1	Using ecological niche-based modeling techniques to predict the establishment, abundance
2	and relative impacts of introduced species: application to the common carp (Cyprinus carpio)
3	
4	Stefanie A. Kulhanek ^{1*} Brian Leung ^{1,2} and Anthony Ricciardi ^{1,3}
5	
6	¹ Department of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada
7	² School of Environment, McGill University, Montreal, Quebec H3A 2A7, Canada
8	³ Redpath Museum, McGill University, Montreal, Quebec, H3A 2K6, Canada
9	
10	
11	*corresponding author: stefanie.kulhanek@mail.mcgill.ca
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	

25 Abstract

26 In order to efficiently manage non-indigenous species (NIS) predictive tools are needed to 27 prioritize locations where they are likely to become established and where their impacts will be 28 most severe. While predicting the impact of a NIS has generally proved challenging, forecasting 29 its abundance patterns across potential recipient locations should serve as a useful surrogate 30 method of estimating the relative severity of the impacts to be expected. Yet such approaches have 31 rarely been applied in invasion biology. We used long-term monitoring data for lakes within the 32 state of Minnesota and artificial neural networks to model both the occurrence as well as the 33 abundance of a widespread aquatic NIS, common carp (*Cyprinus carpio*). We then tested the 34 ability of the resulting models to (i) interpolate to new sites within our main study region, (ii) 35 extrapolate to lakes in the neighboring state of South Dakota and (iii) assessed the relative 36 contribution of each variable to model predictions. Our models correctly identified over 83% of 37 sites where carp are either present or absent and explained 73% of the variation in carp abundance 38 for validation lakes in Minnesota (i.e. lakes not used to build the model). When extrapolated to 39 South Dakota, our models correctly classified carp occurrence in 79% of lakes and explained 32% 40 of the variation in carp abundance. Variables related to climate and water quality were found to be 41 the most important predictors of carp distribution. These results demonstrate that ecological niche-42 based modeling techniques can be used to forecast both the occurrence and abundance patterns of 43 invasive species at a regional scale. Models also yielded sensible predictions when extrapolated to 44 neighboring regions. Such predictions, when combined, should provide more useful estimates of 45 the overall risk posed by NIS on potential recipient systems.

46

Keywords: Non-indigenous species, impact, *Cyprinus carpio*, common carp, ecological nichebased modeling, artificial neural networks

49 Introduction

50 Non-indigenous species (NIS) are an increasing management priority for governments 51 worldwide owing to their potential to cause severe ecological and economic impacts. The 52 ecological effects of invasive species can range from modifications in ecosystem function and 53 community structure to the extirpation or extinction of native species (Lodge 1993, Clavero and 54 Garcia-Berthou 2005). Consequently, NIS are currently recognized as a major threat to 55 biodiversity (Chapin et al. 2000). To mitigate this global problem, many ecologists are aiming to 56 develop tools that enable predictions regarding the invasion process. 57 Environmental conditions have frequently been used to assess the potential for NIS to 58 establish in new geographic locations (e.g. Zambrano et al. 2006, Herborg et al. 2007, Kilroy et al. 59 2008). Such predictions are founded in Hutchinson's (1957) classical niche theory, which states 60 that species distribution patterns are governed by a discrete set of ecological conditions delineating 61 the areas in which a given species can establish and maintain populations at particular densities 62 (Peterson 2003, Araujo and Guisan 2006). As such, ecological niche-based modeling (ENM) 63 techniques, which relate various aspects of species distribution to biologically relevant 64 environmental variables, have become valuable tools for forecasting biological invasions (Peterson 65 and Vieglais 2001). 66 ENM approaches are typically used in invasion biology to predict the presence and absence

of certain NIS or to estimate the probability of their establishment at particular sites (i.e. invasibility) (e.g. Buchan and Padilla 2000, Ficetola et al. 2007). Yet risk assessments of greater management value would be achieved by estimating both the probability of establishment and the severity or magnitude of the impact resulting from the invasion. Unfortunately, the factors that determine the effects of introduced species on their recipient communities are the most poorly understood aspect of the invasion process (Parker et al. 1999, Byers et al. 2002). Furthermore, the impact of any individual NIS can be context dependant and may vary greatly across invaded sites
(Ricciardi & Kipp 2008), posing a challenge to prediction (Ricciardi 2003). Consequently, there
have been few attempts to incorporate quantitative estimates of impact into risk assessment
frameworks (but see Vander Zanden and Olden 2008).

A simple concept drawn from the invasion literature suggests that the severity of the 77 78 impact caused by an introduced species is largely a function of its abundance at the invaded site -79 in general the higher the local density, the greater the impact (Parker et al. 1999, Ricciardi 2003). 80 This intuitive principle is supported by both experimental and observational evidence for 81 numerous NIS (Madsen 1998, Ruiz et al. 1999, Chumchal et al. 2005, Ward and Ricciardi 2007, 82 Pintor et al. 2009) and in the absence of more suitable metrics, various measures of invader 83 abundance (e.g. density, biomass) have occasionally been employed as surrogate measures of 84 impact (e.g. Marchetti et al. 2004). According to niche theory, the abundance patterns of 85 introduced species should also relate to various environmental conditions at potential recipient 86 locations. Thus, we should be able to use ENM approaches to predict abundance, in addition to 87 more coarse metrics of distribution – such as presence and absence (VanDerWal et al. 2009). 88 However, the factors and processes that mediate the abundance of a species may differ from those 89 that determine its occurrence or establishment success (e.g. Ramcharan et al. 1992, Neilsen et al. 90 2005, Heinanen et al. 2008). We may therefore obtain a more comprehensive estimate of invasion 91 risk by forecasting both the probability of establishment as well as the abundance patterns of NIS 92 in newly invaded regions.

Although this proposition is straightforward it has not been widely adopted in invasion
biology, which has rarely used ENM to predict the abundance of introduced species at regional
scales (but see Ramcharan et al. 1992, Koutnik and Padilla 1994, Wilson and Sarnelle 2002). Yet
niche-based models for species abundance are frequently developed beyond the scope of invasion

97 biology, ranging from exploratory analyses of the organism-environment relationships (e.g. Lek et 98 al. 1996, Wiley et al. 2004) to predictive applications intended for conservation and management 99 purposes (e.g. Heinanen et al. 2008, Li et al. 2009). The distinction between invasion biology 100 versus other fields of research is important, given their different objectives. In fields such as 101 conservation biology, researchers often use ENM to predict species abundances or to explain the 102 relationship between species distribution and various environmental factors within the same region 103 where the model was parameterized. In contrast, given the nature of biological invasions, we are 104 often most interested in extrapolating our predictions from ENM developed in one region to new 105 geographical locations.

