity and tractability in modelling species and identifying the most reliable and unbiased solution is not a trivial task. The approach presented in Chapter II consists of a fairly simple one since it only required information from some species part of the same freshwater fish community, and it provided a decrease in uncertainty while keeping the simplicity of the different SDM models obtained. The degree of realism of key assumptions can surely be improved by the investigation of other ways of inserting biotic information in the models, and this consists of a future perspective for this work, besides the investigation related to the models' transferability. In addition, the comprehensive comparison of models' results in this chapter showcased the good performance of machine-learning methods, so future efforts in building "better" models (here "better" representing a compromise between complexity and tractability), should focus on this class of models. Another point worth mentioning is that, despite the fact that the models' performance was not the same, there was a certain agreement regarding the most important variables explaining the presence of the different species across models. That observation supports the idea that "there is no bad model, what actually exists is a badly parameterized model". Finally, one appealing avenue to pursue is to conduct a comparison of the results obtained in Chapter II with the ones derived from community-level models (Ferrier and Guisan 2006), establishing similarities and discrepancies between the results from these two model types. Community-level modelling would allow to include all species data available (instead of only the 6 species considered in Chapter II), while also evaluating several ways of conducting this analysis (e.g. community types, species groups, axes of compositional variation, macro-ecological properties – see Ferrier and Guisan 2006). The main advantages of this method are: (1) fast analysis of very large number of species, allowing a better cross-species synthesis; (2) adds value to data for rare species by "pooling"; (3) enhanced capacity to synthesize complex data into a form more readily interpretable by scientists and decision-makers and (4) address interactions between species. This last advantage would allow a more comprehensive exploration and analysis of biotic interactions in species distribution modelling.

Chapter III brought some of the information gotten from Chapter II about the model most successfully explaining deviance and aimed at the development of biomass models using biomass-per-unit effort information from the same six freshwater fish species mentioned in Chapter 2. Results of this chapter helped better understand the influence of the different environmental determinants used as predictors of biomass, allowing the assessment of their interesting interactions and how they control and limit each other while interfering in the biomass of the fish species object of that chapter. The knowledge obtained specific to each fish species can surely help management and conservation authorities to develop specific plans to help ensure the sustainability of these fish populations, by knowing their specific requirements and important environmental determinants helping the successful increase of their stock. Some future improvements to the models involve corrections that consider the bias involving the use of biomass-per-unit-effort, which consist of valuable future studies. The relationships between abiotic features and both species distributions and biomass, observed from models' results in Chapters II and III, contribute to information on which of these features have most impact in the distribution and biomass of all six studied species, which help set proactive conservation priorities for preserving potentially vulnerable fish populations across the province of Ontario.

These investigations contribute to major scientific underpinnings related to the research of species-habitat relationships, while consisting of ecosystem services when promoting information about the processes underlying these relationships. The investigation of interconnected quantitative frameworks to link environmental, spatial and biotic interactions bring to light a greater understanding of the key agents structuring biodiversity and how they interact to provide the delivery of aquatic ecosystem services, while clarifying about the actions that should be taken to mitigate the loss of these services in face of increasing human population.

Models are simplification of reality and often begin life by helping researchers to formalize their understanding of a particular process or pattern of interest. Models are thus primarily important aids to research. Difficulties may therefore arise when such theoretical models are used to guide conservation planning, management and to support the formulation of policy decisions. The magnitude of uncertainties in species' range and biomass assessments is currently so great that it might lead conservation planners, policy makers and other stakeholders to question the overall usefulness of science as an aid to solve real world problems. Bridging the perceived gap between science and societal needs is of paramount importance if one wants to make progress and contribute meaningfully to solve the global environmental change crises. The investigations pursued in this work raise additional questions about modelling aspects that help decrease uncertainty and thus help the conceptualization of even more useful tools that help understand the processes explaining species distributions and their productivity. Some ideas consist of: 1- incorporating spatial information in the models since this aspect was not investigated; 2- development of community biomass models that take into account the structure of the food web; 3- development of production indexes based on the biomass models presented, while incorporating information about the natural mortality and fishing mortality of each species. These surely represent interesting avenues of investigation for future studies.

APPENDIX A

CHAPTER I

Mean adjusted R^2 results for logistic model applied to presence-absence data, Poisson model applied to presence-absence data and Poisson model applied to abundance data, at population and sample level, according to varying sample sizes and number of inserted random N(0,1) covariates.

Scenario		Logistic P-A Pop	Logistic P-A Sample	Poisson P-A Pop	Poisson P-A Sample	Poisson Abundance Pop	Poisson Abundance Sample
Sample size	20	0.3543	0.29162	0.28349	0.18653	0.7039	0.67818
	50	0.3671	0.39666	0.27895	0.2656	0.7039	0.69013
	100	0.35983	0.38484	0.26691	0.26201	0.7039	0.6978
	150	0.35718	0.37439	0.26417	0.26076	0.7039	0.69937
	200	0.35653	0.36894	0.26356	0.26055	0.7039	0.70012
Number of inserted of random N(0,1) covariates	0	0.32241	0.34889	0.23891	0.23531	0.61786	0.61065
	5	0.32988	0.38767	0.24986	0.24366	0.61786	0.6136
	10	0.32923	0.40528	0.25969	0.25047	0.61786	0.61459
	15	0.31103	0.39847	0.25616	0.24035	0.61786	0.61227
	20	0.27078	0.36378	0.23257	0.20969	0.61786	0.61656

APPENDIX B

CHAPTER I

Estimated fractions in R^2 variation partitioning, under the logistic model applied to presence-absence data, at population and sample level according to varying sample sizes and levels of correlation between **X** and **W**.

Correlation level	Sample size	a_{pop}	a_{samp}	b_{pop}	b_{samp}	c_{pop}	c_{samp}	d_{pop}	d_{samp}
0.2	20	0.2265	0.0929	0.0019	-0.0233	0.2319	0.0932	0.5397	0.8373
	50	0.2466	0.2866	-0.0019	-0.0482	0.2528	0.2937	0.5026	0.4679
	100	0.2516	0.2974	-0.0105	-0.0432	0.2492	0.2958	0.5097	0.4500
	150	0.2476	0.2858	-0.0132	-0.0383	0.2448	0.2817	0.5208	0.4708
	200	0.2454	0.2756	-0.0138	-0.0336	0.2424	0.2715	0.5260	0.4865
0.4	20	0.2290	0.1024	0.0182	-0.0108	0.1713	0.0608	0.5815	0.8476
	50	0.2550	0.3031	0.0438	-0.0073	0.1902	0.2357	0.5110	0.4685
	100	0.2506	0.2908	0.0479	0.0200	0.2013	0.2392	0.5003	0.4500
	150	0.2483	0.2741	0.0470	0.0302	0.2033	0.2286	0.5014	0.4671
	200	0.2478	0.2670	0.0469	0.0348	0.2032	0.2212	0.5020	0.4771
0.8	20	0.0708	-0.0019	0.1984	0.1098	0.0911	0.0120	0.6397	0.8801
	50	0.0864	0.1351	0.2608	0.2073	0.0855	0.1360	0.5672	0.5216
	100	0.0879	0.1263	0.2807	0.2565	0.0853	0.1230	0.5461	0.4942
	150	0.0872	0.1163	0.2713	0.2561	0.0849	0.1143	0.5567	0.5133
	200	0.0864	0.1106	0.2639	0.2523	0.0840	0.1080	0.5657	0.5291

Estimated fractions in R^2 variation partitioning, under the Poisson model applied to presence-absence data, at population and sample level according to varying sample sizes and levels of correlation between **X** and **W**.

Correlation level	Sample size	a_{pop}	a_{samp}	b_{pop}	b_{samp}	c_{pop}	c_{samp}	d_{pop}	d_{samp}
0.2	20	0.1703	-0.0133	0.0320	0.0548	0.1795	-0.0112	0.6182	0.9697
	50	0.1622	0.1474	0.0476	0.0401	0.1716	0.1565	0.6186	0.6560
	100	0.1518	0.1493	0.0490	0.0469	0.1536	0.1515	0.6456	0.6522
	150	0.1467	0.1465	0.0472	0.0455	0.1478	0.1469	0.6583	0.6611
	200	0.1448	0.1447	0.0464	0.0450	0.1457	0.1452	0.6631	0.6651
0.4	20	0.1705	-0.0135	0.0458	0.0692	0.1248	-0.0492	0.6589	0.9935
	50	0.1723	0.1600	0.0800	0.0694	0.1218	0.1104	0.6259	0.6602
	100	0.1604	0.1583	0.0874	0.0845	0.1207	0.1183	0.6316	0.6389
	150	0.1575	0.1559	0.0871	0.0855	0.1207	0.1194	0.6347	0.6392
	200	0.1570	0.1563	0.0870	0.0859	0.1206	0.1191	0.6355	0.6387
0.8	20	0.0482	-0.1050	0.1812	0.1692	0.0674	-0.0913	0.7033	1.0271
	50	0.0520	0.0416	0.2281	0.2142	0.0534	0.0445	0.6666	0.6998
	100	0.0486	0.0473	0.2388	0.2341	0.0475	0.0462	0.6650	0.6724
	150	0.0470	0.0460	0.2300	0.2285	0.0460	0.0452	0.6770	0.6803
	200	0.0463	0.0459	0.2239	0.2230	0.0452	0.0449	0.6846	0.6862

Estimated fractions in R^2 variation partitioning, under the Poisson model applied to abundance data, at population and sample level according to varying sample sizes and levels of correlation between **X** and **W.**

