Minnesota Department of Natural Resources
Investigational Report 562, November 2018

Large-scale Habitat Factors Affecting Fish Populations in Minnesota Lakes and a Proposed Habitat Classification

Timothy K. Cross

Minnesota Department of Natural Resources
Division of Fish and Wildlife
20596 Highway 7
Hutchinson, MN 55350, USA

Abstract. — Habitat has long been considered a keystone for fisheries management in Minnesota lakes, but until recently much of the requisite information on regional and watershed lake habitat factors were not amenable to analysis. In this study, current data describing these large-scale habitat (LSH) factors were compiled and matched with data describing populations of fish in Minnesota lakes surveyed by the Minnesota Department of Natural Resources (MNDNR) Section of Fisheries. A hierarchical decision tree classification procedure was used to identify 11 classes of lakes with relatively homogeneous LSH conditions. This classification accounted for 4 to 41 percent of the variation among lakes in catch per effort (CPE) of various fish species commonly assessed with MNDNR lake surveys. Additional analysis of variation remaining within individual lake classes using machine learning methods (regression tree analysis and random forest) applied to selected fish species revealed more precise fish habitat relationships. Regression tree models for individual lake classes were shown to collectively explain 57 percent of the statewide variation among lakes in Walleye CPE. Random forest models revealed the relative importance and response of fish populations to each habitat factor. In combination, lake classification and machine learning tools are potentially useful to managers faced with acquiring information needed to guide decisions on protection, regulation, restoration, or enhancement of fish habitat in Minnesota lakes at both statewide and local levels.

This project was funded in part by the Federal Aid in Sport Fish Restoration (Dingell-Johnson) Program. Completion Report, Study 606, D-J Project F-26-R Minnesota.

INTRODUCTION

Habitat has long been considered a keystone for fisheries management in Minnesota lakes (Moyle 1945, Moyle 1956, Schupp 1992, and MNDNR 1993). A renewed emphasis has been placed on fish habitat including regional and watershed habitat factors that have changed from historical levels (MNDNR 2013a). Problems in downstream water bodies are often symptomatic of problems upstream in the watershed (Williams et al. 1997). However, information on large-scale habitat factors (factors described by watershed, ecoregion, and lake physical-chemical attributes) have historically been inadequate for identifying specific impacts and appropriate management actions (Rahel and Jackson 2007, Soranno et al. 2009). Large-scale approaches using comparisons of otherwise similar lakes have been useful for predicting effects of these habitat factors (Jackson et al. 2001). Such an approach is especially applicable for Minnesota lakes given the large number of lakes and diversity of habitats and fish communities.

Although each lake has a unique combination of habitat attributes that function to determine the composition of its fish community, comparisons among similar lakes using classification and predictive modeling can provide useful information for directing appropriate efforts aimed at protecting or enhancing habitats for sustaining quality angling. Standardized fish and habitat surveys have been done on Minnesota lakes for over a half century. However, few systematic efforts to identify specific effects of key habitat variables determining the composition of fish communities have occurred since survey procedures were revised in 1993. Blann and Cornett (2008) and Osgood et al. (2002) reviewed existing classifications of Minnesota lakes. Blann and Cornett (2008) concluded that "*hierarchical landscape classifications appear to explain the dominant gradients in species presence and lake chemistry and trophic status. The local-scale macrohabitat classification is also important, but explains less variance in lake attribute response variables and species data*". This observation of the importance of landscape scale relative to site-scale macrohabitat (substrate and plant cover) was documented with bluegill abundance in Minnesota lakes by Cross and McInerny (2005). Minnesota lakes occur across very heterogeneous environments that potentially influence their fish communities. A landscape classification accounts for much of the variation that groups of similar lakes have in common, while the remainder is assumed to represent part of the range of natural variation of local habitat factors associated with a given lake. However, a more lake-specific modeling landscape approach can also be used to address additional variation (Hawkins et al. 2010). Efforts to predict benchmarks for ecological and water-quality assessments have increasingly moved toward site-specific modeling approaches as a way to improve both accuracy and precision of predictions (Hawkins et al. 2010).

Minnesota provides a useful setting for classifying lakes and modeling specific habitat influences due to numerous lakes occurring over diverse large-scale habitat conditions. Furthermore, standardized assessments of fish abundance and lake habitats are available for most lakes with fish communities managed by the MNDNR (1993). This allows for more informative analysis than available with presence-absence data in that differences following abundance gradients likely relate to habitat requirements for fish. Schupp (1992) identified key physico-chemical variables that could be used to classify lakes based on ecological similarities provided a basis for exploring environmental influences on fish communities. Emmons et al. (1999) found that using a hierarchical decision tree methodology to classify lakes by environmental factors was more effective at explaining variation among Wisconsin lakes than a classification scheme modeled after Schupp (1992). More recently Wherely et al. (2012) showed how multivariate regression tree analysis could be used to derive a classification of lakes based on how fish communities in Michigan lakes were affected by environmental factors. Recursive partitioning machine learning algorithms are potentially useful for addressing nonlinear relationships and interactions among environmental factors likely to affect fish communities. Ecologically meaningful decision making requires an understanding of interactions and threshold effects that are rarely modeled (Pittman and Brown 2011). In Minnesota lakes, interactions are likely between fish communities and lake morphology and watershed characteristics. Significant interactions between land use and ecoregions influencing lake phosphorus levels in Minnesota lakes were documented by Cross and Jacobson (2013).

This study focused on differences in long-term (1993-2011) average fish populations among lakes resulting from responses to large-scale habitat variables describing lake morphology, geographic setting, watershed attributes, and other regional characteristics. Study objectives were to describe large-scale habitat (LSH) influences on fish communities affecting Minnesota lakes statewide. Secondly, to derive a spatial hierarchical lake classification to identify groups of lakes with similar LSH influences on fish communities. Finally, to determine some specific influences of habitat on primary fish species assessed in Minnesota lakes. These objectives were addressed using fish catches from standardized surveys analyzed simultaneously with LSH attributes. A decision tree hierarchical classification of lakes was performed using habitat criteria at each step to explain among lake variation in fish catches. This hierarchical classification was then used for identifying more detailed responses of individual species to habitat influences among lakes within each constructed LSH lake class. This process filters through habitat processes operating at various scales down to lake specific physical characteristics significantly affecting fish communities.

METHODS

Fish data.–A dataset of MNDNR standardized lake surveys and population assessments was compiled by querying the statewide MNDNR Fisheries Lake Survey database for catches in experimental gill nets net and ¾ inch double frame trap netting performed for standardized assessments (MNDNR file data). Netting was conducted during the period from June 1 through August 31 from 1993 through 2011 with a resulting mean Julian sampling date of 196 (July 15 for non-leap years). MNDNR standardized trap and gill net sampling is effective for assessing adult populations of larger fish including most sport fish species (MNDNR 1993) and consequentially our analysis focused these species. Only lakes with both gill and trapnet data were used which limited the dataset to 1916 lakes.