106 In order to assess the utility of such models, their predictive power should be tested using 107 data that reflect their intended purpose (Pearce and Ferrier 2000, Vaughan and Ormerod 2005). 108 Yet, several authors have noted that a surprisingly large number of ENMs are only evaluated using 109 the same data that were also used to fit the model (Araujo and Guisan 2006, Ozesmi et al. 2006). 110 This can result in overly optimistic assessments of performance and highly inaccurate predictions 111 (Fielding and Bell 1997, Olden et al. 2002). Where researchers have taken the next step and 112 validated ENMs, they have typically done so by partitioning data from a single region into subsets, 113 using one portion of the data to calibrate the model and using the remainder for validation. This 114 approach allows for confident predictions at new sites within the same range where ENM was 115 parameterized, which we term interpolation. However, models should also be evaluated using data 116 that are spatially or temporally separated from those used for calibration in order to assess the 117 ability to extrapolate predictions to new geographical areas or temporal horizons (Vaughan and 118 Ormerod 2005). While the popularity of niche-based models in invasion biology is increasing, and 119 there is often the implicit assumption that predictions can be applied to new locations, tests of 120 extrapolation remain relatively rare. Further, as ENMs for forecasting the abundance of introduced

species are uncommon in the invasion literature, and given that such models have rarely been developed for similar predictive purposes in other fields, the ability to extrapolated model predictions to new geographic locations needs to be assessed.

124 In this study we develop ecological niche-based models for predicting the occurrence as 125 well as the abundance of a widespread vertebrate invader, common carp (*Cyprinus carpio*). Using 126 data for lakes within the state of Minnesota, we developed artificial neural networks to forecast 127 these two aspects of carp distribution from several limnological and climatic variables. We then 128 examined the accuracy with which our models were able to interpolate to sites within the main 129 study region and extrapolate to independent data for lakes within the neighboring state of South 130 Dakota. We also estimated the relative contribution of environmental variables to model 131 predictions. By incorporating abundance as a surrogate metric for impact and examining the 132 degree to which models for both the occurrence and abundance of a highly invasive species can 133 extrapolate to new sites, this study assesses the viability of using ENM methods to more fully 134 model the risk posed by NIS on their recipient communities.

135

136 Methods

137 Model organism

Common carp was selected as a model species for this study owing to the availability of data regarding its distribution and impacts. The common carp is native to Eurasia, but has been introduced across the globe both deliberately, for aquaculture, recreational and ornamental purposes, and unintentionally due to live bait release and other vectors (Balon 1995, Koehn 2004). Carp were first introduced into the United States in the late 1870s and were subsequently spread throughout the country, reaching the mid-west, including Minnesota, shortly after the initial introduction (Cole 1905). The common carp is currently considered to be one of the world's most ecologically harmful invasive species (Lowe et al. 2004). Its impacts arise mainly from the ability
to alter aquatic habitats through high levels of excretion and by disturbing the bottom sediments of
lakes and other waterbodies to which it had been introduced; often resulting in increased turbidity,
degraded water quality and reduced macrophyte and benthic invertebrate densities (e.g. Zambrano
and Hinojosa 1999, Parkos et al. 2003, Matsuzaki et al. 2007). Furthermore, the severity of these
effects has been shown to be highly dependant on local carp biomass (Robel 1961, Crivelli 1983,
Lougheed et al. 1998, Chumchal et al. 2005).

152 The physiological tolerances and habitat preferences of carp have been investigated by 153 several authors (Crivelli 1981, Balon 1995, Garcia-Berthou 2001, Penne and Pierce 2008), 154 providing a sound basis for predictor variable selection. Although carp prefer shallow, warm, slow 155 moving waterbodies and have a high tolerance of eutrophic waters, established populations have 156 been found under a broad range of environmental conditions (Koehn 2004, Schade and Bonar 157 2005). Further, while ENM techniques have previously been used to predict the full potential 158 invaded range of carp in North and South America (Zambrano et al. 2006), the occurrence and 159 abundance patterns of this invader have not yet been modeled at a regional scale.

160

161 *Data collection*

162 The abundance and distribution records used to develop the models were obtained from the 163 Minnesota Department of Natural Resources (MNDNR). Carp currently occur throughout most of 164 the southern and some of the north-western areas of Minnesota and are established in over 800 165 lakes that are routinely surveyed by the department. While the MNDNR uses several sampling 166 methods to assess fish populations, we restricted our analysis to biomass catch per unit effort 167 (BPUE) data from standard summer trap net sets, as this method is most effective at capturing 168 various benthivorous fish species, including carp (A. Stevens, MNDNR, pers. comm.).

169 The frequency with which a particular lake is surveyed depends largely on its size, 170 recreational value and various logistic factors. BPUE measures can fluctuate from one survey 171 event to another due to factors such as winter fish kills, atypically large age-0 year classes or other 172 stochastic events. We sought to minimize the potential effects of sampling bias and short term 173 fluctuations in BPUE in our models. We did this by limiting our analysis to lakes that had been 174 sampled for their fish populations a minimum of 3 times between 1980 and 2007 and derived a 175 mean estimate of BPUE from repeat samples. We believe that this metric best reflects long-term 176 equilibrium densities.

177 Independent variables with potential distributional importance were compiled from several 178 main sources. Climatic variables, consisting of 20-year averages, were obtained from the National 179 Climate Data Center. For each lake, climate data was extracted from the nearest sampling station, 180 generally located within 10 to 30 km of the site. Lake morphometry data were provided by the 181 MNDNR, while water chemistry variables, which consisted of 10-year summer averages, were obtained from the Minnesota Pollution Control Agency and the Environmental Protection 182 183 Agency's STORET database. In most cases, variables within each category (i.e. climate, 184 morphometry, water chemistry) formed several tightly correlated subsets. As a high degree of 185 correlation between predictor variables is generally undesirable for ENM applications, the number 186 of predictors was reduced after removing all but one variable within each highly correlated 187 grouping (Spearman's $|\mathbf{r}| > 0.8$), while taking into consideration the availability of data for each 188 predictor.