Correlation level	Sample size	a_{pop}	a_{samp}	b_{pop}	b_{samp}	c_{pop}	c_{samp}	d_{pop}	d_{samp}
0.2	20	0.1876	0.1802	0.3643	0.3541	0.1893	0.1690	0.2588	0.2967
	50	0.1876	0.1812	0.3643	0.3671	0.1893	0.1761	0.2588	0.2755
	100	0.1876	0.1793	0.3643	0.3694	0.1893	0.1843	0.2588	0.2670
	150	0.1876	0.1834	0.3643	0.3666	0.1893	0.1856	0.2588	0.2644
	200	0.1876	0.1831	0.3643	0.3676	0.1893	0.1854	0.2588	0.2639
0.4	20	0.1096	0.1018	0.5678	0.5498	0.1106	0.1064	0.2120	0.2419
	50	0.1096	0.1029	0.5678	0.5645	0.1106	0.1065	0.2120	0.2261
	100	0.1096	0.1056	0.5678	0.5672	0.1106	0.1055	0.2120	0.2216
	150	0.1096	0.1058	0.5678	0.5679	0.1106	0.1086	0.2120	0.2176
	200	0.1096	0.1058	0.5678	0.5700	0.1106	0.1074	0.2120	0.2168
0.8	20	0.0283	0.0289	0.8625	0.8260	0.0274	0.0283	0.0818	0.1168
	50	0.0283	0.0265	0.8625	0.8483	0.0274	0.0272	0.0818	0.0980
	100	0.0283	0.0266	0.8625	0.8556	0.0274	0.0266	0.0818	0.0912
	150	0.0283	0.0274	0.8625	0.8565	0.0274	0.0272	0.0818	0.0889
	200	0.0283	0.0270	0.8625	0.8602	0.0274	0.0266	0.0818	0.0862

APPENDIX C

CHAPTER I

Estimated fractions in R^2 variation partitioning, under the logistic model applied to presence-absence data, Poisson model applied to presence-absence data and Poisson model applied to abundance data, at population and sample level according to varying number of inserted N(0,1) covariates. Samples were based on 100 observations. The degree of correlation between **X** and **W** is equal to 0.4.

Model	Number of inserted N(0,1) covariates	a_{pop}	a_{samp}	b_{pop}	b_{samp}	c_{pop}	c_{samp}	d_{pop}	d_{samp}
Logistic P-A	0	0.2239	0.2659	0.0372	0.0082	0.2179	0.2588	0.5211	0.4672
	5	0.2286	0.2792	0.0395	-0.0078	0.2174	0.2891	0.5144	0.4395
	10	0.2309	0.2903	0.0348	-0.0294	0.2115	0.3046	0.5228	0.4345
	15	0.2241	0.2866	0.0237	-0.0535	0.2025	0.3093	0.5496	0.4576
	20	0.1958	0.2598	0.0149	-0.0700	0.1910	0.3059	0.5982	0.5043
Poisson P-A	0	0.1407	0.1393	0.0772	0.0750	0.1355	0.1332	0.6465	0.6525
	5	0.1490	0.1458	0.0777	0.0732	0.1405	0.1359	0.6327	0.6452
	10	0.1587	0.1544	0.0694	0.0630	0.1452	0.1354	0.6266	0.6473
	15	0.1625	0.1535	0.0540	0.0461	0.1494	0.1317	0.6341	0.6688
	20	0.1500	0.1361	0.0381	0.0316	0.1531	0.1232	0.6588	0.7091
Poisson abundance	0	0.1407	0.1393	0.0772	0.0750	0.1355	0.1332	0.6465	0.6525
	5	0.1490	0.1458	0.0777	0.0732	0.1405	0.1359	0.6327	0.6452
	10	0.1587	0.1544	0.0694	0.0630	0.1452	0.1354	0.6266	0.6473
	15	0.1625	0.1535	0.0540	0.0461	0.1494	0.1317	0.6341	0.6688
	20	0.1500	0.1361	0.0381	0.0316	0.1531	0.1232	0.6588	0.7091

APPENDIX D

CHAPTER II

Calibration and validation metrics of all six SDM approaches for predicting lake trout. “A” corresponds to only abiotic variables, “B” corresponds to only biotic variables and “A+B” to both abiotic and biotic variables.

	GLM			GAM			MARS			RF			BRT			MAXENT		
	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B
MSE	3.891	7.436	4.607	0.072	0.131	0.067	0.081	0.133	0.077	0.072	0.136	0.067	0.353	0.825	0.320	0.079	0.162	0.074
Adj. Deviance	0.336	0.148	0.392	0.515	0.127	0.550	0.494	0.126	0.540	0.520	0.093	0.551	0.611	0.136	0.664	*	*	*
AUC	0.903	0.751	0.923	0.921	0.751	0.937	0.940	0.751	0.952	0.941	0.617	0.956	0.940	0.753	0.950	0.933	0.752	0.944
Sensitivity	89.049	67.435	85.014	89.770	67.435	88.040	90.490	67.435	91.066	93.084	31.412	91.354	89.914	70.317	90.778	90.778	71.758	91.354
Specificity	78.442	69.351	87.468	81.884	69.351	87.760	85.325	69.351	88.052	81.234	92.987	85.974	85.844	67.013	86.429	84.286	65.390	86.039
TSS	0.654	0.382	0.704	0.687	0.382	0.740	0.721	0.382	0.777	0.743	0.227	0.781	0.739	0.382	0.764	0.722	0.378	0.739

* No adj. deviance was computed for MAXENT.

Calibration and validation metrics of all six SDM approaches for predicting lake whitefish. "A" corresponds to only abiotic variables, "B" corresponds to only biotic variables and "A+B" to both abiotic and biotic variables.

	GLM			GAM			MARS			RF			BRT			MAXENT		
	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B
MSE	4.936	7.612	5.752	0.110	0.125	0.105	0.115	0.127	0.111	0.110	0.132	0.105	0.571	0.770	0.548	0.118	0.129	0.112
Adj. Deviance	0.320	0.285	0.360	0.343	0.250	0.364	0.319	0.250	0.348	0.353	0.216	0.374	0.471	0.257	0.471	*	*	*
AUC	0.871	0.819	0.878	0.873	0.820	0.884	0.875	0.820	0.890	0.889	0.758	0.899	0.886	0.822	0.894	0.855	0.822	0.864
Sensitivity	88.337	90.819	85.360	85.112	90.819	84.616	81.886	90.819	83.871	83.375	69.975	85.360	83.127	89.578	83.623	81.638	89.330	84.367
Specificity	69.340	59.097	72.507	72.777	59.097	75.169	76.213	59.097	77.830	77.156	78.976	78.504	77.089	60.445	79.515	74.259	60.647	73.652
TSS	0.579	0.469	0.594	0.586	0.483	0.606	0.593	0.483	0.618	0.618	0.481	0.651	0.612	0.480	0.627	0.560	0.486	0.588

* No adj. deviance was computed for MAXENT.

Calibration and validation metrics of all six SDM approaches for predicting walleye. "A" corresponds to only abiotic variables, "B" corresponds to only biotic variables and "A+B" to both abiotic and biotic variables.

	GLM			GAM			MARS			RF			BRT			MAXENT		
	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B
MSE	3.753	7.347	5.618	0.111	0.125	0.099	0.118	0.127	0.107	0.113	0.137	0.100	0.591	0.758	0.535	0.121	0.125	0.110
Adj. Deviance	0.326	0.388	0.449	0.394	0.344	0.467	0.386	0.343	0.454	0.429	0.309	0.495	0.488	0.353	0.543	*	*	*
AUC	0.870	0.878	0.901	0.888	0.878	0.905	0.886	0.878	0.909	0.890	0.862	0.914	0.878	0.881	0.905	0.873	0.880	0.908
Sensitivity	83.140	83.721	88.76	85.659	83.333	87.21	81.977	83.333	85.659	87.016	85.271	86.047	88.372	84.109	89.535	75.775	85.078	88.76
Specificity	74.052	81.633	79.009	76.531	82.289	81.451	79.738	82.289	83.892	75.875	79.446	83.382	70.408	81.122	78.28	81.633	80.248	80.977
TSS	0.590	0.641	0.657	0.619	0.644	0.669	0.614	0.644	0.682	0.615	0.639	0.674	0.596	0.639	0.666	0.592	0.634	0.678

* No adj. deviance was computed for MAXENT.

Calibration and validation metrics of all six SDM approaches for predicting northern pike. “A” corresponds to only abiotic variables, “B” corresponds to only biotic variables and “A+B” to both abiotic and biotic variables.

	GLM			GAM			MARS			RF			BRT			MAXENT		
	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B
MSE	1.635	1.894	1.646	0.118	0.162	0.108	0.128	0.165	0.118	0.114	0.174	0.105	0.574	0.976	0.537	0.156	0.182	0.150
Adj. Deviance	0.295	0.225	0.326	0.447	0.276	0.494	0.439	0.275	0.484	0.543	0.297	0.574	0.582	0.290	0.609	*	*	*
AUC	0.878	0.811	0.902	0.891	0.812	0.910	0.903	0.812	0.918	0.911	0.813	0.924	0.899	0.817	0.914	0.862	0.818	0.877
Sensitivity	76.623	64.581	79.693	78.631	63.754	82.527	80.638	63.754	85.360	81.464	64.109	82.999	87.249	64.109	82.881	76.151	64.109	80.874
Specificity	82.788	88.365	85.962	83.510	89.808	84.231	84.231	89.808	82.500	85.577	89.712	87.981	76.442	89.712	83.462	81.923	89.712	81.635
TSS	0.600	0.532	0.647	0.622	0.540	0.667	0.643	0.540	0.687	0.680	0.542	0.704	0.639	0.542	0.663	0.587	0.541	0.617

Calibration and validation metrics of all six SDM approaches for predicting white sucker. “A” corresponds to only abiotic variables, “B” corresponds to only biotic variables and “A+B” to both abiotic and biotic variables.