Catch-per-effort (CPE) was summarized for each gear type and lake by calculating the mean of all surveys and assessments completed during the study period. Commonly each lake had three to five annual surveys performed during the study period, but they ranged from a single survey to annual surveys depending on management objectives of each lake. CPE calculations averaged over the period were assumed to be representative of equilibrium densities in each lake. The study period was assumed to be recent enough to characterize current habitat conditions, primarily with respect to climate and land use.

Catch-per-effort data was compiled and analyzed using correlation and principal component analysis (PCA). First, descriptive statistics were calculated separately for trapnet and gillnet datasets for all species captured. Mean CPE data was transformed prior to correlation and PCA using (log10+1) to normalize distributions and increase linearity of relationships. Correlation analyses were performed separately on trapnet and gillnet datasets for all commonly captured fish species (usually captured in at least 1% of the lakes). Both ordered and hierarchical correlation analysis procedures contained in the rattle package in R (Williams 2009) were performed to identify fish community patterns. Principal components analysis was performed on a correlation matrix of CPE data on 14 fish species (5 species with gillnet CPE and 9 with trapnet CPE). The 14 species selected were field identifiable, vulnerable to capture by either gill or trap nets, and common enough to be captured in at least 30% of surveyed lakes. In addition we used a combined CPE of all salmonids and coregonids which we coded as COLD to represent cold-water species. PCA was performed in JMP 10 using orthogonal varimax rotation on two factors to reduce the dimensionality of the dataset (SAS 2012).

Large-scale habitat data.–A comprehensive dataset of large-scale variables describing habitat conditions was compiled for Minnesota fishing lakes. Criteria for a lake to be included were that they be naturally formed with a surface area > 4 ha, have a watershed contained entirely within Minnesota, and at least one MNDNR Fisheries standardized survey between 1993 and 2011. Lakes created by impoundment or artificially constructed (excavated) were excluded, but lakes with minor water level control structures were included. Habitat variables selected for the study were restricted to those judged as most readily available, interpretable, and reliably measured. For each lake, maximum depth, surface area, shoreline perimeter, shoreline development index (SDI), and Secchi depth were obtained from MNDNR and MPCA files. Geographic lake center points were used to assign estimates of lake temperature, groundwater recharge, and ecoregion scheme to each lake. Air temperatures used as a surrogate for lake water temperature were the estimated July average daily maximum temperature for the month of July for the period 1981-2010 obtained from the Oregon State University PRISM model (Daly et al. 2008). Also, the USGS base-flow index (BFI) estimated for each lake location was used as a surrogate for ground-water contribution to lakes since direct estimates were not available for most Minnesota lakes. The BFI is the ratio of estimated annual base flow to total flow volume for a given year modeled on data from unregulated rivers and streams (Wolock 2003). Omernik Environmental Protection Agency (EPA) ecoregion classifications (Omernik 2004), were assigned to each lake to characterize the influence of the surrounding template of climate, geomorphology, and land cover/land use.

Lake watersheds were identified using GIS catchment delineations developed for the Midwest Glacial Lake Program by Minnesota DNR staff (Lyn Bergquist MNDNR, personal communication; MNDNR 2012). These catchments were delineated based on height of land using GIS to derive hydrologic corrected digital elevation models and flow networks. Individual lake catchments are hydrologic units nested in a multi-level, hierarchical drainage system. These same hydrologic units fold into the national standards defining the Watershed Boundary Dataset (USGS 2012). Total lake watersheds were identified for each lake by summing all upstream hydrologic units draining into the lake. The total area of the watershed was calculated for each lake and used to calculate the ratio of total watershed area to lake area. The surface area of each land cover type defined in the 2001 National Land Cover Dataset (NLCD; Homer et al. 2004) were extracted for each hydrologic unit polygon. NLCD classes 21, 22, 23, and 24 were used to represent development; 41, 42, 43, and 52 to represent forest; and 82 to represent cropped agricultural land use. Anthropogenically disturbed land use was calculated as the sum of developed, mining, and cultivated agriculture land use. Stream contribution to each lake was computed by summing all stream segments within contributing catchments for each lake from the Strahler stream order layer (MNDNR 2013b). Finally, a lake order was assigned each lake based on a determination of the Strahler stream order (Strahler 1952) of the largest outlet. If a lake was landlocked with no outlet the lake order was assigned to zero. These assignments were based on the DNR Strahler stream layer along with inspection of USGS 24K digital topographical maps and FSA areal imagery for multiple years digitally overlayed on the stream and catchment layers using GIS. An attempt was also made to characterize lakes with outflows and inflows through wetlands or bogs, but this proved too difficult to consistently classify.

Highly correlated habitat variables were identified to guide subsequent modeling efforts for which it was helpful to limit the redundancy of explanatory habitat variables. As with fish community data, both ordered and hierarchical correlation analysis procedures were performed using the rattle package (Williams 2009) to facilitate identification of variables correlated with each other. Also, for both categorical variables (ecoregion and lake order) a series of one-way analysis of variance tests were performed to determine R^2 fit with each continuous habitat variable.

Data analysis.–The 1916 lakes in the dataset were randomly divided into a model development dataset of 1533 lakes and a model validation dataset of 383 lakes. First, a statewide LSH classification of lakes based on fish communities was performed to aggregate lakes with LSH resulting in similar fish communities. A decision tree approach was used combining multivariate modeling of fish communities simultaneously with a series of dichotomous splits using regression tree analysis of habitat factors. This analysis was patterned after previously successful lake classification strategies developed for Wisconsin lakes by Emmons et al. (1999) and Michigan lakes by Wehrly et al. (2012). The first Principal Component (PC) calculated from PCA describing fish communities was used to guide each step of a hierarchical decision tree instead of the environmental variables used by Emmons et al. (1999). In regard to using lake habitat factors to classify lakes based on fish community criteria with a RTA algorithm our classification procedure is similar to the multivariate regression tree approach used by Wehrly et al. (2012). The initial split of the data was based on PC1 scores resulting from a run on the entire model dataset (n=1533).

PC1 is the linear combination of the standardized original variables that has the greatest possible variance (SAS 2012). These PC1 scores were then assigned as the dependent variable to base a split using RTA performed with LSH habitat explanatory variables (Outlet, DEPTH, L-AREA, Region, W-AREA, SDI, JTEM, BFLOW, SECCHI, ALK, FOR, DEVL, AG, and WET). RTA was performed in JMP 10 (SAS 2012) and checked for agreement with the rpart package implemented in R (The R Development Core Team 2013). Variable importance values used to evaluate the contribution of each explanatory variable were generated from the RF procedure implemented in the rattle package (Williams 2009) in R (The R Development Core Team 2013). However, in all cases the habitat variable with the highest RF importance provided the first split of the data with the JMP RT model, which matched rpart model results. For subsequent splits different assemblages of fish usually became prominent. Therefore, PCA was repeated on each set using fish CPE data including only species captured in over half of the lakes to avoid problems of using species with high numbers of zero catches from biasing PCA results (McCune and Grace 2002). The final number of nodes (lake classes) was determined using criteria recommended for RTA analysis by Williams (2011). Ideally, study objectives are best addressed with a model that explains a maximum amount of variability in fish communities with a minimum number of LSH lake classes and maintain class sizes with enough lakes to provide an adequate dataset for species specific CPE models within each class. Consequently we set a minimum decision node size of 50 and an R^2 of 0.12 explaining variation in PC1 (fish component) and RTA (habitat component) models used for each split. In addition five-fold cross validated R^2 values were used to determine if the decision tree split was over-fitting the data (Williams 2011). Finally, because the primary modeling objective was to develop habitat criteria for key fish species managed by the MNDNR we used the CPE of these key fish species in a one-way analysis of variance and used the R^2 statistic to judge the fit and used this as criteria to trim or "prune" the classification. The final decision tree model was evaluated using the validation dataset of 383 lakes excluded from the model dataset. Again we used the CPE of key fish species in a one-way analysis of variance and calculated R^2 to judge the fit of the LSH classification to the verification dataset. In addition we calculated AIC and BIC statistics for comparison to those calculated using a one-way analysis of variance with Ecological Lake Classification (Schupp 1993) instead of LSH class as the independent variable.