189 The final set of variables included maximum and minimum annual air temperature, annual 190 precipitation, growth degree days (annual sum of daily air temperature above 5°C), lake surface 191 area and maximum depth, total nitrogen concentrations, total alkalinity, and Carlson's trophic state 192 index (TSI). TSI is a common measure of lake productivity and can be derived from Chlorophyll-*a* 193 concentrations, total phosphorus, secchi depth measurements, or a combination of these factors194 (Carlson 1977). Variables are summarized in Table 1.

The final Minnesota database consisted of 285 carp-invaded lakes and 238 lakes where carp have been historically absent, after removing all entries that did not meet our criteria or for which data on predictor variables were unavailable. These data were used to derive (i) the abundance data set which contained mean BPUE values and respective environmental predictors for the 285 lakes containing carp and (ii) the occurrence data set which contained information on all 523 water bodies, where carp BPUE was converted to a dichotomous variable representing presence or absence.

202

203 External validation data

204 To assess the degree to which we might generalize from our sample of Minnesota lakes, we 205 compiled information for an additional 38 lakes in eastern South Dakota. Carp distribution data 206 were obtained from the South Dakota Game, Fish and Parks Commission (SDGFP), which 207 employs a similar sampling protocol and equipment to that used by the MNDNR. This data set 208 consisted of mean BPUE values and presence-absence data derived from the past 10 years of trap 209 net sampling. Environmental data corresponding to this set of lakes were compiled from a 210 database maintained by the South Dakota Department of Environment and Natural Resources and 211 additional sources cited above.

212

213 Model development and validation: Artificial neural networks

We used multi-layer feed-forward artificial neural networks (ANN) trained by backpropagation (Rumelhart et al. 1986) to model the occurrence and abundance of carp. We chose to

216 work with ANN because this method requires no prior assumptions about underlying distributions,

217 can account for nonlinearity and interactions between variables, and has demonstrated a high 218 predictive power compared to several common ENM methods, including generalized linear 219 models, discriminant analysis, classification and regression trees and other machine learning 220 methods (Baran et al. 1996, Ozesmi and Ozesmi 1999, Tan et al. 2006, Olden et al. 2008). ANN 221 have previously been applied to many problems in ecology (e.g. Lek and Guegan 1999, Ozesmi et 222 al. 2006, Goethals et al. 2007), and have shown promising performance in their ability to model 223 both the occurrence (Vander Zanden et al. 2004) and abundance patterns (Baran et al. 1996, 224 Brosse and Lek 2002) of several freshwater fish species. For more information concerning ANN, 225 their implementation, comparison to other modeling methods and available software we refer 226 readers to Bishop (1995) as well as Olden et al. (2008).

227 Our feed-forward networks consisted of multiple interconnected layers of processing 228 elements, often termed neurons. These included an input layer, representing each of our 9 229 predictor variables, one to two hidden layers and an output node, each connected by a set of 230 adjustable parameters (i.e. weights). During the training process, variables associated with each set 231 of observations are fed through the network, multiplied by their respective weights, summed and 232 transformed into an output signal by applying a transformation function. This process is repeated 233 at each hidden layer until the signal reaches the output neurons where the values, corresponding to 234 the variable being predicted, are calculated. Using this output, the mean squared error (MSE) 235 between predicted and observed values was computed and used to adjust the weights between the 236 neurons in each layer, by applying the Levenberg-Marquardt backpropagation algorithm (Hagan 237 and Menhaj 1994). This process was repeated for multiple epochs (i.e. training iterations) to 238 sequentially minimize the MSE between observed and predicted outputs.

Prior to training, input variables were proportionally scaled to a range of -1 to 1. This was
done in order to standardize the units of measurement associated with different predictor variables,

241 ensuring that each receives equal attention during training (Goethals et al. 2007). We then 242 separated the data into three components: the calibration set, and the internal and the external 243 validation sets. The calibration data, which consist of ~80% of lakes randomly selected from the 244 complete Minnesota dataset, were used to fit the models and optimize network architecture. The 245 remaining 20% of the Minnesota data was allotted to the internal validation set, which was later 246 used to evaluate the ability of our models to interpolate within the main study region. Data for the 247 South Dakota lakes (i.e. the external validation set) were used to test the model's ability to 248 extrapolate to independent sites.

To limit over-fitting the networks, we used a form of cross validation known as early stopping (Prechelt 1998). Before training, the calibration data were randomly divided into training and test sets, corresponding to roughly 60% and 20% of the full Minnesota data respectively. During the training process, the test data were employed to limit the number of training iterations by terminating training when test set MSE did not decrease during 5 subsequent epochs. This procedure did not directly influence weight adjustments but rather was used to improve the ability of the fitted model to generalize to new data.

256 The optimal number of hidden layers and neurons within them was determined empirically 257 by creating multiple networks, with all other parameters held constant. Network performance can 258 be sensitive to the random initial weight values set prior to training (Ozesmi et al. 2006, Olden et 259 al. 2008). For this reason, 100 networks were run based on the same architecture, after resetting 260 the initial weights to small random values, and performance was assessed based on the averaged 261 predictions across all 100 runs. For the occurrence data we tested networks with no hidden layer 262 and a single hidden layer containing between 1 and 15 neurons. Models with no hidden layer were 263 included to evaluate the performance of ANN against what is essentially a generalized linear 264 model (Ozesmi et al. 2006). For the abundance data, in addition to testing models with zero and

265 one hidden layer, we also examined models containing 2 hidden layers, since during the 266 optimization procedure networks with a single layer had a consistent bias toward underestimating 267 high BPUE values. All networks were constructed and trained using the version 5 of the Neural 268 Network Toolbox in MATLAB (The MathWorks, Natick, Massachusetts). 269 The relative importance of each predictor variable was assessed using Garson's algorithm 270 (Garson 1991). This procedure consisted of calculating the product of the weighting factors 271 between each input-output neuron connection, summing the products across all hidden neurons, 272 and calculating a percentage representing the individual contribution of each variable to network 273 predictions. For more information on calculating variable importance using Garson's algorithm, 274 we refer readers to Goh (1995). 275 276 Model Performance criteria 277 For the occurrence model, network outputs take on continuous values ranging between 0 278 and 1, representing the probability of carp presence at each site. In order to assess model accuracy, 279 we opted to employ a threshold selection procedure that maximized the sum of model sensitivity 280 (the proportion of correctly classified presences) and model specificity (the proportion of correctly 281 classified absences), as the arbitrary choice of a threshold probability at which the carp were 282 deemed either present or absent (e.g. 0.5) could greatly influence our results (Manel et al. 2001). 283 This procedure consisted of developing Receiver-Operating Characteristic (ROC) curves (i.e. plots 284 of model sensitivity against 1-specificity across different thresholds) and determining the threshold 285 at which the sum of these values was at its maximum (Manel et al. 2001, Jimenez-Valverde and 286 Lobo 2007). Using this threshold, we then generated confusion matrices: 2 x 2 tables containing 287 the true positive, false positive, true negative and false negative rate associated with each network 288 (Fielding and Bell 1997). These matrices as well as the ROC itself were used to derive four