	GLM			GAM			MARS			RF			BRT			MAXENT		
	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B
MSE	0.601	0.553	0.609	0.162	0.185	0.160	0.164	0.188	0.162	0.162	0.189	0.158	0.886	1.083	0.865	0.225	0.241	0.224
Adj. Deviance	0.138	0.082	0.137	0.220	0.129	0.224	0.220	0.129	0.222	0.268	0.134	0.278	0.304	0.139	0.311	*	*	*
AUC	0.797	0.746	0.801	0.800	0.748	0.806	0.804	0.748	0.812	0.806	0.737	0.823	0.810	0.751	0.815	0.787	0.748	0.780
Sensitivity	77.230	60.016	71.518	77.191	62.363	75.470	77.152	62.363	79.421	75.430	59.624	80.203	80.438	67.449	84.664	77.621	62.285	79.421
Specificity	68.637	80.460	75.369	69.540	79.803	72.496	70.443	79.803	69.622	73.071	80.460	70.443	68.473	75.041	65.189	67.980	77.340	66.667
TSS	0.461	0.394	0.441	0.464	0.394	0.465	0.468	0.394	0.489	0.451	0.394	0.483	0.467	0.430	0.490	0.454	0.393	0.455

* No adj. deviance was computed for MAXENT.

Calibration and validation metrics of all six SDM approaches for predicting smallmouth bass. “A” corresponds to only abiotic variables, “B” corresponds to only biotic variables and “A+B” to both abiotic and biotic variables.

	GLM			GAM			MARS			RF			BRT			MAXENT		
	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B	A	B	A+B
MSE	7.711	5.606	7.732	0.090	0.140	0.086	0.103	0.140	0.101	0.088	0.139	0.085	0.467	0.869	0.447	0.101	0.199	0.095
Adj. Deviance	0.320	0.047	0.348	0.370	0.038	0.397	0.354	0.038	0.382	0.395	0.041	0.413	0.498	0.067	0.520	*	*	*
AUC	0.855	0.643	0.867	0.863	0.643	0.877	0.871	0.643	0.888	0.872	0.555	0.887	0.874	0.665	0.887	0.871	0.661	0.888
Sensitivity	80.480	45.045	78.679	83.033	45.045	78.229	85.586	45.045	77.778	83.334	14.414	83.634	83.784	78.979	84.084	82.883	77.477	83.183
Specificity	77.992	74.196	79.987	76.802	74.196	82.014	75.611	74.196	84.041	79.119	94.530	79.955	78.700	42.407	80.502	79.537	43.308	79.408
TSS	0.547	0.216	0.576	0.564	0.216	0.592	0.581	0.216	0.608	0.585	0.107	0.616	0.589	0.230	0.620	0.582	0.224	0.612

* No adj. deviance was computed for MAXENT.

REFERENCES

Abell, R., Allan, J.D., Lehner, B. 2007. Unlocking the potential of protected areas for freshwaters. Biol. Conserv. 134:48-63.

Ali, M.A., Ryder, R.A., and Anctil, M. 1977. Photoreceptors and visual pigments as related to behavioral responses and preferred habitats of perches (*Perca spp.*) and pikeperches (*Stizostedion spp.*). J. Fish. Res. Board. Can. 34:1475-1480.

Allan, J.D., and Flecker, A.S. 1993. Biodiversity conservation in running waters. BioScience 43:32-43.

Allan, J.D. 2004. Landscapes and riverscapes: The influence of land use on stream ecosystems. Annual Review of Ecology, Evolution, and Systematics. 35:257-284.

Allouche, O., Tsoar, A., Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa, and the true skill statistic (TSS). Journal of Applied Ecology 43: 1223-1232.

Alofs, K.M., and Jackson, D.A. 2015. The abiotic and biotic factors limiting establishment of predatory fishes at their expanding northern range boundaries in Ontario, Canada. Global Change Biology 21: 2227-2237.

Amarasekare, P. 2003. Competitive coexistence in spatially structured environments: a synthesis. Ecol. Lett. 6, online publication.

Anderson, J. M. 1993. Soil organisms as engineers: Microsite modulation of macroscale process, in Jones, C.G., Lawton, J.H. (eds.) Linking species and ecosystems. Chapman & Hall, New York, pp. 94-106.

Anderson, R.P., Peterson, A.T., Gómez-Laverde, M. 2002. Using niche-based GIS modelling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. Oikos 98:3-16.

Angermeier, P.L., and Winston, M.R. 1999. Characterizing fish community diversity across Virginia landscapes: prerequisite for conservation. Ecological Applications 9(1): 335-349.

Araújo, M.B., and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. J. Biogeogr. 33:1677-1688.

Araújo, M.B., and Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. Global Ecol. Biogeogr. 16: 743-753.

Araújo, M.B., Alagador, D., Cabeza, M., Nogués-Bravo, D., Thuiller, W. 2011. Climate change threatens European conservation áreas. Ecology Letters 14: 484-492.

Araújo, M.B., and Townsend Peterson, A. 2012. Uses and misuses of bioclimatic envelope modeling. Ecology 93(7):1527-1539.

Arlinghaus, R., Cooke, S.J., Coleman, F.C., Figueira, W.F., Ueland, J.S., and Crowder, L.B. 2005. Global impact of recreational fisheries. Science. 307:1561-1563.

Baiser, B., Gotelli, N.J., Buckley, H.L., Miller, T.E., and Ellison, A.M. 2012. Geographic variation in network structure of a nearctic aquatic food web. Global Ecology and Biogeography 21:579-591.

Barbour, C., and Brown, J. 1974. Fish species diversity in lakes. The American Naturalist, 108(962): 473-489.

Barry, S., and Elith, J. 2006. Error and uncertainty in habitat models. Journal of Applied Ecology 43:413-423.

Beale, C.M., Brewer, M.J., Lennon, J.J., Kriticos, D. 2014. A new statistical framework for the quantification of covariate associations with species distributions. Methods Ecol Evol 5: 421-432.

Beard, T.D., Jr., Hansen, M.J., and Carpenter, S.R. 2003. Development of a regional stock-recruitment model for understanding factors affecting walleye recruitment in northern Wisconsin lakes. Trans. Am. Fish. Soc. 132(2): 382-391.

Beever, E.A., Huso, M., Pyke, D.A. 2006. Multiscale responses of soil stability and invasive plants to removal of non-native grazers from an arid conservation reserve. Div. Distr. 12:258-268.

Begg, M.D., and Lagakos, S. 1990. On the consequences of model misspecification in logistic regression. Environmental Health Perspectives 87:69-75.

Beitinger, T.L., Bennett, W.A, and McCauley, R.W. 2000. Temperature tolerances of North American freshwater fishes exposed to dynamic changes in temperature. Env. Biol. Fish. 58:237-275.

Benndorf, J., Kneschke, H., Kossatz, K., Penz, E. 1984. Manipulation of the pelagic food web by stocking with predacious fishes. Int. Rev. Gesamten Hydrobiol. 69:407-428.

Bonar, S.A., Hubert, W.A., and Willis, D.W.(Editors). 2009. Standard methods for sampling North American freshwater. American Fisheries Society, Bethesda, Maryland.

Bond, N., Thomson, J., Reich, P., Stein, J. 2011. Using species distribution models to infer potential climate change-induced range shifts of freshwater fish in south-eastern Australia. Marine and Freshwater Research 62:1043-1061.

Borcard, D., Legendre, P., and Drapeau, P. 1992. Partialling out the spatial component of ecological variation. Ecology 73:1045-1055.

Boulangeat, I., Gravel, D., Thuiller, W. 2012. Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. Ecology Letters 15:584-593.

Box, E.O. 1995. Factors determining distributions of tree species and plant functional types. Vegetatio 121:101-116.

Bradford, M.J., and Irvine, J. 2000. Land-use, fishing, climate change and the decline of Thompson River, British Columbia, Coho Salmon. Can. J. Fish. Aquat. Sci. 57:13-16.

Breiman, L. et al. 1984. Classification and regression trees. – Chapman and Hall.

Brett, J.R. 1979. Environmental factors and growth. In: Hoar, W.S., Randall, D.J., Brett, J.R. (Eds.), Fish Physiology Volume 8: Bioenergetics and Growth. Academic Press, New York, pp. 599-675.

Brooker, R.W. 2006. Plant-plant interactions and environmental change. New Phytologist 171:271-284.

Buisson, L., Thuiller, W., Lek, S., Lim, P., Grenouillet, G. 2008. Climate change hastens the turnover of stream fish assemblages. Global Change Biology 14:2232-2248.

Callaghan, D.T., Blanchfield, P.J., and Cott, P.A. Lake trout (*Salvelinus namaycush*) spawning habitat in a northern lake: the role of wind and physical characteristics on habitat quality. Journal of Great Lakes Research 42: 299-307.

Cameron, A.C., and Windmeijer, F.A.G. 1996. R2 measures for count data regression models with applications to health-care utilization. J. Bus. Econom. Statist. 14:209-220.

Carlander, K.D. 1955. The standing crop of fish in lakes. J. Fish. Res. Board Can. 12: 543-570.

Carpenter, S.R., Kitchell, J.F., Hodgson, J.R. 1985. Cascading trophic interactions and lake productivity-fish predation and herbivory can regulate lake ecosystems. BioScience 35(1998):634-639.

Carpenter, S.R., Cole, J.J., Kitchell, J.F., and Pace, M.L. 1998. Impact of dissolved organic carbon, phosphorus, and grazing on phytoplankton biomass and production in experimental lakes. Limnol. Oceanogr. 43:73-80.

Casselman, J.M., and Lewis, C.A. 1996. Habitat requirements of northern pike (*Esox lucius*). Can. J. Fish. Aquat. Sci. 53(1): 161-174.

Chevalier, J.R. 1977. Changes in walleye (*Stizostedion vitreum vitreum*) population in Rainy Lake and factors in abundance, 1924-75. J. Fish. Res. Board Can. 34(10): 1696-1702.

Chezik, K.A., Lester, N.P., and Venturelli, P.A. 2014. Fish growth and degree-days I: selecting a base temperature for a within-population study. Can. J. Fish Aquat. Sci. 71:47-55.

Chow-Fraser, P. 1991. Use of the morphoedaphic index to predict nutrient status and algal biomass in some Canadian lakes. Can. J. Fish. Aquat. Sci. 48: 1909-1918.

Christie, G.C. 1986. Measures of optimal thermal habitat and their relationship to yields for four commercial fish species. M.Sc. thesis, University of Toronto, Toronto, Ont.