LSH effects on Walleye *Sander vitreus* and Largemouth Bass *Micropterus salmoides* CPE were modeled using RF and RTA. These species were selected because much of their habitat requirements are known and because of the high intensity of management efforts directed at these popular sportfish. All species habitat modeling was done using 10 LSH variables (Table 9) selected on the basis of data availability and minimizing redundancy (variables highly correlated with each other). Hence, we used only 2 land cover variables; land cover disturbance (PDIST), which has a strong inverse correlation with forest cover (FOR), and a land cover factor indicative of the influence of the influence of the total area of open water in the watershed (WATER). Maximum depth instead of geometry ratio and total watershed area instead watershed:lake area were used to avoid redundant use of lake area in these ratios. Also, stream length and outlet stream order (lake order) were not used because they were strongly correlated with watershed area. In addition to habitat variables, sampling date was included to reduce spurious habitat relationships that might be influenced by seasonal gear bias. Both RF and RTA models were applied to the entire statewide dataset as well as individual LSH lake classes. Random forest modeling was performed using the RF routine implemented in R (The R Development Core Team 2013) using Rattle (Williams 2009) with model defaults for the number of trees and variables. The overall model fit (R^2) and importance values (percent mean square error and included node purity) of each habitat variable were generated from the model to use in evaluating the contribution of habitat to variation in Walleye and Largemouth Bass CPE. Also, partial dependence plots based on RF models were used to illustrate the relationship between the response and a specific predictor after accounting for effects of other predictors. Regression tree analysis is well suited for analysis of ecological data because it is a robust nonparametric method that handles both categorical and continuously distributed data while providing the ability to reveal interactions in the data (Olden et al. 2008). RTA was implemented with the rpart package contained in the Rattle routine (Williams 2009) using a minimum split size of 20, minimum bucket of 7, and complexity parameter set to 0.03. The overall model fit was assessed using R^2 and a second R^2 generated with five-fold cross-validation implemented using the partition procedure in JMP 10 (SAS 2013). Regression tree analysis provided a quantitative model that accounts for interactions and is easy to understand, however, RF models produce models exhibiting less bias and variance than a single decision tree (RTA model; Williams 2011; Cutler et al. 2007). Finally, individual LSH lake class RT models of Walleye CPE were combined to form a statewide predictive model of Walleye CPE in lakes based on LSH (LSH-RT). Walleye CPE predicted by the LSH-RT model were compared to predictions made using a multiple regression model using the same 10 LSH variables used as input to RTA, ELC lake classes, and LSH lake classes. Suitability of each model was judged based on the amount of variation explained by the models (R2), the distribution of residual values plotted on maps, and the amount of spatial autocorrelation (nonrandom spatial distribution of model residuals) quantified by global Moran's I index values and Getis-Ord Gi* statistics calculated using ArcMap 10 (ESRI 2012). In addition z-scores were calculated for both Moran's index and Getis-Ord Gi* scores to evaluate statistical significance of spatial correlation.

RESULTS

Fish.–A total of 32 fish species and hybrids were captured in at least one percent of the Minnesota lakes sampled with both standardized trap and gill net assessments and are listed in Tables 1 and 2 along with abbreviated species codes and scientific names. Species occurring most frequently in lakes sampled with gillnets were NOP and YEP and with trapnets the most frequently sampled fish species were BLG and BLC. Highest catch rates were found for yellow perch in gill nets and bluegill in trap nets. Fifteen species sampled in over 30% of the lakes were considered for inclusion in subsequent PCA analysis. In addition, because of the paucity of cold water species, we combined Salmonid and Coregonid species as a cold water species group (COLD). Also, SMB were included because of the status as a prominent piscivorous sport species despite occurring in < 30% of lakes. Gill net data was used for indexing LMB, SMB, NOP, WAE, WTS, YEP, and COLD (Table 1). Trap net CPE was used for indexing BLB, YEB, BLC, BLG, BOF, GSF, CAP, PKS, and RKB (Table 2). GOS, BRB, and HSF were not selected for subsequent PCA analysis despite their occurrence in >30% of the surveyed lakes. Correlation matrices and PCA revealed species with similar distributions among lakes (Figures 1, 2; Appendix I). The first two PC's explain over one third of the overall variation among lakes (Table 3). The most obvious assemblage was shown with positive correlations among LMB, BLG, PMK, YEB, NOP, and BOF that resulted in high PC1 scores (Table 3). A second group with highly correlated species abundances, BLB-CAP-BLC, was shown with low PC1 scores, but high PC2 scores (Table 3). Two less distinguishable groups are characterized by low PC2 scores. Those with a tendency for low PC1 scores WAE-WTS-SMB and those with higher PC1 scores COLD-RKB. Additional correlation matrices calculated using all species captured in at least one percent of Minnesota lakes showed less ubiquitous species often correlate with those used in PCA which provides an indication of the robustness of the fish community gradients defined by PCA.

Habitat.–Large-scale habitats are highly variable among Minnesota fishing lakes. Lakes are located across three major Ecoregions (EPA level I) and run the gamut from soft to hard water and from oligotrophic to hyper-eutrophic. Lake area ranges over 3 orders of magnitude and differs in maximum depth by more than 64 m and in Secchi transparency more than 15 m (Table 4). Watershed area ranges over 4 orders of magnitude and land cover ranges from highly forested (97%) to highly agricultural (95%) to highly developed (92%; Table 4).

TABLE 1. Statistical summary of average lake gill net catch-per-effort (number/net) 1993-2010 for Minnesota fishing lakes (N=1916).