289 performance measures including: (1) the percentage of correct classification instances (CCI) (i.e. 290 the percentage of sites where the model correctly predicted either presence or absence); (2) model 291 sensitivity; (3) model specificity; and (4) the area under the ROC curve (AUC), a measure that is 292 independent of threshold selection (Fielding and Bell 1997, Pearce and Ferrier 2000). 293 In contrast to the occurrence model, the outputs of the abundance networks are inherently 294 continuous. Performance was therefore evaluated based on regression analysis of the fit between observed and predicted values. The slope (m), intercept (b) and goodness of fit (R^2) of this 295 296 relationship were used to derive the model non-ideality index (δ): the Euclidean distance between the observed, $|m, b, R^2|$, and the ideal agreement vectors |1, 0, 1|, which would be obtained from a 297 298 perfect fit (Plumb et al. 2005). Networks with the lowest δ were considered to provide the best predictions. 299 300 Once the optimal network architectures were determined based on the performance for the 301 calibration data, the resulting models were used to generate predictions of expected carp 302 occurrence and abundance for both the internal and external validation sets. As with the calibration 303 data, predictions were obtained by averaging outputs across all 100 runs of the optimal networks. 304 Performance on the validation data was assessed as described above, with the exception that, for

305 occurrence predictions, the threshold probability for carp presence was derived from calibration306 data results.

307

308 **Results**

309 *Occurrence model*

The optimal presence-absence network consisted of 9 input neurons, corresponding to each of the predictor variables (Table 1), one hidden layer containing 5 neurons, and a single output node. Sigmoid transfer functions were used at each layer. All networks that included a hidden

313 layer outperformed those with no hidden neurons (i.e. GLM). For the calibration data, the model 314 correctly identified 197 of 228 (86 %) lakes where carp are present and 171 of 191 (90 %) lakes 315 where carp have been historically absent. When applied to the internal validation set, the model 316 correctly predicted carp presence in 49 of 57 (86 %) lakes and carp absence in 39 of 48 (81 %) 317 lakes. When extrapolated to South Dakota lakes, carp presence and absence was correctly predicted at 25 of 33 (76 %) and 4 out of 5 sites (80 %), respectively (Figure 1). 318 319 Performance criteria including CCI, sensitivity, specificity, AUC and respective *P*-values for each data set are presented in Table 2. For all indices, performance on the internal validation 320 321 set was comparable to that on the calibration data but slightly reduced when applied to the South 322 Dakota lakes. For AUC, values of 0.5 or lower indicate that model predictions are no better than random, while values above 0.5 are increasingly accurate. According to Pearce and Ferrier (2000), 323 324 the AUC values obtained indicate very good discrimination for the Minnesota lakes and 325 reasonable predictive power when applied to the external validation data.

326

327 *Abundance model*

328 The optimized network architecture for carp biomass per unit effort (BPUE) consisted of 9 329 input neurons, two sigmoid hidden layers (containing 6 and 3 neurons, respectively), and one 330 linear output neuron. As with the occurrence model, networks with a hidden layer performed better 331 than those with no hidden neurons but optimal performance was achieved when two hidden layers 332 were included. Observed carp BPUE was significantly related to model predictions for the calibration set ($R^2 = 0.76$, n = 228, F = 735.0, p < 0.0001), the internal validation set ($R^2 = 0.73$, n333 = 57, F = 145.8, p < 0.0001) and for the South Dakota data ($R^2 = 0.32$, n = 33, F = 15.90, p = 15334 335 0.0004), Figure 2 a, b and c.

336	The non-ideality index (δ) was 0.24, 0.41 and 2.03 for the calibration data, internal
337	validation set and South Dakota lakes, respectively. For each dataset the intercepts of the fitted
338	regression line did not differ significantly from zero ($t = 0.06, 0.73, 1.49, p = 0.95, 0.47, 0.15$ for
339	the calibration, internal and external validation data respectively). The slope of the relationship
340	between predicted and observed BPUE did not differ significantly from 1 for both the calibration
341	(t = -0.75, p = 0.45) and the internal validation data $(t = 0.1, p = 0.92)$; although for the South
342	Dakota dataset there was a significant departure from a one to one relationship ($t = -2.10$, $p = 0.04$).
343	However, despite a loss of predictive power, the 3 lakes with the highest BPUE observed in the
344	South Dakota dataset, were among the 5 lakes predicted to support the highest carp densities. The
345	model also correctly identified 8 of 10 lakes with the lowest observed carp BPUE.
346	
347	Predictor variable contribution
348	The relative contribution of each predictor variable is illustrated in Figure 3. For both the
349	carp occurrence and abundance models, limnological variables were weighted more heavily than
350	climate variables, with the exception of minimum annual temperature. For the occurrence model,
351	the variables with the largest contributions to network predictions included trophic state index
352	(TSI), minimum annual temperature, and total alkalinity. For the abundance model, the most
353	highly weighted variables included TSI and total nitrogen, with lake morphometry and climate
354	variables contributing more evenly to network outputs.
355	
356	Discussion
357	Using a neural network approach, we were able to predict both the occurrence and the

358 abundance of common carp and to extrapolate model predictions to new data. Our occurrence

359 model correctly identified over 75% of all sites where this invader is either present or absent, even

360 when generating predictions in locations independent from those where the model was 361 parameterized. Predictions at new sites within the same region and in independent locations 362 explained 73% and 32% of variation in carp biomass, respectively. Extrapolation to new 363 geographic regions, while rarely done, represents the strictest test of ecological niche-based 364 models intended for management application. Although the predictive power of our carp 365 abundance network was reduced when applied to South Dakota lakes, the model still explained a 366 significant portion of biological variation. At a minimum, this degree of transferability would 367 allow for a relative ranking of sites vulnerable to carp invasion within the area.