Christie, G.C., and Regier, H.A. 1988. Measures of optimal thermal habitat and their relationship to yields for 4 commercial fish species. Can. J. Fish. Aquat. Sci. 45:301-314.

Chu, C., Mandrak, N.E., Minns, C.K. 2005. Potential impacts of climate change on the distributions of several common and rare freshwater fishes in Canada. Diversity Distrib. 11:299-310.

Chu, C., Lester, N.P., Giacomini, H.C., Shuter, B.J., and Jackson, D.A. 2016. Catch-per-unit-effort and size spectra of lake fish assemblages reflect underlying patterns in ecological conditions and anthropogenic activities across regional and local scales. Can. J. Fish. Aquat. Sci. 73:535-546.

Clark, B.J., Dillon, P.J., and Molot, L.A. 2004. Lake trout (*Salvelinus namaycush*) habitat volumes and boundaries in Canadian Shield lakes. Boreal shield waters: Lake trout ecosystems in a changing environment. Edited by JM Gunn, RJ Steedman, and RA Ryder. Lewis Publishers, CRC Press, Boca Raton, Fla, 111-117.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20(1):37-46.

Cole, J.J., Carpenter, S.R., Kitchell, J., Pace, M.L., Solomon, C.T., and Weidel, B. 2011. Strong evidence for terrestrial support of zooplankton in small lakes based on stable isotopes of carbon, nitrogen, and hydrogen. Proc. Natl Acad. Sci. USA 108:1975-1980.

Coleman, F.C., Figueira, W.F., Ueland, J.S., and Crowder, L.B. 2004. The impact of United States recreational fisheries on marine fish populations. 305:1958-1960.

Comte, L., Buisson, L., Daufresne, M., and Grenouillet, G. 2013. Climate-induced changes in the distribution of freshwater fish: observed and predicted trends. Freshwat. Biol. 58:625-639.

Cornell, H.V. and Lawton, J.H. 1992. Species interactions, local and regional processes, and limits to the richness of ecological communities: a theoretical perspective. J. Anim. Ecol. 61: 1-12.

Cottenie, K. 2005. Integrating environmental and spatial processes in ecological community dynamics. Ecology Letters 8:1175-1182.

Coutant, C.C. 1987. Thermal preference: when does an asset become a liability. Environ. Biol. Fish. 18:161-172.

Daily, G.C. 1997. Nature's Services: Societal Dependence on Natural Ecosystems. Washington, DC: Island Press.

Darwall, W.R.T., Holland, R.A., Smith, K.G., Allen, D., Brooks, E.G.E., Katarya, V., Pollock, C.M., Shi, Y., Clausnitzer, V., Cumberlidge, N., Cuttelod, A., Dijkstra, K.-D.B., Diop, M.D., García, N., Seddon, M.B., Skelton, P.H., Snoeks, J., Tweddle, D., and Vié, J.-C. 2011. Implications of bias in conservation research and investment for freshwater species. Conservation Letters. 4:474-482.

Davey, C.M., Chamberlain, D.E., Newson, S.E., Noble, D.G., Johnson, A. 2012. Rise of the generalists: evidence for climate driven homogenization in avian communities. Global Ecology and Biogeography 21:568-578.

Davis, A.J., Jenkinson, L.S., Lawton, J.H., Shorrocks, B., and Wood, S. 1998. Making mistakes when predicting shifts in species range in response to global warming. Nature. 391:783-786.

De'ath, G. 2002. Multivariate regression trees: a new technique for modelling species-environment relationships. Ecology 83: 1105-1117.

Decoursey, D.G. 1992. Developing models with more detail: do more algorithms give more truth? Weed Technol. 6:709-715.

De Groot, R.S., Wilson, M.A., Boumans, R.M.J. 2002. A typology for the classification, description and valuation of ecosystem functions, goods and services. Ecol. Econ. 41:393-408.

De Groot, R., Brander, L., van der Ploeg, S., Costanza, R., Bernard, F., Braat, L., Christie, M., Crossman, N., Ghermandi, A., Hein, L., Hussain, S., Kumar, P., McVittie, A., Portela, R., Rodriguez, L.C., ten Brink, P., van Beukering, P. 2012. Global estimates of the value of ecosystems and their services in monetary units. Ecosystem Services 1:50-61.

Dodge, D.P., Goodchild, G.A., Tilt, J.C., and Waldriff, D.C. 1984. Manual of instructions: aquatic habitat inventory surveys. Ontario Ministry of Natural Resources, Fisheries Branch, Toronto.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M., Wilson, R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30:609-628.

Downing, J.A., Plante, C., and Lalonde, S. 1990. Fish production correlated with primary productivity, not the morphoedaphic index. Can. J. Fish. Aquat. Sci. 47: 1929-1936.

Downing, J.A., and Plante, C. 1993. Production of fish populations in lakes. Can. J. Fish. Aquat. Sci. 50:110-120.

Duinker, P.N., and Greig, L.A. 2006. The impotence of cumulative effects assessment in Canada: ailments, and ideas for redeployment. Environ. Manage. 37(2):153-161.

Eadie, J.McA., Keast, A. 1984. Resource heterogeneity and fish species diversity in lakes. Can. J. Zool. 62: 1689-1695.

Edwards, B.A., Southee, F.M., and McDermid, J.L. 2016. Using climate and a minimum set of local characteristics to predict the future distributions of freshwater fish in Ontario, Canada, at the lake-scale. Global Ecology and Conservation 8: 71-84.

Efron, B., and Tibshirani, R.J. An introduction to the bootstrap. Chapman & Hall/CRC, 1993.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S., Richardson, K., Schachetti Pereira, R., Schapire, R.E., Soberón, J., Williams, S.E., Wisz, M., and Zimmermann, N.E. 2006. Novel methods improve predictions of species' distributions from occurrence data. Ecography 29:129-151.

Elith, J., Leathwick, J.R., and Hastie, T. 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77(4):802-813.

Elith, J., and Leathwick, J.R. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40: 677-697.

Emmrich, M., Pédron, S., Brucet, S., Winfield, I.J., Jeppensen, E., Volta, P., Argillier, C., Lauridsen, T.L., Holmgren, K., Hesthagen, T., and Mehner, T. 2014.Geographical patterns in the body-size structure of European lake fish assemblages along abiotic and biotic gradients. J. Biogeogr. 41:2221-2233.

Ettema, C.H., and Wardle, D.A. 2002. Spatial soil ecology. Trends Ecol. Evol. 17:177-183.

Evans, D.O., and Oliver, C.H. 1995. Introduction of lake trout to inland lakes of Ontario, Canada: Factors contributing to successful colonization. Journal of Great Lakes Research. 21(Suppl. 1):30-53.

Evans, D.O. 2007. Effects of hypoxia on scope-for-activity and power capacity of lake trout (*Salvelinus namaycush*). Can. J. Fish. Aquat. Sci. 64:345-361.

Feria, A.T.P., and Peterson, A.T. 2002. Prediction of bird community composition based on point-occurrence data and inferential algorithms: a valuable tool in biodiversity assessments. Diversity and Distribution 8:49-56.

Ferrier, S., and Guisan, A. 2006. Spatial modelling of biodiversity at the community level. Journal of Applied Ecology 43:393-404.

Ficke, A.D., Myrick, C.A., Hansen, L.J. 2007. Potential impacts of global climate changeon freshwater fishes. Rev. Fish Biol. Fish. 17:581-613.

Fielding, A.H., and Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24(1):38-49.

Finstad, A.G., Helland, I.P., Ugedal, O., Hesthagen, T., and Hessen, D.O. 2014. Unimodal response of fish yield to dissolved organic carbon. Ecology Letters 17:36-43.

Fisheries and Oceans Canada. 2012. Survey of Recreational Fishing in Canada 2010. Government of Canada, Ottawa, Canada. 33 pp.

Filipe, A.F., Markovic, D., Pletterbauer, F., Tisseuil, C., De Wever, A., Schmutz, S., Bonada, N., Freyhof, J. 2013. Forecasting fish distribution along stream networks: brown trout (*Salmo trutta*) in Europe. Diversity Distrib. 19:1059-1071.

Friedman, J. 1991. Multivariate adaptive regression splines. – Ann. Stat. 19:1-141.

Friedman, J.H., and Meulman, J.J. 2003. Multiple additive regression trees with application in epidemiology. Statist. Med. 22:1365-1381.

Gelwick, F.P., Stock, M.S., Matthews, W.J. 1997. Effects of fish, water depth, and predation risk on patch dynamics in a north-temperate river ecosystem. Oikos 80:382-398.

Gilman, S.E., Urban, M.C., Tewksbury, J., Gilchrist, G.W., Holt, R.D. 2010. A framework for community interactions under climate change. Trends in Ecology and Evolution 25:325-331.

Gilpin, M.E., and Diamond, J.M. 1982. Factors contributing to non-randomness in species co-occurrences on islands. Oecologia 52: 75-84.

Graham, C.T., and Harrod, C. 2009. Implications of climate change for the fishes of the British Isles. J. Fish Biol. 74:1143-1205.

Goldman, R.L., Tallis, H., Kareiva, P., Daily, G.C. 2008. Field evidence that ecosystem service projects support biodiversity and diversify options. P. Nat. Acad. Sci. USA. 105(27):9445-9448.

Goodchilde, G.A., and Gale, G.E. 1982. Aquatic habitat inventory: the Ontario approach to lake surveys. Pages 304-311 in N.B. Armantrout, editor. Acquisition and utilization of aquatic habitat inventory information. American Fisheries Society, Western Division, Bethesda, Maryland, USA.

Grant, E.H.C., Lowe, W.H., Fagan, W.F. 2007. Living in the branches: population dynamics and ecological processes in dendritic networks. Ecol. Lett. 10:165-175.

Griffith, D.A., and Peres-Neto, P.R. 2006. Spatial modelling in ecology: the flexibility of eigenfunction spatial analyses in exploiting relative location information. Ecology 87:2603-2613.

Guisan, A., Theurillat, J.-P., Kienast, F. 1998. Predicting the potential distribution of plant species in an alpine environment. J. Veg. Sci. 9:65-74.