Code	Common name	Scientific name	Occurrence (%)	Median	Mean	SD
NOP	Northern Pike	Esox lucius	93.37	5.871	4.888	1.183
YEP	Yellow Perch	Perca flavescens	89.04	5.950	5.982	2.744
BLC	Black Crappie	Pomoxis nigromaculatus	80.22	1.113	1.742	1.716
BLG	Bluegill	Lepomis macrochirus	76.88	1.661	2.148	1.882
WTS	White Sucker	Catostomus commersoni	76.88	0.750	1.158	1.228
WAE	Walleye	Sander vitreus	75.99	1.618	1.838	1.501
LMB	Largemouth Bass	Micropterus salmoides	64.14	0.208	0.429	0.557
BLB	Black Bullhead	Amieurus natalis	60.75	0.197	2.184	3.918
PMK	Pumpkinseed Sunfish	Lepomis gibbosus	58.72	0.167	0.574	0.845
YEB	Yellow Bullhead	Amieurus melas	54.18	0.111	0.954	1.591
BRB	Brown Bullhead	Amieurus nebulosus	47.44	0	0.374	0.811
RKB	Rock Bass	Ambloplites rupestris	32.36	0	0.294	0.716
HSF	Hybrid Sunfish	Lepomis hybrid	31.94	0	0.135	0.380
CAP	Common Carp	Cyprinus carpio	23.75	0	0.250	0.897
COLD	Cold-water species	Cold-water species COLD	23.20	0	0.135	1.047
TLC	Tullibee (Cisco)	Coregonus artedi	21.09	0	0.274	0.832
GOS	Golden Shiner	Notemigonus crysoleucas	20.62	0	0.132	0.500
SMB	Smallmouth Bass	Micropterus dolomieu	12.63	0	0.064	0.268
GSF	Green Sunfish	Lepomis cyanellus	8.72	0	0.019	0.159
RHS	Redhorse spp.	Moxostoma spp.	8.72	0	0.033	0.182
WHC	White Crappie	Pomosix annularis	6.84	0	0.054	0.326
CCF	Channel Catfish	Ictalurus punctatus	6.00	0	0.042	0.255
BIB	Bigmouth Buffalo	Ictiobus cyprinellus	5.58	0	0.049	0.068
BOF	Bowfin	Amia calva	5.58	0	0.050	0.345
TMUE	Muskellunge hybrid	Esox hybrid	4.91	0	0.012	0.093
FRD	Freshwater Drum	Aplodinotus grunniens	4.02	0	0.069	0.502
BUB	Burbot	Lota lota	3.86	0	0.002	0.028
MUE	Muskellunge	Exox masquinongy	3.44	0	0.009	0.085
LKW	Lake Whitefish	Coregonus clueaformis	2.61	0	0.014	0.143
RBT	Rainbow Trout	Oncorhynchus mykiss	1.57	0	0.023	0.241
WHB	White Bass	White Bass WHB	1.46	0	0.009	0.121
OSS	Orange-spotted Sunfish	Orange-spotted Sunfish	1.41	0	0.007	0.076
BNT	Brown Trout	Salmo trutta	0.68	0	0.002	0.054
BKT	Brook Trout	Salvelinus fontinalis	0.57	0	0.006	0.034
LAT	Lake Trout	Salvelinus namaycush	0.26	0	0.001	0.049

TABLE 2. Statistical summary of average lake trap net catch-per-effort (number/net) 1993-2010 for Minnesota fishing lakes (N=1916).

Code	Common name	Scientific name	Occurrence (%)	Median	Mean	SD
NOP	Northern Pike	Esox lucius	90.66	0.556	0.603	0.396
BLG	Bluegill	Lepomis macrochirus	88.83	15.811	11.365	3.131
YEP	Yellow Perch	Perca flavescens	86.12	0.556	0.923	1.046
BLC	Black Crappie	Pomoxis nigromaculatus	86.06	1.317	1.917	1.573
PMK	Pumpkinseed Sunfish	Lepomis gibbosus	80.85	1.649	0.104	1.240
LMB	Largemouth Bass	Micropterus salmoides	74.22	0.222	0.387	0.469
WAE	Walleye	Sander vitreus	66.18	0.137	0.300	0.448
HSF	Hybrid Sunfish	Lepomis hybrid	62.73	0.111	0.626	1.038
YEB	Yellow Bullhead	Amieurus melas	62.27	0.062	0.977	1.236
BLB	Black Bullhead	Amieurus natalis	56.99	0.444	1.244	3.047
WTS	White Sucker	Catostomus commersoni	56.26	0.042	0.274	0.567
BRB	Brown Bullhead	Amieurus nebulosus	52.92	0.037	0.327	0.696
BOF	Bowfin	Amia calva	38.99	0	0.225	0.382
RKB	Rock Bass	Ambloplites rupestris	38.15	0	0.324	0.656
GSF	Green Sunfish	Lepomis cyanellus	31.68	0	0.125	0.428
CAP	Common Carp	Cyprinus carpio	30.38	0	0.247	0.730
GOS	Golden Shiner	Notemigonus crysoleucas	28.76	0	0.094	0.352
SMB	Smallmouth Bass	Micropterus dolomieu	8.87	0	0.025	0.148
RHS	Redhorse spp.	Moxostoma spp.	8.14	0	0.040	0.213
WHC	White Crappie	Pomosix annularis	7.93	0	0.064	0.400
BIB	Bigmouth Buffalo	Ictiobus cyprinellus	7.67	0	0.288	0.173
CCF	Channel Catfish	Ictalurus punctatus	4.91	0	0.014	0.094
FRD	Freshwater Drum	Aplodinotus grunniens	4.12	0	0.040	0.277
OSS	Orange-spotted Sunfish	Orange-spotted Sunfish	3.86	0	0.023	0.236
TMUE	Muskellunge hybrid	Esox hybrid	3.44	0	0.004	0.028
BUB	Burbot	Lota lota	1.20	0	0.001	0.006
RBT	Rainbow Trout	Oncorhynchus mykiss	1.15	0	0.005	0.065
WHB	White Bass	White Bass WHB	1.15	0	0.004	0.051
BKT	Brook Trout	Salvelinus fontinalis	0.37	0	0.002	0.065