368 Indeed, based on the 5 sites predicted to support the highest carp biomasses, we were able 369 to identify the 3 most heavily carp-infested lakes in our South Dakota dataset. Cumulatively, these 370 3 waterbodies account for over 36% of total observed carp BPUE. It is also interesting to note that 371 the lake for which carp biomass was most substantially overestimated is a small eutrophic 372 waterbody that experiences frequent fishkills and is also commercially harvested for carp, which 373 may partially explain the discrepancy between observed and predicted BPUE for this site (D. 374 Lucchesi, SDGFP, pers. comm.). When this data point is removed, the strength of the relationship between observed and predicted BPUE increases substantially ($R^2 = 0.43$, n = 32, F = 22.97, p < 100375 376 0.0001) and the non-ideality index decreases to 1.66. Further, our model also correctly identified 8 377 of the 10 lakes where carp impacts would be expected to be minimal. Thus, while rarely employed 378 in the study of biological invasions, using ENM techniques to predict the abundance of introduced 379 species can be a viable approach to gaining additional insight into the risk they pose on potential 380 recipient habitats, even when extrapolating predictions to new locations.

381

382 Joint model approach

383 By combining models for predicting the establishment success of introduced species with 384 those for forecasting abundance, we should be able to generate more comprehensive measures of 385 the relative risk posed by particular NIS, and allocate prevention efforts accordingly. Such joint-386 model approaches have recently been advocated in the literature, with the general consensus that, 387 where possible, multiple elements of the invasion process should be examined in concert to 388 provide improved predictions. Most existing studies have focused on combining models of NIS 389 introduction effort (i.e. propagule pressure) with those for predicting invasibility (e.g. ENM for 390 occurrence) to estimate the overall risk of establishment (e.g. Herborg et al. 2007, Leung and 391 Mandrak 2007, Jacobs and Macisaac 2009), but few have explicitly considered impacts. Yet by 392 combining models for occurrence, surrogate measures of propagule pressure, and quantitative analyses of the effects caused by introduced species, several researchers have demonstrated the 393 394 addevalue of integrating impacts into predictive models of invasion risk (Vander Zanden et al. 395 2004, Mercado-Silva et al. 2006, Vander Zanden and Olden 2008).

396 By prioritizing management of the small subset of invasible South Dakota lakes expected 397 to support high carp densities, we would be able to mitigate a substantial portion of the impact to 398 be expected in the area. Thus, recent approaches that have been developed to incorporate impact, 399 and additional aspects of the invasion process, should improve predictive power and significantly 400 reduce the number of sites thought to require management intervention, rather than limiting 401 analyses to a single component. As demonstrate here, models for forecasting the abundance of 402 particular NIS can provide insightful predictions and identify sites are most likely to at risk of 403 negative impacts, beyond those which are simply invasible. The utility of such models is however 404 dependant upon their ability to predict invasions in new areas. For invasive species in particular, 405 tests of extrapolation are critical when evaluating the utility of ENMs, given that we are primarily

406 interested in generating predictions for presently uninvaded locations where invasive species may407 eventually spread.

408

409 *Predictor variable contribution*

410 Of the factors that may influence the transferability, as well as the overall accuracy of 411 niche-based models, the choice of predictor variables is of particular importance (Vaughan and 412 Ormerod 2003, Araujo and Guisan 2006). Thus, it is relevant to further examine the relative 413 contribution of specific variables to our model predictions. For both the carp occurrence and 414 abundance models, most of the variables contributing strongly to network predictions are directly 415 interpretable. For example, minimum annual temperature, which was found to be a significant 416 predictor of carp occurrence, affects carp spawning activity and can also cause winter fishkills, 417 both of which can limit the distribution of this species (Balon 1995). Low winter temperatures may 418 currently restrict carp to the southern two thirds of Minnesota but may not be as important a factor 419 in determining their distribution in South Dakota, where minimum annual temperatures are 420 somewhat less extreme.

Alkalinity, another relevant predictor for carp occurrence, is essentially a measure of
buffering capacity, and influences many important limnological parameters, primarily by
controlling pH. For example, water acidity can influence carp by affecting larval development
(Korwinkossakowski 1988) and macro-invertebrate densities – an important food source for carp
(Garcia-Berthou 2001). Alkalinity may therefore influence carp distribution both directly and
indirectly by affecting numerous biotic processes.

427 Variables related to trophic state (i.e. TSI, TN) were found to be relevant predictors for
428 both carp occurrence and abundance. The contribution of these factors likely reflects the high
429 tolerance of carp to eutrophic conditions and their tendency to establish and become dominant in

even heavily degraded waterbodies (Schade and Bonar 2005). However, while carp have an
affinity for eutrophic conditions, they also tend to perpetuate this state through their feeding
activities and excretion (Parkos et al. 2003, Chumchal et al. 2005, Matsuzaki et al. 2007). As such,
the contribution of these variables may also partly be explained by the impacts that carp have had
on the waterbodies to which they have been introduced. Unfortunately, given the historic
introduction of carp to Minnesota, resulting in a lack of data prior to invasion, we were unable to
investigate of the relative importance of this factor.

The contribution of most variables suggests that, to some extent, different environmental conditions are important in determining carp occurrence and its abundance. Within the context of ENM, similar results have been reported by several authors (e.g. Ramcharan et al. 1992, Neilson et al. 2005, Heinanen et al. 2008). Although models developed from occurrence data have been shown to predict the upper limit of abundance for a variety of species (VanDerWal et al. 2009), we would nonetheless caution against the interpretation of the probability of NIS occurrence alone as representative of its likely abundance or potential impacts.

444

445 Conclusion

446 Our results suggest that ecological niche-based modeling methods, similar to those 447 commonly employed to forecast the occurrence of invasive species, can be used to develop 448 accurate predictions for measures of invader abundance. Such models can be particularly useful 449 for forecasting the relative severity of the impacts of NIS across sites yet to be invaded. When the 450 relationship between an invader's abundance and its impacts is known, models that predict 451 abundance from local environmental variables could also be used to quantitatively forecast the 452 magnitude of particular impacts at potentially invasible sites, thereby contributing to a more 453 informative assessment of future invasion threats. Further, when combined with predictions

regarding establishment success, such models can lead to more informed measures for the overall
risk posed by particular NIS and can allow for identification of sites where management
interventions are most needed.

457 The accuracy and transferability of such models must be rigorously tested in order to assess 458 their utility for practical management applications. As ENM for invasive species are often 459 intended to inform management decisions at currently uninvaded sites, the influence of various 460 factors that can affect the ability of such models to extrapolate to new areas merits more thorough 461 investigation. Finally, the ability of many NIS to modify conditions that influence their own 462 abundance through positive feedbacks (Gonzalez et al. 2008) must be also addressed, as such 463 effects may compromise the ability of models developed from current abundance data to generate 464 useful predictions in uninvaded habitats.