Guisan, A., and Harrell, F.E. 2000. Ordinal response regression models in ecology. Journal of Vegetation Science 11:617-626.

Guisan, A., and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135:147-186.

Guisan, A., and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8:993-1009.

Gunn, J.M., and Keller, W. 1990. Biological recovery of an acid lake after reductions in industrial emissions of sulphur. Nature 345:431-433.

Hanchet, S.M., Blackwell, R.G., and Dunn, A. 2005. Development and evaluation of catch per unit effort indices for southern blue whiting (*Micromesistius australis*) on the Campbell Island Rise, New Zealand. ICES Journal of Marine Science 62:1131-1138.

Hansen, M.J., Bozek, M.A., Newby, J.R., Newman, S.P., and Staggs, M.D. 1998. Factors affecting recruitment of walleyes in Escanaba Lake, Wisconsin, 1958-1996. N. Am. J. Fish. Manage. 18(4): 764-774.

Hansen, G.J.A., Carpenter, S.R., Gaeta, J.W., Hennessy, J.M., and Vander Zanden, M.J. 2015. Predicting walleye recruitment as a tool for prioritizing management actions. Can. J. Fish. Aquat. Sci. 72: 661-672.

Hansen, G.J.A., Midway, S.R., and Wagner, T. 2017. Walleye recruitment success is less resilient to warming water temperatures in lakes with abundant largemouth bass populations. Can. J. Fish. Aquat. Sci. 75(1): 106-115.

Hanson, J.M., and Leggett, W.C. 1982. Empirical prediction of fish biomass and yield. Can. J. Fish. Aquat. Sci. 39: 257-263.

Hanson, P., Johnson, T., Schindler, D., and Kitchell, J. 1997. Fish bioenergetics 3.0 [digital application]. University of Wisconsin Seagrant Institute, Madison, Wisc.

Harrell, F.E., Lee, K.L., Mark, D.B. 1996. Multivariate prognostic models: issues in developing models, evaluating assumptions and accuracy, and measuring and reducing errors. Statistics in Medicine 15:361-387.

Harvey, H.H. 1978. The fish communities of the Manitoulin Island lakes. Int. Ver. Theor. Angew. Limnol. 20:2031-2038.

Hasnain, S.S, Minns, C.K., Shuter, B.J. 2010. Key ecological temperature metrics for Canadian freshwater fishes. Applied Research and Development Branch, Ontario Ministry of Natural Resources, Canada. 54 pp.

Hastie, T.J., and Tibshirani, R.J. 1990. Generalized Additive Models. Chapman & Hall/CRC. Boca Raton.

Heikkinen, R., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W., Sykes, M.T. 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. Prog. Phys. Geog. 30:751-777.

Heikkinen, R.K., Luoto, M., Virkkala, R., Pearson, R.G., Körber, J.-H. 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. Global Ecol. Biogeogr. 16:754-763.

Heino, J. and Alahuhta, J. 2014. Elements of regional beetle faunas: faunal variation and compositional breakpoints along climate, land cover and geographical gradients. J. Anim. Ecol. 84(2):427-441.

Heinzl, H. and Mittlböck, M. 2003. Pseudo R-squared measures for Poisson regression models with over- or underdispersion. Computational Statistics & Data Analysis 44:253-271.

Henderson, H.F., and Welcomme, R.L. 1974. The relationship of yield to morpho-edaphic index and numbers of fishermen in African inland fisheries. Food and Agriculture Organization of the United Nations, CIFA (Committee for Inland Fisheries of Africa) Occasional Paper 1, Rome, Italy.

Hessen, D.O., Andersen, T., Larsen, S., Skjelkvale, B.L., and de Wit, H.A. 2009. Nitrogen deposition, catchment productivity, and climate as determinants of lake stoichiometry. Limnol. Oceanogr. 54:2520-2528.

Hobbs, R. 1997. Future landscapes and the future of landscape ecology. Landscape and Urban Plan. 37:1-9.

Hof, C., Brändle, M., Dehling, D.M., Munguía, M., Brandl, R., Araújo, M.B., Rahbek, C. 2012. Habitat stability affects dispersal and the ability to track climate change. Biol. Lett. 8(4):639-43.

Holmlund, C.M., and Hammer, M. 1999. Ecosystem services generated by fish populations. Ecological Economics 29:253-268.

Howard, C., Stephens, P.A., Pearce-Higgins, J.W., Gregory, R.D., Willis, S.G. 2014. Improving species distribution models: the value of data on abundance. Methods in Ecology and Evolution 5:506-513.

Huntley, B., Berry, P.M., Cramer, W., McDonald, A.P. 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. Journal of Biogeography 22:967-1001.

Hutchinson, G.E. 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology 22(2):415-427.

Jablonski, D. 2008. Biotic interactions and macroevolution: extensions and mismatches across scales and levels. Evolution 62 (4): 715-739.

Jackson, D.A., Somers, K.M., and Harvey, H.H. 1992. Null models and fish communities: evidence of non-random patterns. American Naturalist 139: 930-951.

Jackson, D.A., and Harvey, H.H. 1993. Fish and benthic invertebrates: community concordance and community-environment relationships. Can. J. Fish. Aquat. Sci. 50:2641-2651.

Jackson, D.A., and Harvey, H.H. 1997. Qualitative and quantitative sampling of lake fish communities. Canadian Journal of Fisheries and Aquatic Sciences 54: 2807-2813.

Jackson, R.B., Carpenter, S.R., Dahm, C.N., McKnight, D.M., Naiman, R.J., Postel, S.L., Running, S.W. 2001. Water in a changing world. Ecological Applications 11:1027-1045.

Jenkins, R.M. 1967. The influence of some environmental factors on standing crop and harvest of fishes in U.S. reservoirs, p. 298-321. In Proc. Reservoir Fisheries Symp., Southern Div. Am. Fish. Soc.

Jenkins, D.G., and Buikema, A.L. 1998. Do similar communities develop in similar sites? A test with zooplankton structure and function. Ecological Monographs 68:421-443.

Jeppesen, E., Mehner, T., Winfield, I.J., Kangur, K., Sarvala, J., Gerdeaux, D., Rask, M., Malmquist, H.J., Holmgren, K., Volta, P., Romo, S., Eckmann, R., Sandstrom, A., Blanco, S., Kangur, A., Stabo, H.R., Tarvainen, M., Ventela, A.M., Sondergaard, M., Lauridsen, T.L., Meerhoff, M. 2012. Impacts of climate warming on the long-term dynamics of key fish species in 24 European lakes. Hydrobiologia 694: 1-39.

Jobling, M. 1994. Fish Bioenergetics. Chapman and Hall, London, UK.

Johnson, M.G., Leach, J.H., Minns, C.K., and Olver, C.H. 1977. Limnological characteristics of Ontario lakes in relation to associations of walleye (*Stizostedion vitreum vitreum*), northern pike (*Esox Lucius*), lake trout (*Salvelinus namaycush*), and smallmouth bass (*Micropterus dolomieui*). J. Fish. Res. Board Can. 34:1592-1601.

Johnston, A., Ausden, M., Dodd, A.M., Bradbury, R.B., Chamberlain, D.E., Jiguet, F. et al. 2013. Observed and predicted effects of climate change on species abundance in protected areas. Nature Climate Change 3: 1055-1061.

Jones, J.R., and Hoyer, M.V. 1982. Sportfish harvest predicted by summer chlorophyll-a concentration in midwestern lakes and reservoirs. Trans. Am. Fish. Soc. 111: 176-179.

Kao, Y.-C., Madenjian, C.P., Bunnell, D.B., Lofgren, B.M., and Perroud, M. 2015. Potential effects of climate change on the growth of fishes from different thermal guilds in Lakes Michigan and Huron. Journal of Great Lakes Research 41: 423-435.

Karlsson, J., Bystrom, P., Ask, J., Ask, P., Persson, L., and Jansson, M. 2009. Light limitation of nutrient-poor lake ecosystems. Nature 460:506-509.

Karr, J.R. 1981. Assessment of biotic integrity using fish communities. Fisheries 6:21-27.

Kearney, M. and Porter, W. 2009. Mechanistic nice modelling: combining physiological and spatial data to predict species'ranges. Ecology Letters 12:1-17.

King, J.R., Shuter, B.J., and Zimmerman, A.P. 1999. Empirical links between thermal habitat, fish growth, and climate change. Trans. Am. Fish. Soc. 128:656-665.

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., McInerny, G.J., Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.-C., Zimmermann, N.E. & O'Hara, R.B. 2012. Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial scales. Journal of Biogeography 39: 2163-2178.

Kling, G.W., Hayhoe, K., Johnson, L.B., Magnuson, J.J., Polasky, S., Robinson, S.K., Shuter, B.J., Wander, M.M., Wuebbles, D.J., and Zak, D.R. 2003. Confronting climate change in the Great Lakes. Union of Concerned Scientists, Cambridge, Massachusetts, and Ecological Society of America, Washington, D.C.

Leathwick, J.R., and Austin, M.P. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. Ecology 82:2560-2573.

Leathwick, J.R. 2002. Intra-generic competition among Nothofagus in New Zealand's primary indigenous forests. Biodiv. Conserv. 11:2177-2187.

Le Cessie, S. and Van Houwelingen, J.C. 1992. Ridge estimators in logistic regression. Appl. Statist. 41(1):191-201.

Lee, W.-S., Mangel, M., and Peres-Neto, P. 2016. Environmental integration: patterns of correlation between environmental factors, early life decisions, and their long-term consequences. Evolutionary Ecology Research 17: 1-19.

Legendre, P., and Fortin, M.-J. 1989. Spatial pattern and ecological analysis. Vegetatio 80:107-138.

Legendre, P., and Legendre, L. Numerical Ecology. Amsterdan: Elsevier Science; 1998.

Lennon, J.J. 2000. Red-shifts and red herrings in geographical ecology. Ecography 23: 101-113.

Legendre, P., Borcard, D., and Peres-Neto, P.R. 2005. Analyzing beta diversity: partitioning the spatial variation of community composition data. Ecological Monographs. 75: 435-450.