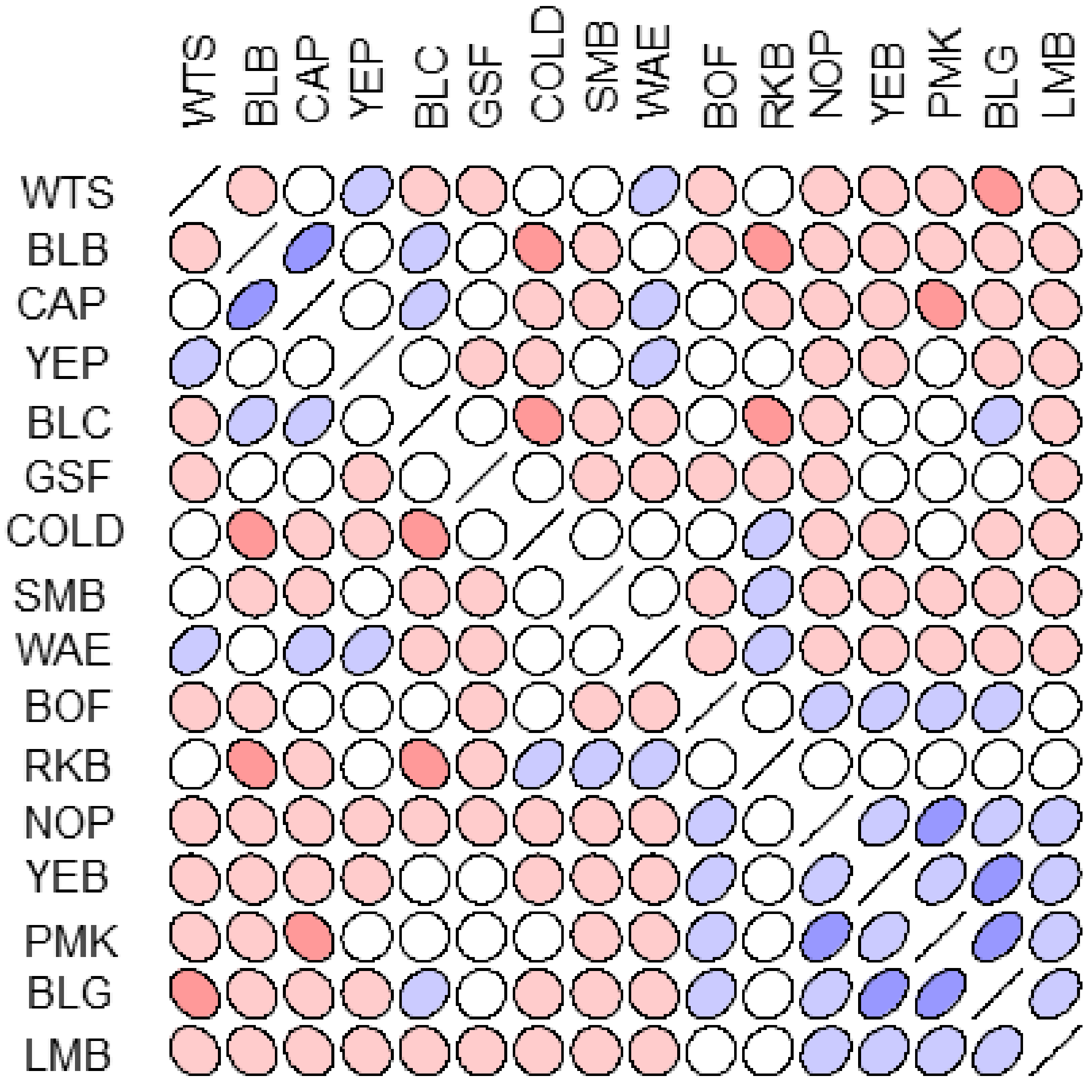


FIGURE 1. A matrix of Pearson correlation coefficients for primary species sampled by standardized gill and trap net assessments in 1916 Minnesota lakes. Negative values depicted in red and positive values depicted in blue. Stronger correlations are indicated more color intensity and symbol elongation. Actual correlation coefficients are listed in Appendix I.

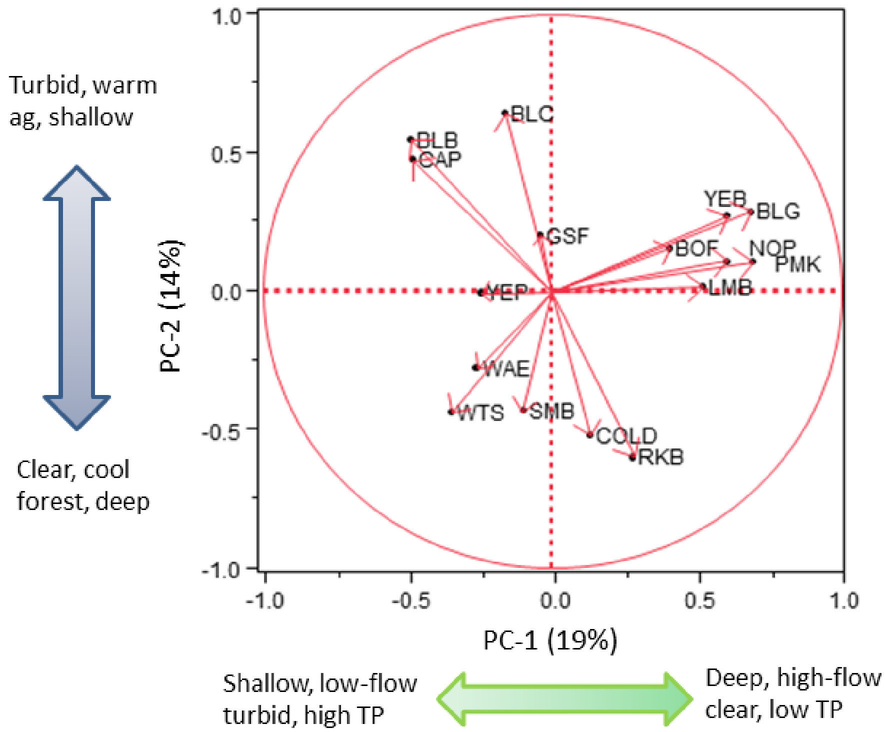


FIGURE 2. The first two principal components calculated on fish CPE dataset with corresponding LSH variables mostly highly correlated with each PC.

TABLE 3. Pearson correlations between CPE of fish species and principal components 1 and 2.

Species	PC1	PC2
LMB	**0.52**	0.01
NOP	**0.60**	0.11
WAE	-0.26	-0.28
WTS	-0.34	**-0.44**
YEP	-0.25	-0.01
SMB	-0.10	**-0.44**
COLD	0.13	**-0.52**
BLB	**-0.49**	**0.55**
BLC	-0.16	**0.64**
BLG	**0.69**	0.29
BOF	**0.41**	0.15
CAP	**-0.48**	**0.47**
GSF	-0.04	0.20
RKB	0.28	**-0.60**
YEB	**0.60**	0.27
PMK	**0.69**	0.11
Percent of variance explained	18.6	14.2

TABLE 4. A statistical summary of large-scale habitat variables characterizing Minnesota fishing lakes. Temperature is the average daily maximum air temperature for July (1981-2010) from the PRISM model (Daly et al. 2008).

Variable	Abbreviation	N	Minimum	Maximum	Median	Mean	SD
Lake							
Max. depth (ft)	DEPTH	1856	4	213	31	37.2	25.21
Area (acres)	L-AREA	1916	10	39,272	214	500.3	1266.73
Geometry Ratio	-	1160	0.47	34.63	3.32	4.82	4.249
Perimeter (km)	PERIM	1916	0.87	549.6	5.6	8.60	15.505
Shoreline development	SDI	1916	1.02	12.98	1.68	1.87	0.762
Temperature (C)	JTEMP	1905	22.24	28.78	26.50	26.62	1.158
Alkalinity (mg/l $CaCO_3$)	ALK	1688	1.2	786	111.4	107.9	69.1
Total phosphorus (ppb)	TP	1477	4	722	27	49.5	64.52
Secchi (m)	SECCHI	1903	0.14	15.29	2.40	2.60	1.492
Watershed							
Area (acres)	W-AREA	1916	40	950,764	3084	19,697	68,110.8
Watershed:Lake area	W:L	1916	1.47	2356.73	12.47	56.58	169.478
Baseflow %	BFLOW	1916	31	76	55	55.9	7.00
Stream length (km)	STRM	1916	0	3067.03	5.87	63.81	228.83
Water %	WATER	1916	0	66.3	13.6	15.3	10.25
Developed %	DEVL	1916	0	92.3	3.7	7.2	14.41
Bare %	-	1916	0	24.5	0	0.1	0.95
Forest %	FOR	1916	0	97.0	45.7	44.0	28.80
Grass %	GRASS	1916	0	30.4	1.6	2.4	3.14
Agriculture %	AG	1916	0	94.8	10.1	23.7	26.36

Large-scale lake habitat variables are frequently inter-correlated (Table 5; Figure 3). Secchi transparency is strongly correlated with TP. Lakes tend to have higher TP and alkalinity in areas with warmer mean July maximum temperatures and with higher percentages of agriculture and development land cover. Conversely, lakes tend to have lower TP when they are deeper and have more forested watersheds with greater groundwater input (baseflow). The physical dimensions of lakes and watersheds are also frequently correlated. Lake area is strongly correlated with lake perimeter, SDI, and watershed area which was strongly correlated with stream length and watershed to lake area ratio. Lake order defined by outlet stream order corresponds closely to length of tributary streams and watershed area. Level IV EPA ecoregions are strongly associated with watershed land cover and ground water inputs explaining their strong association with lake trophic status variables Secchi transparency and TP.