465

Acknowledgements - We thank the Minnesota Department of Natural Resources, South Dakota
Game, Fish and Parks commission, Minnesota Pollution Control Agency and South Dakota
Department of Environment and Natural Resources for providing the monitoring data used in this
study. We also thank K. Hasegawa for assistance in data compilation and two anonymous
reviewers for their valuable comments on the manuscript. This work was funded by le Fonds
Québécois de la Recherche sur la Nature et les Technologies and by the Natural Sciences and
Engineering Research Council of Canada through grants awarded to Dr. Brian Leung.

474 Literature Cited

475 Araujo, M. B. and A. Guisan. 2006. Five (or so) challenges for species distribution modelling.
476 Journal of Biogeography 33:1677-1688.

- Balon, E. K. 1995. Origin and domestication of the wild carp, *Cyprinus carpio*: from Roman
 gourmets to the swimming flowers. Aquaculture 129:3-48.
- Baran, P., S. Lek, M. Delacoste, and A. Belaud. 1996. Stochastic models that predict trout
 population density or biomass on a mesohabitat scale. Hydrobiologia 337:1-9.
- 481 Bishop, C. M. 1995. Neural networks for pattern recognition. Clarendon Press, Oxford, UK
- Brosse, S. and S. Lek. 2002. Relationships between environmental characteristics and the density
 of age-0 Eurasian perch *Perca fluviatilis* in the littoral zone of a lake: A nonlinear approach.
- 484 Transactions of the American Fisheries Society **131**:1033-1043.
- 485 Buchan L. A. J. and D. K. Padilla. 2000. Predicting the likelihood of Eurasian watermilfoil
- 486 presence in lakes, a macrophyte monitoring tool. Ecological Applications **10**:1442-1455.
- 487 Byers, J. E., S. Reichard, J. M. Randall, I. M. Parker, C. S. Smith, W. M. Lonsdale, I. A. E.
- 488 Atkinson, T. R. Seastedt, M. Williamson, E. Chornesky, and D. Hayes. 2002. Directing
 489 research to reduce the impacts of nonindigenous species. Conservation Biology 16:630-640.
- 490 Carlson, R. E. 1977. A trophic state index for lakes. Limnology and Oceanography 22:361-369.
- 491 Chapin, F. S., E. S. Zavaleta, V. T. Eviner, R. L. Naylor, P. M. Vitousek, H. L. Reynolds, D. U.
- 492 Hooper, S. Lavorel, O. E. Sala, S. E. Hobbie, M. C. Mack, and S. Diaz. 2000.
- 493 Consequences of changing biodiversity. Nature **405**:234-242.
- Chumchal, M. M., W. H. Nowlin, and R. W. Drenner. 2005. Biomass-dependent effects of
 common carp on water quality in shallow ponds. Hydrobiologia 545:271-277.
- 496 Clavero, M. and E. Garcia-Berthou. 2005. Invasive species are a leading cause of animal
 497 extinctions. Trends in Ecology & Evolution 20:110-110.
- 498 Cole, L. J. 1905. The Status of the Carp in America. Transactions of the American Fisheries
 499 Society 34:201-206.

- 500 Crivelli, A. J. 1981. The biology of the common carp, *Cyprinus carpio* L. in the Camargue,
 501 Southern France. Journal of Fish Biology 18:271-290.
- 502 Crivelli, A. J. 1983. The destruction of aquatic vegetation by carp *Cyprinus carpio* a comparison
 503 between Southern France and the USA. Hydrobiologia 106:37-42.
- Ficetola, G. F., W. Thuiller, and C. Miaud. 2007. Prediction and validation of the potential global
 distribution of a problematic alien invasive species the American bullfrog. Diversity and
 Distributions 13:476-485.
- Fielding, A. H. and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in
 conservation presence/absence models. Environmental Conservation 24:38-49.
- Garcia-Berthou, E. 2001. Size and depth-dependent variation in habitat and diet of the common
 carp (*Cyprinus carpio*). Aquatic Sciences **63**:466-476.
- 511 Garson, G. D. 1991. Interpreting neural-network connection weights. AI Expert **6**:46-51.
- 512 Goethals, P. L. M., A. P. Dedecker, W. Gabriels, S. Lek, and N. De Pauw. 2007. Applications of
 513 artificial neural networks predicting macroinvertebrates in freshwaters. Aquatic Ecology
 514 41:491-508.
- 515 Goh, A. T. C. 1995. Backpropagation neural networks for modeling complex systems. Artificial
 516 Intelligence in Engineering 9:143-151.
- 517 Gonzalez, A., A. Lambert, and A. Ricciardi. 2008. When does ecosystem engineering facilitate
 518 invasion and species replacement? Oikos 117:1247–1257.
- Hagan, M. T. and M. B. Menhaj. 1994. Training feedforward networks with the Marquardt
 algorithm. IEEE Transactions on Neural Networks 5:989-993.
- 521 Heinanen, S., M. Ronka, and M. von Numers. 2008. Modelling the occurrence and abundance of a
- 522 colonial species, the arctic tern *Sterna paradisaea* in the archipelago of SW Finland.
- 523 Ecography **31**:601-611.

- Herborg, L. M., C. L. Jerde, D. M. Lodge, G. M. Ruiz, and H. J. MacIsaac. 2007. Predicting
 invasion risk using measures of introduction effort and environmental niche models.
 Ecological Applications 17:663-674.
- 527 Hutchinson, G. E. 1957. Concluding Remarks. Cold Spring Harbor Symposia on Quantitative
- 528 Biology **22**:415-427.
- Jacobs, M. J. and H. J. Macisaac. 2009. Modelling spread of the invasive macrophyte *Cabomba caroliniana*. Freshwater Biology 54:296-305.
- Jimenez-Valverde, A. and J. M. Lobo. 2007. Threshold criteria for conversion of probability of
 species presence to either-or presence-absence. Acta Oecologica **31**:361-369.
- Kilroy, C., T. H. Snelder, O. Floerl, C. C. Vieglais, and K. L. Dey. 2008. A rapid technique for
 assessing the suitability of areas for invasive species applied to New Zealand's rivers.
 Diversity and Distributions 14:262-272.
- Koehn, J. D. 2004. Carp (*Cyprinus carpio*) as a powerful invader in Australian waterways.
 Freshwater Biology 49:882-894.
- Korwinkossakowski, M. 1988. Larval development of carp, *Cyprinus carpio* L, in acidic water.
 Journal of Fish Biology **32**:17-26.
- 540 Koutnik, M. A. and D. K. Padilla. 1994. Predicting the spatial distribution of Dreissena
- 541 *polymorpha* (zebra mussel) among inland lakes of Wisconsin: modeling with a GIS
- 542 Canadian Journal of Fisheries and Aquatic Sciences **51**:1189-1196.
- 543 Lek, S., A. Belaud, P. Baran, I. Dimopoulos, and M. Delacoste. 1996. Role of some environmental
 544 variables in trout abundance models using neural networks. Aquatic Living Resources
 545 9:23-29.
- Lek, S. and J. F. Guegan. 1999. Artificial neural networks as a tool in ecological modelling, an
 introduction. Ecological Modelling 120:65-73