Lester, N.P., Marshall, T.R., Armstrong, K., Dunlop, W.I., Ritchie, B. 2003. A broad-scale approach to management of Ontario's recreational fisheries. N. Am. J. Fish. Manage. 23:1312-1328.

Lester, N.P., Dextrase, A.J., Kushneriuk, R.S., Rawson, M.R., Ryan, P.A. 2004. Light and temperature: key factors affecting Walleye abundance and production. Transactions of the American Fisheries Society 133:588-605.

Lévêque, C., Oberdorff, T., Paugy, D., Stiassny, M.L.J., and Tedesco, P.A. 2007. Global diversity of fish (Pisces) in freshwater. Hydrobiologia. 595:545-567.

Levin, S.A. 1992. The problem of pattern and scale in ecology. Ecology 73:1943-1967.

Levins, R. 1966. The strategy of model building in population biology. Am. Sci. 54:421-431.

Lienesch, P.W., McDonald, M.E., Hershey, A.E., O'Brien, W.J., and Bettez, N.D. 2005. Effects of a whole-lake, experimental fertilization on lake trout in an oligotrophic arctic lake. Hydrobiologia, 548:51-66.

Lin, D.Y., Wei, L.J., Ying, Z.L. 2002. Model-checking techniques based on cumulative residuals. Biometrics 58:1-12.

Mackenzie-Grieve, J.L., and Post, J.R. 2006. Projected impacts of climate warming on production of lake trout (*Salvelinus namaycush*) in southern Yukon lakes. Can. J. Fish. Aquat. Sci. 63: 788-797.

Mac Nally, R., and Fleishman, E. 2004. A successful predictive model of species richness based on indicator species. Conservation Biology 18:646-654.

Magnuson, J.J., Crowder, L.B., and Medvick, P.A. 1979. Temperature as an ecological resource. Am. Zool. 19:331-343.

Magnuson, J.J., Meisner, J.D., and Hill, D.K. 1990. Potential changes in the thermal habitat of the Great Lakes fish after global climate warming. Transactions of the American Fisheries Society 119(2): 254-264.

Makarenkov, V., and Legendre, P. 2002. Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. Ecology 83(4): 1146-1161.

Malakoff, D. 1997. Extinction on the high seas. Science 277:486-488.

Maloney, K.O., Mitchell, R.M., and Feminella, J.W. 2006. Influence of catchment disturbance on *Pteronotropis euryzonus* (Broadstripe shiner) and *Semotilus thoreauianus* (Dixie chub). Southeastern Naturalist. 5(3):393-412.

Malmqvist, B., and Rundle, S. 2002. Threats to running water ecosystems of the world. Environmental Conservation 29:134-153.

Mandrak, N.E., and Crossman, E.J. 1992. A checklist of Ontario freshwater fishes: annotated with distribution maps. Royal Ontario Museum, Toronto, Ontario, Canada.

Matuszek, J.E. 1978. Empirical predictions of fish yields of large North American lakes. Trans. Am. Fish. Soc. 107:385-394.

Mazumder, A., McQueen, D.J., Taylor, W.D., Lean, D.R.S. 1988. Effects of fertilization and planktivorous fish (yellow perch) predation on size-distribution of particulate phosphorus and assimilated phosphate: large enclosure experiments. Limnology and Oceanography 33:421-430.

McCullagh, P., and Nelder, J.A. 1989. Generalized Linear Models: Second Edition. Chapman Hall: New York.

McKenzie, E., Irwin, F., Ranganathan, J., Hanson, C.E., Kousky, C., Bennett, K., Ruffo, S., Conte, M., Salzman, J., Paavola, J. 2011. Incorporating ecosystem services in decisions. Pages 339-355. In: Kareiva, P., Tallis, H., Ricketts, T.H., Daily, G.C., Polasky, S. (Eds.), Natural Capital: Theory and Practice of Mapping Ecosystem Services. Oxford University Press, Oxford, UK.

McMahon, T.E., and Bennett, D.H. 1996. Walleye and northern pike. Fisheries 21(8): 6-13.

McQueen, D.J., Johannes, M.R.S., Lafontaine, N.R., Young, A.S., Longbotham, E., Lean, D.R.S. 1990. Effects of planktivore abundance on chlorophyll-a and Secchi depth. Hydrobiologia 200/201:337-341.

Meier, E.S., Edwards, T.C. Jr, Kienast, F., Dobbertin, M., Zimmermann, N.E. 2011. Co-occurrence patterns of trees along macro-climatic gradients and their potential influence on the present and future distribution of Fagus sylvatica L. Journal of Biogeography 38: 371-382.

Mertz, G., and Myers, R.A. 1998. A simplified formulation for fish production. Can. J. Fish. Aquat. Sci. 55: 478-484.

Minns, C.K. 1995. Allometry of home range size in lake and river fishes. Canadian Journal of Fisheries and Aquatic Sciences 52:1499-1508.

Minns, C.K., and Moore, J.E. 1992. Predicting the impact of climate change on the spatial pattern of freshwater fish yield capability in eastern Canadian lakes. Climate Change 22: 327-346.

Mittlböck, M., and Heinzl, H. 2001. A note on R2 measures for Poisson and logistic regression models when both models are applicable. Journal of Clinical Epidemiology 54:99-103.

Mittlböck, M., and Schemper, M. 1996. Explained variation for logistic regression. Statistics in Medicine 15:1987-1997.

Mittlböck, M., and Waldhör, T. 2000. Adjustments for R2¬ –measures for Poisson regression models. Comput. Statist. Data Anal. 34:461-472.

Mod, H.K., le Roux, P.C., and Luoto, M. 2014. Outcomes of biotic interactions are dependent on multiple environmental variables. Journal of Vegetation Science, 25: 1024-1032.

Moyle, J.B. 1956. Relationship between the chemistry of Minnesota surface waters and wildlife management. J. Wildl. Manage. 20: 303-320.

Murtaugh, P.A. 1998. Methods of variable selection in regression modeling. Commun. Statist.-Simula. 27(3):711-734.

Musick, J.A. 1999. Criteria to define extinction risk in marine fishes. Fisheries 24:6-12.

Naiman, R.J., Magnuson, J.J., McKnight, D.M., Stanford, J.A. 1995. The Freshwater Imperative: A Research Agenda. Island Press, Washington D.C., U.S.A.

Naiman, R.J., and Turner, M.G. 2000. A future perspective on North America's freshwater ecosystems. Ecological Applications 10:958-970.

Nate, N.A., Bozek, M.A., Hansen, M.J., and Hewett, S.W. 2000. Variation in walleye abundance with lake size and recruitment source. N. Am. J. Fish. Manage. 20(1): 119-126.

Nikolsky, G.V. 1963. The ecology of fishes. Academic Press, London. 352 p.

Oberdorff, T., Pont, D., Hugueny, B., Chessel, D. 2001. A probabilistic model characterizing fish assemblages of French rivers: a framework for environmental assessment. Freshwater Biology 46:399-415.

O'Connor, D.V., Rottiers, D.V., Berlin, W.H. 1981. Food consumption, growth rate, conversion efficiency, and proximate composition of yearling lake trout. Administrative Report No. 81-5, U.S. Fish and Wildlife Service. 21 pp.

Oglesby, R.T. 1977. Relationship of fish yield to lake phytoplankton standing crop, production and morphoedaphic factors. J. Fish. Res. Board Can. 34:2271-2279.

O'Grady, J.J., Reed, D.H., Brook, B.W., Frankham, R. 2004. What are the best correlates of predicted extinction risk? Biological Conservation 118: 513-520.

Olden, J.D., and Jackson, D.A. 2001. Fish-habitat relationships in lakes: gaining predictive and explanatory insight by using artificial neural networks. Transactions of the American Fisheries Society 130:878-897.

Olden, J.D., and Jackson, D.A. 2002. A comparison of statistical approaches for modelling fish species distributions. Freshwater Biology 47: 1976-1995.

Olden, J.D., Lawler, J.J., and Poff, N.L. 2008. Machine learning methods without tears: a primer for ecologists. The Quarterly Review of Biology 83:171-193.

Owen, J.G. 1989. Patterns of herpetofaunal species richness: relation to temperature, precipitation and variance in elevation. J. Biogeogr. 16:141-150.

Pankhurst, N.W., and Munday, P.L. 2011. Effects of climate change on fish reproduction and early life history stages. Marine and Freshwater Research 62:1015-1026.

Pearce, J., and Ferrier, S. 2000a. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. Ecol. Model. 128:127-147.

Pearce, J., and Ferrier, S. 2000b. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Model. 133:225-245.

Pearce, J., and Ferrier, S. (2001). The practical value of modelling relative abundance of species for regional conservation planning: a case study. Biological Conservation 98:33-43.

Pearson, R.G., and Dawson, T.P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Glob. Ecol. Biogeogr. 12: 361-371.

Pellissier, L., Brathen, K.A., Pottier, J., Randin, C.F., Vittoz, P., Dubuis, A. et al. 2010. Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. Ecography 33: 1004-1014.

Peres-Neto, P.R., Legendre, P., Dray, S., and Borcard, D. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology 87(10):2614-2625.

Perrings, C.A., Mäler, K.-G., Folke, C., Holling, C.S., Jansson, B.-O. 1995. Biodiversity conservation and economic development: the policy problem. In: Perrings, C.A., Mäler, K.-G., Folke, C., Holling, C.S., Jansson, B.-O. (Eds.). Biodiversity Conservation. Kluwer, Dordrecht, pp. 3-21.

Peterson, A.T., and Vieglais, D.A. 2001. Predicting species invasions using ecological niche modelling: new approaches from bioinformatics attack a pressing problem. Bioscience 51:363-371.

Peterson, A.T. 2003. Predicting the geography of species'invasions via ecological niche modelling. Quarterly Review of Biology 78:419-433.

Peterson, A.T., Martinez-Meyer, E., Gonzalez-Salazar, C. 2004. Reconstructing the Pleistocene geography of the Aphelocoma jays (Corvidae). Biodiv. Res. 10:237-246.