Fish community gradients shown with PC1 and PC2 (Figure 2) were strongly correlated lake trophic and climatic variables (Secchi, TP, July temperatures) in addition to variation in watershed landcover and groundwater input (Table 5). PC1 which differentiated lakes with higher abundances of sunfish species, LMB, NOP, and YEB from lakes with higher abundances of BLB and CAP was strongly correlated with trophic status (TP and Secchi), depth, and groundwater contribution. Lakes that favored high abundances of the sunfish group tend to be deeper with lower fertility and more groundwater input (Figure 2). With variation attributed to PC1 removed, PC2 differentiated between lakes with higher abundances of BLC from those characterized by higher abundances of RKB and COLD water species. PC2 was most correlated with temperature and land cover in addition to trophic status. Lakes that favored higher abundances of COLD and RKB were found in less fertile lakes exposed to cooler July temperatures and more forest cover (Figure 2).

TABLE 5. Pearson correlation coefficients estimated by pairwise method for dataset of 1916 Minnesota fishing lakes.

Variable	PC2	JTEMP	DEVL	AG	ALK	TP	GRASS	PC1	WET	WATER	BFLOW	SECCHI	W:L	DEPTH	FOR	W-SHED	L-AREA	PERIM	SDI
PC2	1.00	0.64	0.46	0.49	0.25	0.52	0.26	0.00	0.07	-0.07	-0.21	-0.48	-0.04	-0.31	-0.58	-0.17	-0.20	-0.24	-0.21
JTEMP	0.64	1.00	0.67	0.69	0.57	0.51	0.31	0.03	-0.02	-0.08	-0.10	-0.38	0.03	-0.03	-0.79	0.05	0.04	-0.05	-0.18
DEVL	0.46	0.67	1.00	0.28	0.39	0.30	0.12	0.07	-0.10	-0.03	-0.10	-0.23	0.02	0.01	-0.57	-0.05	-0.10	-0.15	-0.16
AG	0.49	0.69	0.28	1.00	0.59	0.48	0.49	-0.03	-0.02	-0.20	-0.09	-0.35	0.08	-0.08	-0.78	0.17	0.16	0.05	-0.15
ALK	0.25	0.57	0.39	0.59	1.00	0.26	0.35	0.08	-0.07	-0.09	0.13	-0.10	0.16	0.15	-0.53	0.29	0.25	0.13	-0.13
TP	0.52	0.51	0.30	0.48	0.26	1.00	0.07	-0.56	-0.06	-0.14	-0.45	-0.78	0.12	-0.55	-0.58	0.08	-0.04	-0.08	-0.08
GRASS	0.26	0.31	0.12	0.49	0.35	0.07	1.00	0.18	0.14	-0.13	0.02	-0.05	0.14	0.07	-0.30	0.17	0.09	0.04	-0.07
PC1	0.00	0.03	0.07	-0.03	0.08	-0.56	0.18	1.00	0.20	0.02	0.53	0.57	-0.02	0.50	0.18	0.03	0.07	0.03	-0.06
WET	0.07	-0.02	-0.10	-0.02	-0.07	-0.06	0.14	0.20	1.00	-0.11	-0.03	-0.06	0.16	0.02	0.02	0.24	0.17	0.11	-0.02
WATER	-0.07	-0.08	-0.03	-0.20	-0.09	-0.14	-0.13	0.02	-0.11	1.00	-0.10	0.13	-0.52	0.03	-0.05	-0.28	0.19	0.14	-0.01
BFLOW	-0.21	-0.10	-0.10	-0.09	0.13	-0.45	0.02	0.53	-0.03	-0.10	1.00	0.46	0.12	0.38	0.34	0.16	0.10	0.08	0.01
SECCHI	-0.48	-0.38	-0.23	-0.35	-0.10	-0.78	-0.05	0.57	-0.06	0.13	0.46	1.00	-0.12	0.65	0.47	-0.09	0.01	0.03	0.05
W:L	-0.04	0.03	0.02	0.08	0.16	0.12	0.14	-0.02	0.16	-0.52	0.12	-0.12	1.00	-0.02	0.13	0.73	-0.07	-0.02	0.08
DEPTH	-0.31	-0.03	0.01	-0.08	0.15	-0.55	0.07	0.50	0.02	0.03	0.38	0.65	-0.02	1.00	0.19	0.12	0.21	0.20	0.09
FOR	-0.58	-0.79	-0.57	-0.78	-0.53	-0.58	-0.30	0.18	0.02	-0.05	0.34	0.47	0.13	0.19	1.00	0.03	-0.10	0.01	0.19
W-AREA	-0.17	0.05	-0.05	0.17	0.29	0.08	0.17	0.03	0.24	-0.28	0.16	-0.09	0.73	0.12	0.03	1.00	0.63	0.61	0.32
L-AREA	-0.20	0.04	-0.10	0.16	0.25	-0.04	0.09	0.07	0.17	0.19	0.10	0.01	-0.07	0.21	-0.10	0.63	1.00	0.92	0.38
PERIM	-0.24	-0.05	-0.15	0.05	0.13	-0.08	0.04	0.03	0.11	0.14	0.08	0.03	-0.02	0.20	0.01	0.61	0.92	1.00	0.69
SDI	-0.21	-0.18	-0.16	-0.15	-0.13	-0.08	-0.07	-0.06	-0.02	-0.01	0.01	0.05	0.08	0.09	0.19	0.32	0.38	0.69	1.00
Outlet*	0.00	0.02	0.01	0.02	0.00	0.04	0.02	0.01	0.06	0.09	0.01	0.05	0.42	0.00	0.01	0.58	0.15	0.15	0.06
Region*	0.54	0.91	0.59	0.72	0.57	0.45	0.40	0.45	0.23	0.12	0.67	0.41	0.04	0.27	0.77	0.06	0.10	0.07	0.07

*Coefficients for these categorical variables are R^2 values from least squares one-way analysis of variance.