548	Leung, B. and N. E. Mandrak. 2007. The risk of establishment of aquatic invasive species: joining
549	invasibility and propagule pressure. Proceedings of the Royal Society B-Biological
550	Sciences 274 :2603-2609.
551	Li, J., D. W. Hilbert, T. Parker, and S. Williams. 2009. How do species respond to climate change
552	along an elevation gradient? A case study of the grey-headed robin (Heteromyias
553	albispecularis). Global Change Biology 15:255-267.
554	Lodge, D. M. 1993. Biological invasions - lessons for ecology. Trends in Ecology & Evolution
555	8 :133-137.
556	Lougheed, V. L., B. Crosbie, and P. Chow-Fraser. 1998. Prediction on the effect of common carp
557	(Cyprinus carpio) exclusion on water quality, zooplankton, and submergent macrophytes
558	in a Great Lakes wetland. Canadian Journal of Fisheries and Aquatic Sciences 55:1189-
559	1197.
560	Lowe, B., M. Browne, S. Boudjelas and M. De Poorter. 2004. 100 of the World's Worst Invasive
561	Alien Species. The Invasive Species Specialist Group (ISSG) of the World Conservation
562	Union (IUCN). Auckland, New Zealand.
563	Madsen, J. D. 1998. Predicting invasion success of Eurasian watermilfoil. Journal of Aquatic Plant
564	Management 36 :28-32.

Manel, S., H. C. Williams, and S. J. Ormerod. 2001. Evaluating presence-absence models in
 ecology: the need to account for prevalence. Journal of Applied Ecology 38:921-931.

- 567 Marchetti, M. P., P. B. Moyle, and R. Levine. 2004. Alien fishes in California watersheds:
- 568 Characteristics of successful and failed invaders. Ecological Applications **14**:587-596.
- 569 Matsuzaki, S. S., N. Usio, N. Takamura, and I. Washitani. 2007. Effects of common carp on
- 570 nutrient dynamics and littoral community composition: roles of excretion and bioturbation.
- 571 Fundamental and Applied Limnology **168**:27-38.

- 572 Mercado-Silva, N., J. D. Olden, J. T Maxted, T. R. Hrabik, and M. J. Vander Zanden. 2006.
- 573 Forecasting the spread of invasive rainbow smelt in the Laurentian Great Lakes region of
 574 North America. Conservation Biology 20:1740-1749.
- 575 Neilsen, S. E., C. J. Johnson, D. C. Heard, and M. S. Boyce. 2005. Can models of presence-
- absence be used to scale abundance? Two case studies considering extremes in life history.
 Ecography 28:197-208.
- Olden, J. D., D. A. Jackson, and P. R. Peres-Neto. 2002. Predictive models of fish species
 distributions: A note on proper validation and chance predictions. Transactions of the
 American Fisheries Society 131:329-336.
- 581 Olden, J. D., J. J. Lawler, and N. L. Poff. 2008. Machine learning methods without tears: A primer
 582 for ecologists. Quarterly Review of Biology 83:171-193.
- 583 Ozesmi, S. L. and U. Ozesmi. 1999. An artificial neural network approach to spatial habitat
 584 modelling with interspecific interaction. Ecological Modelling 116:15-31.
- Ozesmi, S. L., C. O. Tan, and U. Ozesmi. 2006. Methodological issues in building, training, and
 testing artificial neural networks in ecological applications. Ecological Modelling 195:83 93.
- 588 Parker, I. M., D. Simberloff, W. M. Lonsdale, K. Goodell, M. Wonham, P. M. Kareiva, M. H.

589 Williamson, B. Von Holle, P. B. Moyle, J. E. Byers, and L. Goldwasser. 1999. Impact:

- 590 toward a framework for understanding the ecological effects of invaders. Biological
- 591 Invasions **1**:3–19.
- Parkos, J. J., V. J. Santucci, and D. H. Wahl. 2003. Effects of adult common carp (*Cyprinus carpio*)
 on multiple trophic levels in shallow mesocosms. Canadian Journal of Fisheries and
 Aquatic Sciences 60:182-192.

- 595 Pearce, J. and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed
 596 using logistic regression. Ecological Modelling 133:225-245.
- 597 Penne, C. R. and C. L. Pierce. 2008. Seasonal distribution, aggregation, anal habitat selection of
 598 common carp in Clear Lake, Iowa. Transactions of the American Fisheries Society
 599 137:1050-1062.
- Peterson, A. T. 2003. Predicting the geography of species' invasions via ecological niche modeling.
 Quarterly Review of Biology **78**:419-433.
- Peterson, A. T. and D. A. Vieglais. 2001. Predicting species invasions using ecological niche
 modeling: New approaches from bioinformatics attack a pressing problem. Bioscience
 51:363-371.
- Pintor, L. M., A. Sih, and J. L. Kerby. 2009. Behavioral correlations provide a mechanism for
 explaining high invader densities and increased impacts on native prey. Ecology 90:581587.
- Plumb, A. P., R. C. Rowe, P. York, and M. Brown. 2005. Optimisation of the predictive ability of
 artificial neural network (ANN) models: A comparison of three ANN programs and four
 classes of training algorithm. European Journal of Pharmaceutical Sciences 25:395-405.
- 611 Prechelt, L. 1998. Automatic early stopping using cross validation: quantifying the criteria. Neural
 612 Networks 11:761-767.
- Ramcharan, C. W., D. K. Padilla, and S. I. Dodson. 1992. Models to predict potential occurrence
 and density of the zebra mussel, *Dreissena polymorpha* Canadian Journal of Fisheries and
 Aquatic Sciences 49:2611-2620.
- 616 Ricciardi, A. 2003. Predicting the impacts of an introduced species from its invasion history: an
- 617 empirical approach applied to zebra mussel invasions. Freshwater Biology **48**:972-981.