Peterson, A.T., Lash, R.R., Carroll, D.S., Johnson, K.M. 2006. Geographic potential for outbreaks of Marburg hemorrhagic fever. American Journal of Tropical Medicine and Hygiene 75:9-15.

Peterson, A.T., Papes, M., Carroll, D.S., Leirs, H., Johnson, K.M. 2007. Mammal taxa constituting potential coevolved reservoirs of filoviruses. Journal of Mammalogy 88:1544-1554.

Phillips, S.J., Anderson, R.P. and Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190(3-4): 231-259.

Plante, C., and Downing, J.A. 1989. Production of freshwater invertebrate populations in lakes. Can. J. Fish. Aquat. Sci. 46:1489-1498.

Poesch, M.S., Chavarie, L., Chu, C., Pandit, S.N., and Tonn, W. 2016. Climate change impacts on freshwater fishes: a Canadian perspective. Fisheries 41(7): 385-391.

Portner, H.O., and Farrell, A.P. 2008. Physiology and climate change. Science 322:690-692.

Portner, H.O., and Peck, M.A. 2010. Climate change effects on fishes and fisheries: towards a cause-and-effect understanding. J. Fish Biol. 77:1745-1779.

Post, D.M., Pace, M.L., and Hairston Jr, N.G. 2000. Ecosystem size determines food-chain length in lakes. Nature 405:1047-1049.

Postel, S., and Richter, B. 2003. Rivers for Life: Managing Water for People and Nature. Island Press, Washington D.C., U.S.A.

Prairie, Y.T., Duarte, C.M., and Kalff, J. 1989. Unifying nutrient-chlorophyll relationships in lakes. Can. J. Fish. Aquat. Sci. 46: 1176-1182.

Prasad, A.M., Iverson, L.R. and Liaw, A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9:181-199.

Pratchett, M.S., Bay, L.K., Gehrke, P.C., Koehn, J.D., Osborne, K., Pressey, R.L., Sweatman, H.P.A., Wachenfeld, D. 2011. Contribution of climate change to degradation and loss of critical fish habitats in Australian marine and freshwater environments. Marine and Freshwater Research 62:1062-1081.

Prentice, I.C. 1986. Some concepts and objectives of forest dynamics research. In: Fanta, J. (Ed.), Forest Dynamics Research in Western and Central Europe. PUDOC, Wageningen, pp. 32-41.

Quiros, R. 1990. Predictors of relative fish biomass in lakes and reservoirs of Argentina. Can. J. Fish. Aquat. Sci. 47: 928-939.

Quiros, R. 1991. Empirical relationships between nutrients, phyto- and zooplankton and relative fish biomass in lakes and reservoirs of Argentina. Verh. Internat. Verein. Limnol. 24:1198-1206.

Quist, M.C., Guy, C.S., Schultz, R.D., and Stephen, J.L. 2003. Latitudinal comparisons of walleye growth in North America and factors influencing growth of walleyes in Kansas reservoirs. N. Am. J. Fish. Manage. 23(3): 677-692.

Quist, M.C., Guy, C.S., Bernot, R.J., and Stephen, J.L. 2004. Factors related to growth and survival of larval walleyes: implications for recruitment in a southern great plains reservoir. Fish. Res. 67(2): 215-225.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Raat, A.J.P. 1988. Synopsis of biological data on the northern pike, *Esox lucius linnaeus*, 1758. FAO Fish. Synop., (30)Rev.2:178 p.

Rahel, F.J. 2002. Homogenization of freshwater faunas. Annual Review of Ecology and Systematics 33:291-315.

Randall, R.G., Kelso, J.R.M., and Minns, C.K. 1995. Fish production in freshwaters: Are rivers more productive than lakes? Can. J. Fish. Aquat. Sci. 52: 631-643.

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M., Guisan, A. 2006. Are niche-based species distribution models transferable in space? J. Biogeogr. 33:1689-1703.

Rao, C.R. 1964. The use and interpretation of principal component analysis in applied research. Sankhyaá, Series A. 26:329-358.

Rawson, D.S. 1952. Mean depth and the fish production of large lakes. Ecology 33:513-521.

Rawson, D.S. 1955. Morphometry as a dominant factor in the productivity of large lakes. Int. Ver. Theor. Angew. Limnol. Verh. 12: 164-175.

Rawson, D.S. 1960. A limnological comparison of twelve large lakes in northern Saskatchewan. Limnol. Oceanogr. 5: 195-211.

Raxworthy, C.J., Martinez-Meyer, E., Horning, N., Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A., Peterson, A.T. 2003. Predicting distributions of known and unknown reptile species in Madagascar. Nature 426:837-841.

Revenga, C., Campbell, I., Abell, R., de Villiers, P., Bryer, M. 2005. Prospects for monitoring freshwater ecosystems towards the 2010 targets. Philosophical Transactions of the Royal Society B. 360:397-413.

Ricciardi, A., and Rasmussen, J.B. 1999. Extinction rates of North American freshwater fauna. Conservation Biology 13:1220-1222.

Ricker, W.E. 1975. Computation and interpretation of biological statistics of fish populations. Fish. Res. Bd. Canada, Bull. No. 191.

Ritz, K. 2005. Underview: Origins and consequences of belowground biodiversity, in Bardgett, R.D., Usher, M.B., Hopkins, D.W. (eds.) Biological diversity and function in soils. Cambridge University Press, Cambridge, pp. 381-401.

Robinson, L.D., and Jewell, N.P. 1991. Some surprising results about covariate adjustment in logistic regression models. International Statistical Review 58(2): 227-240.

Romanuk, T.N., Hayward, A., and Hutchings, J.A. 2010. Trophic level scales positively with body size in fishes. Global Ecology and Biogeography 20:231-240.

Rosenzweig, M.L. 1995. Species diversity in space and time. Cambridge: Cambridge University Press. 436pp.

Rounsefell, G.A. 1946. Fish production in lakes as a guide for estimating production in proposed resevoirs. Copeia 1:29-40.

Rushton, S.P., Ormerod, S.J., and Kerby, G. 2004. New paradigms for modelling species distributions? Journal of Applied Ecology 41:193-200.

Ryan, P.M., and Harvey, H.H. 1977. Growth of the rock bass, *Ambloplites rupestris*, in relation to the morphoedaphic index as an indicator of an environmental stress. Journal of the Fisheries Research Board of Canada 34: 2079-2088.s

Ryder, R.A. 1965. A method for estimating the potential fish production of north-temperate lakes. Trans. Am. Fish. Soc. 94:214-218.

Ryder, R.A. 1977. Effects of ambient light variations on behavior of yearling, subadult, and adult walleyes (*Stizostedion vitreum vitreum*). J. Fish. Res. Board. Can. 34:1481-1491.

Ryder, R.A., Kerr, S.R. 1989. Environmental priorities: Placing habitat in hierarchic perspective. In: Levings, C.D., Holtby, L.B., Henderson, M.A. (eds.), Proceedings of the National Workshop on Effects of Habitat Alteration on Salmonid Stocks. Can. Spec. Publ. Fish. Aquat. Sci. 105:2-12.

Rykiel, E.J., Jr. 1996. Testing ecological models: the meaning of validation. Ecological Modeling 90:229-244.

Sandstrom, S., Rawson, M., and Lester, N. 2011. Manual of instructions for broad-scale fish community monitoring using North American (NA1) and Ontario small mesh (ON2) gillnets. Ontario Ministry of Natural Resources. Peterborough, Ontario. Version 2011.1.

Sawatzky, C.D., Michalak, D., Reist, J.D., Carmichael, T.J., Mandrak, N.E. and Heuring, L.G. 2007. Distributions of freshwater and anadromous fishes from the mainland Northwest Territories, Canada. Can. Manuscr. Rep. Fish. Aquat Sci. 2793: xiv + 239 p.

Schaefer, M.B. 1968. Methods of estimating effects of fishing on fish populations. Trans. Am. Fish. Soc. 97:231-241.

Schindler, D.E. 1992. Nutrient regeneration by Sockeye salmon (*Oncorhynchus nerka*) fry and subsequent effects on zooplankton and phytoplankton. Can. J. Fish. Aquat. Sci. 49:2498-2506.

Schlesinger, D.A., and Regier, H.A. 1982. Climatic and morphoedaphic indices of fish yields from natural lakes. Trans. Am. Fish. Soc. 111:141-150.

Schlesinger, D.A., and McCombie, A.M. 1983. An evaluation of climate, morphoedaphic and effort data as predictors of yields from Ontario sport fisheries. Ont. Fish. Tech. Rep. Ser. 10: 14 p.

Schoener, T.W. 1989. Food webs from the small to the large. Ecology 70:1559-1589.

Schröder, B. 2008. Challenges of species distribution modelling belowground. J. Plant Nutr. Soil Sci. 171:325-337.

Schuster, A. 1994. Regional distribution of breeding birds elaborated by a geographic information system – Possibilities and limitations. In: Hagenmeijer, E. J. M., Vertrael, T. J. (Eds.), Bird Numbers 1992 – Distribution, monitoring and ecological aspects, Statistics Netherlands, Voorburg/Heerlen and SOVON, Beek-Ubbergen, pp. 493-501.

Scott, W.B., and Crossman, E.J. 1998. Freshwater fishes of Canada. Galt House Publications, Oakville, Ontario.

Scotts, D., and Drielsma, M. 2003. Developing landscape frameworks for regional conservation planning: an approach integrating fauna spatial distributions and ecological principles. Pacific Conservation Biology 8:235-254.

Serns, S.L. 1982. Influence of various factors on density and growth of age-0 walleyes in Escanaba Lake, Wisconsin, 1958-1980. Trans. Am. Fish. Soc. 111(3): 299-306.

Sharma, S., and Jackson, D.A. 2007. Fish assemblages and environmental conditions in lower reaches of northeastern Lake Erie tributaries. J. Great. Lakes Res. 33(1):15-27.

Sharma, S., Jackson, D.A., Minns, C.K., and Shuter, B.J. 2007. Will northern fish populations be in hot water because of climate change? Global Change Biology 13: 2052-2064.