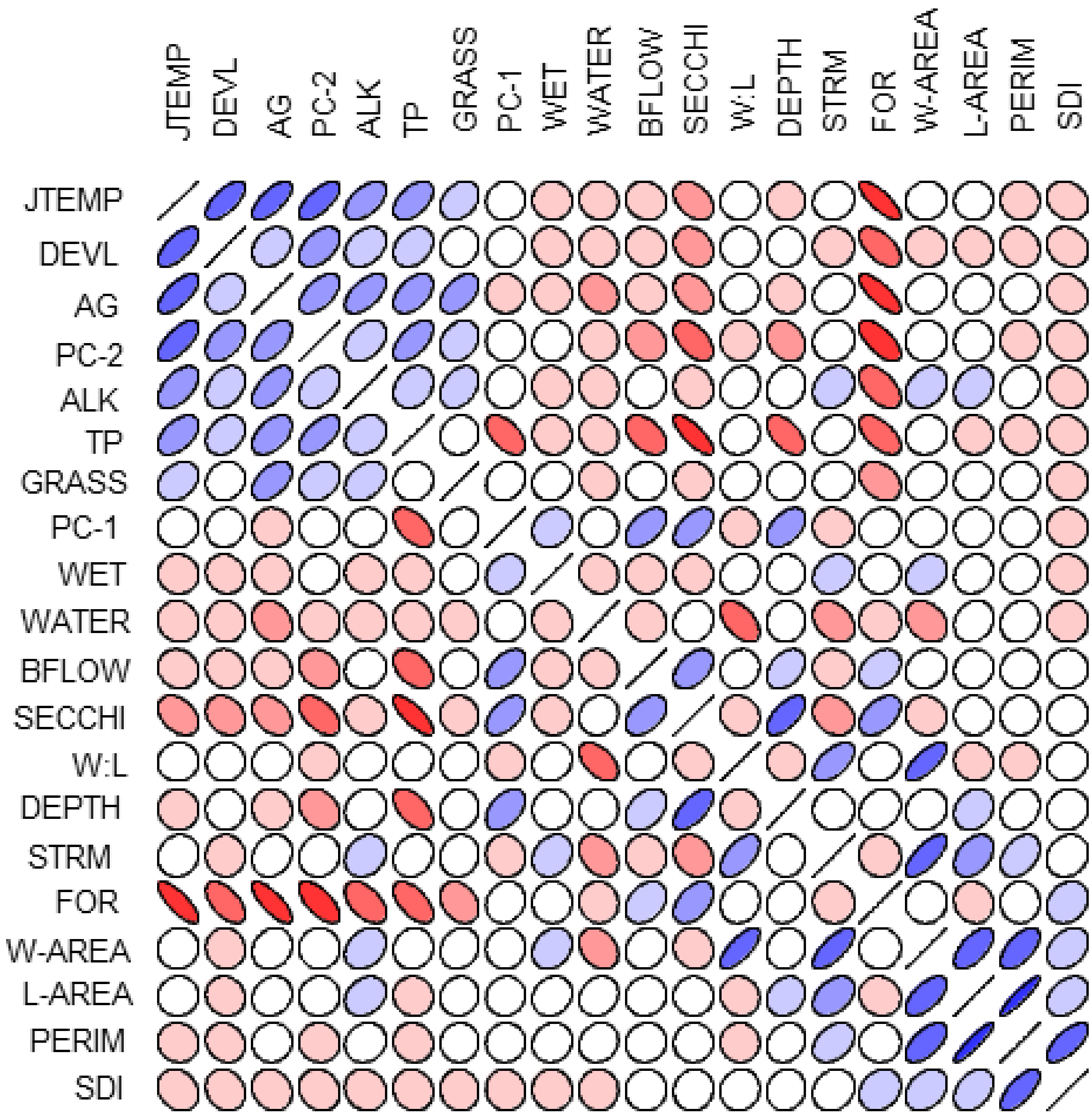


FIGURE 3. A matrix of Pearson correlation coefficients among large-scale habitat variables using 1916 Minnesota lakes. Negative values depicted in red and positive values depicted in blue. Stronger correlations are indicated with more color intensity and symbol elongation.

Classification.–The first PC calculated on each dataset formed with binary splits used for constructing a lake classification decision tree had R^2 values ranging from 16.0 to 28.7 (Table 6). A total of 16 species (including COLD) were used for PCA at each binary split, of which seven (NOP, WAE, WTS, YEP, BLG, BLC, and PMK) were ubiquitous and used at every split. Among the 10 splits, five different LSH variables were selected with RTA for splitting criteria. Ecoregion criteria were involved with three of the initial splits, lake area for three splits, latitude for two splits, and temperature and Secchi for a single split. Additional exploratory analysis revealed that temperature could be substituted for the two latitude splits while explaining only slightly less variation with a similar split of lakes.

Variation in fish communities among lakes defined by PC1 at each split explained by LSH variables ranged from 13.1% for split 2A1 to 76.8% for split 2A (Table 7). Split 2A1 also had the poorest cross-validation R^2, but cross-validation R^2 values at other splits matched closely to the model R^2 values. The NLF ecoregion contains the highest density of surveyed lakes in our dataset and this was reflected by inclusion of six of ten LSH lake classes. Geographic centers of class H and F lakes (split from the other NLF lakes based on lower Secchi measurements) are located more southerly than other LSH classes in the ecoregion (Figure 4). Classes M (clear lakes > 450 a) and K (clear lakes 63 to 450 a) had geographic centers in close proximity near the northernmost edge of the NLF ecoregion. In contrast, similar lakes classified with less water clarity (N and L) had geographic centers near the center of the NLF ecoregion (Figure 4) indicating a latitudinal gradient in water clarity.

TABLE 6. Pearson correlation coefficients of fish species catch-per-effort (CPE) to principal component one (PC1) used as the dependent value subjected to regression tree analysis at each split of the data. The number of fish species used at each split is totaled at the bottom along with the number of lakes (N) and the amount of fish CPE variation explained by PC1.

	Decision Tree Split									
Species	1	2A	2A1	2A11	2B	2B1	2B2	2B21	2B212	2B22
Gillnet										
LMB	0.32				0.28	0.20	-0.07	0.20	0.01	0.63
NOP	0.35	-0.13	0.48	-0.09	0.28	0.58	-0.11	0.55	0.51	0.45
SMB			-0.08	0.49						-0.23
WAE	-0.12	0.21	-0.14	0.07	0.14	0.31	0.77	0.34	-0.58	-0.53
WTS	-0.18	-0.30	-0.17	-0.39	0.04	0.17	0.76	0.17	-0.65	-0.66
YEP	-0.11	0.28	0.26	-0.4	0.01	0.55	0.63	0.21	-0.49	-0.65
COLD	0.09		-0.34	0.3	0.25		0.46			-0.38
Trapnet										
BLB	-0.30	0.52			-0.38	-0.12	-0.08	-0.17		0.06
BLG	0.40	0.47	0.35	0.78	-0.33	0.49	-0.08	0.67	0.39	0.67
BLC	-0.10	0.17	0.45	0.55	0.24	0.14	-0.32	-0.01	0.08	0.32
BOF	0.23				0.12	0.55	0.10	0.43	0.34	0.13
CAP	-0.29	0.49			-0.33	0.01				0.20
GSF	-0.02					-0.31	0.02	0.12	0.19	0.14
PMK	0.40	-0.09	0.45	-0.06	0.35	0.59	0.10	0.63	0.55	0.25
RKB	0.18		0.07	0.75	0.36		0.70	0.34	0.46	-0.36
YEB	0.34				0.26	0.69	-0.10	0.61	0.44	0.55
Number	15	9	10	10	14	13	14	13	12	16
Variation %	19.7	28.7	22.3	21.4	18.3	17.7	17.3	16.0	17.5	19.2
N	1533	313	214	157	1220	388	832	608	512	224

TABLE 7. Decision tree splitting criteria and fit to principal component 1 (PC1) at each hierarchical split of 1533 Minnesota fishing lakes using large-scale habitat variables. The number of lakes at each split, mean PC1 value, percent of variation accounted for with the split (R^2), and an R^2 calculated with 5-fold cross validation of the model. Split identification code can be referenced to Table 1 and Figure 4.