- Ricciardi, A. and R. Kipp. 2008. Predicting the number of ecologically harmful exotic species in
 an aquatic system. Diversity and Distributions 14:374–380.
- 620 Robel, R. J. 1961. The effect of carp populations on the production of waterfowl food plants on a
- western waterfowl marsh. Transactions of the North American Wildlife and Natural
 Resources Conference 26:147-159.
- 623 Ruiz, G. M., P. Fofonoff, A. H. Hines, and E. D. Grosholz. 1999. Non-indigenous species as
- 624 stressors in estuarine and marine communities: Assessing invasion impacts and interactions.
 625 Limnology and Oceanography 44:950-972.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. Learning representations by backpropagating errors. Nature 323:533-536.
- Schade, C. B. and S. A. Bonar. 2005. Distribution and abundance of nonnative fishes in streams of
 the Western United States. North American Journal of Fisheries Management 25:13861394.
- Tan, C. O., U. Ozesmi, M. Beklioglu, E. Per, and B. Kurt. 2006. Predictive models in ecology:
 Comparison of performances and assessment of applicability. Ecological Informatics
 1:195-211.
- The MathWorks Inc. 2007. MATLAB version 7.5.0. Natick, Massachusetts.
- 635 VanDerWal, J., L. P. Shoo, C. N. Johnson, and S. E. Williams. 2009. Abundance and the
- environmental niche: Environmental suitability estimated from niche models predicts the
 upper limit of local abundance. The American Naturalist 174:282-291.
- Vander Zanden, M. J. and J. D. Olden. 2008. A management framework for preventing the
 secondary spread of aquatic invasive species. Canadian Journal of Fisheries and Aquatic
- 640 Sciences **65**:1512-1522.

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

644 Vaughan, I. P. and S. J. Ormerod. 2003. Improving the quality of distribution models for

645 conservation by addressing shortcomings in the field collection of training data.

646 Conservation Biology **17**:1601-1611.

- Vaughan, I. P. and S. J. Ormerod. 2005. The continuing challenges of testing species distribution
 models. Journal of Applied Ecology 42:720-730.
- Ward, J. M. and A. Ricciardi. 2007. Impacts of *Dreissena* invasions on benthic macroinvertebrate
 communities: a meta-analysis. Diversity and Distributions 13:155-165.
- Wiley, D. J., R. P. Morgan, R. H. Hilderbrand, R. L. Raesly, and D. L. Shumway. 2004. Relations
 between physical habitat and American eel abundance in five river basins in Maryland.
 Transactions of the American Fisheries Society 133:515-526.
- 654 Wilson, A. E. and O. Sarnelle. 2002. Relationship between zebra mussel biomass and total

655 phosphorus in European and North American lakes. Archiv für Hydrobiologie **153**:339-351.

656 Zambrano, L. and D. Hinojosa. 1999. Direct and indirect effects of carp (Cyprinus carpio L.) on

657 macrophyte and benthic communities in experimental shallow ponds in central Mexico.

658 Hydrobiologia **408**:131-138.

- Zambrano, L., E. Martinez-Meyer, N. Menezes, and A. T. Peterson. 2006. Invasive potential of
 common carp (*Cyprinus carpio*) and Nile tilapia (*Oreochromis niloticus*) in American
- freshwater systems. Canadian Journal of Fisheries and Aquatic Sciences **63**:1903-1910.
- 662
- 663
- 664

Table 1 Summary of the nine environmental variables used for modeling the occurrence and the
abundance and of *Cyprinus carpio*. Descriptive statistics were derived from the full Minnesota

dataset which included the 285 lakes containing carp and 238 lakes where carp are absent.

Variable	Abbreviation	Min	Mean	Median	Max
Climate					
Maximum air temperature (°C)	MAXT	35.56	37.27	37.22	41.11
Minimum air temperature (°C)	MINT	-46.67	-35.14	-35.00	-31.67
Growth degree days	GDD	3536	4369	4496	4890
Annual precipitation (mm)	MP	571.20	741.90	744.00	878.30
Morphometry					
Lake area (ha)	LA	0.86	316.43	164.19	4025.42
Maximum depth (m)	MXD	1.00	10.51	8.20	42.70
Water Chemistry					
Total nitrogen (ppm)	TN	0.38	1.52	1.40	5.07
Total alkalinity (ppm)	ALK	41.00	149.00	145.00	324.00
Carlson's trophic state index	TSI	41.30	62.67	62.59	93.29

676	Table 2 Performance of the carp occurrence model for the 485 lakes on which the model was
677	trained (calibration data), the 57 lakes within Minnesota withheld from model development
678	(internal validation data) and the 38 lakes from eastern South Dakota (external validation data).
679	Criteria include CCI (the percentage of sites where the model correctly predicted either presence
680	or absence), model sensitivity and specificity, AUC (area under the receiver-operating
681	characteristic curve) and corresponding <i>P</i> -value.

Data set	CCI	Sensitivity	Specificity	AUC	Р
Calibration	87.83	86.40	89.53	0.94	> 0.0001
Internal validation	83.81	85.96	81.25	0.90	> 0.0001
External validation	78.95	78.79	80.00	0.75	0.04
582					
583 504					
584 505					
085					
586					
587					
588					
589					
590					
591					
592					
593					
594					
95					

696 Figure legend	ds
-------------------	----

698	Figure 1 Geographic locations and model classifications for the 523 Minnesota lakes and the 38
699	South Dakota lakes used in this study. Circles represent sites where carp are currently established
700	while triangles indicate sites from which carp are absent. Lakes correctly classified by the
701	occurrence model are illustrated in black and incorrectly classifies lakes are shown in grey.
702	
703	Figure 2 Relationship between predicted and observed carp abundance, measured as biomass (kg)
704	per unit effort (BPUE), for A the calibration data, B the internal validation set and C the external
705	validation data. Solid lines represent the fitted regression line between observed and predicted
706	BPUE values and dotted lines represent a one to one fit.
707	
708	Figure 3 Percent relative contribution (mean \pm standard deviation, derived from 100 runs of the
709	optimal networks) of the nine environmental predictor variables for the carp occurrence (P/A) and
710	abundance (BPUE) models, assessed using Garson's algorithm. Abbreviations for each variable
711	can be found in Table 1.











Figure 3