Sharma, S., and Jackson, D.A. 2008. Predicting smallmouth bass (*Micropterus dolomieu*) occurrence across North America under climate change: a comparison of statistical approaches. Can. J. Fish. Aquat. Sci. 65:471-481.

Sharma, S., Walker, S.C., and Jackson, D.A. 2008. Empirical modelling of lake water-temperature relationships: a comparison of approaches. Freshwater Biology 53: 897-911.

Sharma, S., Jackson, D.A., and Minns, C.K. 2009. Quantifying the potential effects of climate change and the invasion of smallmouth bass on native lake trout populations across Canadian lakes. Ecography 32: 517-525.

Sharma, S., Legendre, P., Boisclair, D., and Gauthier, S. 2012. Effects of spatial scale and choice of statistical model (linear versus tree-based) on determining species-habitat relationships. Can. J. Fish. Aquat. Sci. 69:2095-2111.

Sharpe, P.J.A. 1990. Forest modeling approaches: compromises between generality and precision. In: Dixon, R.K., Meldahl, R.S., Ruark, G.A., Warren, W.G. (Eds.), Process Modeling of Forest Growth Responses to Environmental Stress. Timber Press, Portland, OR, pp. 180-190.

Sheridan, J.A., and Bickford, D. 2011. Shrinking body size as an ecological response to climate change. Nat. Clim. Chang. 1:401-406.

Shuter, B.J., Maclean, J.A., Fry, F.E.J., and Regier, H.A. 1980. Stochastic simulation of temperature effects on first-year survival of smallmouth bass. Transactions of the American Fisheries Society 109(1): 1-34.

Shuter, B.J., Schlesinger, D.A., and Zimmerman, A.P. 1983. Empirical predictors of annual surface water temperature cycles in North American lakes. Can. J. Fish. Aquat. Sci. 40:1838-1845.

Soberon, J., and Nakamura, M. 2009. Niches and distributional areas: Concepts, methods, and assumptions. PNAS 106 (2): 19644-19650.

Soininen, J. 2016. Spatial structure in ecological communities – a quantitative analysis. Oikos 125: 160-166.

Somers, K.M., and Harvey, H.H. 1984. Alteration of fish communities in lakes stressed by acid deposition and heavy metals near Wawa, Ontario. Can. J. Fish. Aquat. Sci. 41:20-29.

Stauffer, B.A., Patrick, C.J., Robinson, K.L., and Peter, H. 2014. Temporal scales of drivers of community dynamics: from microbes to macrofaunal across the salinity gradient. In: PF Kemp (Ed.) Eco-DAS X Symposium Proceedings, Waco, TX: Association for the Sciences of Limnology and Oceanography, 14-24.

Stefan, H.G., Fang, X., and Eaton, J.G. 2001. Simulated fish habitat changes in North American lakes in response to projected climate change. Transactions of the American Fisheries Society. 130:459-477.

Strubbe, D., Matthysen, E., Graham, C.H. 2010. Assessing the potential impact of invasive ring-necked parakeets Psittacula krameri on native nuthatches Sitta europeae in Belgium. Journal of Applied Ecology 47:549-557.

Svenning, J.-C., Gravel, D., Holt, R.D. et al. 2014. The influence of interspecific interactions on species range expansion rates. Ecography. EV:1-EV.

Swezey, S.L., and Heizer, R.F. 1977. Ritual management of salmonid fish resources in California. Journal of California Anthropology 4:7-29.

Tanentzap, A.J., Szkokan-Emilson, E.J., Kielstra, B.W., Arts, M.T., Yan, N.D., and Gunn, J.M. 2014. Forests fuel fish growth in freshwater deltas. Nat. Commun. 5:4077.

Thompson, R.M., and Townsend, C.R. 2005 Energy availability, spatial heterogeneity and ecosystem size predict food-web structure in streams. Oikos 108:137-148.

Thuiller, W., Lavorel, S. Araújo, M., Sykes, M.T., Prentice, I.C. 2005. Climate change threats to plant diversity in Europe. Proc. Natl. Acad. Sci. USA. 102:8245-8250.

Tobler, W.R. 1970. A computer movie simulating urban growth in the Detroit region. Econ. Geogr. 46:234-240.

Vadeboncoeur, Y., Peterson, G., Vander Zanden, M.J., and Kalff, J. 2008. Benthic algal production across lake size gradients: interactions among morphometry, nutrients, and light. Ecology 89(9): 2542-2552.

Vander Zanden, M.J., Shuter, B.J., Lester, N., and Rasmussen, J.B. 1999. Patterns of food chain length in lakes: a stable isotope study. American Naturalist 154:406-416.

Vander Zanden, M.J., Olden, J.D., Thorne, J.H., and Mandrak, N.E. 2004. Predicting occurrences and impacts of smallmouth bass introductions in north temperate lakes. Ecological Applications 14(1): 132-148.

Vander Zanden, M.J., Vadeboncoeur, Y., and Chandra, S. 2011. Fish reliance on littoral-benthic resources and the distribution of primary production in lakes. Ecosystems 14: 894-903.

Van Houwelingen, J.C., and Le Cessie, S. 1990. Predictive value of statistical models. Statistics in Medicine 9:1303-1325.

Van Zuiden, T.M., Chen, M.M., Stefanoff, S., Lopez, L., and Sharma, S. 2016. Projected impacts of climate change on three freshwater fishes and potential novel competitive interactions. Diversity and Distributions 22: 603-614.

Vaughan, I.P., and Ormerod, S.J. 2005. The continuing challenges of testing species distribution models. Journal of Applied Ecology 42:720-730.

Villemant, C., Barbet-Massin, M., Perrard, A., Muller, F., Gargominy, O., Jiguet, F., Rome, Q. 2011. Predicting the invasion risk by the alien bee-hawking Yellowlegged hornet *Vespa velutina nigrithorax* across Europe and other continents with niche models. Biological Conservation. 144:2142-2150.

Waldhör, T., Haidinger, G., Schober, E. 1998. Comparison of R2 measures for Poisson regression by simulation. J. Epidemiol. Biostatist. 3:209-215.

Walker, W.W. 1979. Use of hypolimnetic oxygen depletion rate as trophic state index for lakes. Water Resources Research 15: 1463-1470.

Walker, L.R. Landau, F.H., Velazquez, E., Shiels, A.B., Sparrow, A.D. 2010. Early successional woody plants facilitate and ferns inhibit forest development on Puerto Rican landslides. J. Ecol. 98:625-635.

Wang, L., and Jackson, D.A. 2014. Shaping up model transferability and generality of species distribution modelling for predicting invasions: implications from a study on *Bythotrephes longimanus*. Biol. Invasions. 16:2079-2103.

Weisberg, S. (1980). Applied Linear Regression. Wiley, New York.

Wenger, S.J., and Olden, J.D. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods in Ecology and Evolution 3:260-267.

Wetzel, R.G. 1983. Limnology. 2d ed. Saunders College Publishing, Philadelphia.

Wherry, R. J. 1931. A New Formula for Predicting the Shrinkage of the Coefficient of Multiple Correlation. Ann. Math. Statist. 2(4): 440--457.

Whittaker, R.H. 1975. Communities and Ecosystems. MacMillan, New York.

Wiens, J.A. 1984. On understanding a non-equilibrium world: myth and reality in community patterns and processes. Pages 439-457 *in* D.R. Strong, D. Simberloff, L.G. Abele, and A.B. Thistle (editors). Ecological communities: conceptual issues and the evidence. Princeton University Press, Princeton, New Jersey.

Wiens, J.A. 1989. Spatial scaling in ecology. Funct. Ecol. 3:385-397.

Wiens, J.A. 1992. Ecology 2000: an essay on future directions in ecology. Bulletin of the Ecological Society of America 73:165-170.

Wiens, J.J. 2011. The niche, biogeography and species interactions. Philosophical Transactions of the Royal Society of London Series B 366: 2336-2350.

Williams, P.H., and Araújo, M.B. 2000. Using probability of persistence to identify important areas for biodiversity conservation. Proceedings of the Royal Society of London B. 267:1959-1966.

Williamson, C.E., Stemberger, R.S., Morris, D.P., Frost, T.M., and Paulsen, S.G. 1996. Ultraviolet radiation in North American lakes: attenuation estimates from DOC measurements and implications for plankton communities. Limnol. Oceanogr. 41:1024-1034.

Wismer, D.A., and Christie, A.E. 1987. Temperature relationships of Great Lakes fishes. Great Lakes Fishery Commission, Special Publication 87-3, Ann Arbor, Michigan.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.-A., Guisan, A., Heikkinen, R.K., Høye, T.T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.-C., Normand, S., Öckinger, E., Schmidt, N.M., Termansen, M., Timmermann, A., Wardle, D.A., Aastrup, P., Svenning, J.-C. 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biol. Rev. 88:15-30.

Wood, C.M., and McDonald, D.G. 1997. Global warming: Implications for freshwater and marine fish. Cambridge University Press, Cambridge. UK.

Woodward, G., Perkins, D.M., Brown, L.E. 2010. Climate change and freshwater ecosystems: impacts across multiple levels of organization. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365:2093-2106.

Woolnough, D.A., Downing, J.A., and Newton, T.J. 2008. Fish movement and habitat use depends on water body size and shape. Ecology of Freshwater Fish 18:83-91.

Yee, T.W., and Dirnböck, T. 2009. Models for analyzing species' presence/absence data at two time points. Journal of Theoretical Biology 259(4):684-694.

Yin, P. and Fan, X. 2001. Estimating R2 shrinkage in multiple regression: a comparison of different analytical methods. The Journal of Experimental Education 69(2):203-224.

Zimmermann, N.E., Edwards, T.C. Jr, Graham, C.H., Pearman, P.B., Svenning, J.-C. 2010. New trends in species distribution modelling. Ecography 33: 985-989.

Zou, G. 2004. A modified Poisson regression approach to prospective studies with binary data. Am. J. Epidemiol. 159:702-706.