Split	Criteria	N	Mean PC1	R^2	X-fold R^2
1A	Ecoregion (46k,48a,47b,46e,47c,47g,50t,50n,48d,50p,49b)	313	-1.967	0.336	0.334
	(50a,51h,49a,51i,51a,50o,50s,50m,50r,51k,51j,50b,51l,50q)	1220	0.505		
2A	Ecoregion (50t,50n,50p,49b,48d)	214	-0.956	0.768	0.763
	(47g,46e,48a,46k,47b,47c)	99	2.066		
2A1	Ecoregion (50n,50t)	157	-0.334	0.138	0.0877
	Level4(50p,48d,49b)	57	0.919		
2A11	Average July maximum daily temperature <25.08 C	92	-0.710	0.336	0.324
	>=25.08 C	65	1.005		
2B	Secchi <1.93 m	388	-1.373	0.343	0.340
	>=1.93 m	832	0.640		
2B1	Lake area <150.36 acres	148	-0.841	0.190	0.181
	>=150 acres	240	0.519		
2B2	Lake area <450.29 acres	608	-0.497	0.277	0.272
	>=450.29 acres	224	1.349		
2B21	Lake area <63.22 acres	96	-1.352	0.165	0.157
	>=63.22 acres	512	0.254		
2B212	UTM-northing>=5183025	239	-0.638	0.170	0.163
	<5183025	273	0.559		
2B22	UTM-northing>=5199308	93	-1.182	0.324	0.320
	<5199308	131	0.840		

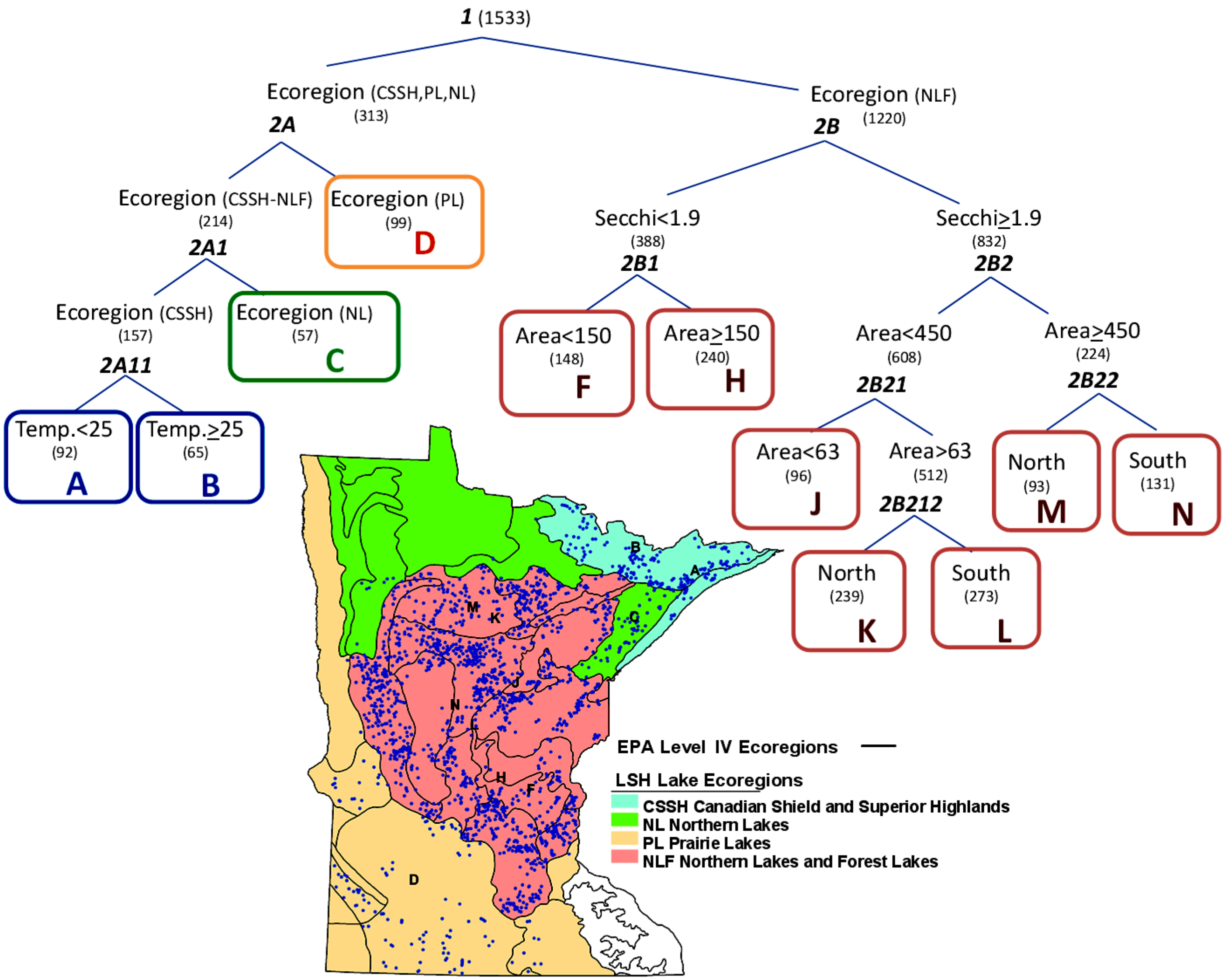


FIGURE 4. Decision tree of Minnesota lakes classified by LSH attributes. Lake classes shown as terminal nodes in rectangles color coded to match ecoregion membership.

Aside from ecoregional and geographical differences, LSH lake classes differed widely in other habitat attributes. Six key habitat variables including 3 LSH variables used as splitting criteria (lake area, Secchi, and temperature) characterize these habitat differences (Figure 5). Compared to other lake classes A, B, and C had lower alkalinities and temperatures and class D lakes which had higher alkalinities, temperatures, and TP levels. Considerable differences in depth also occur among lake classes with Class D lakes being very shallow. Differences in Secchi transparency mirror TP for lake classes including the large difference between NLF classes F, H and other lake classes split at 2B. Also, large differences in NLF ecoregion lakes are apparent from splits on lake area differentiating small and larger lakes with low Secchi transparency and small, mid, and large size lakes with high Secchi transparency. In the NLF ecoregion, lakes with low Secchi transparency (F, H) tend to be shallower than lakes in other lake classes mirroring differences in watershed land cover (Figures 5 and 6).

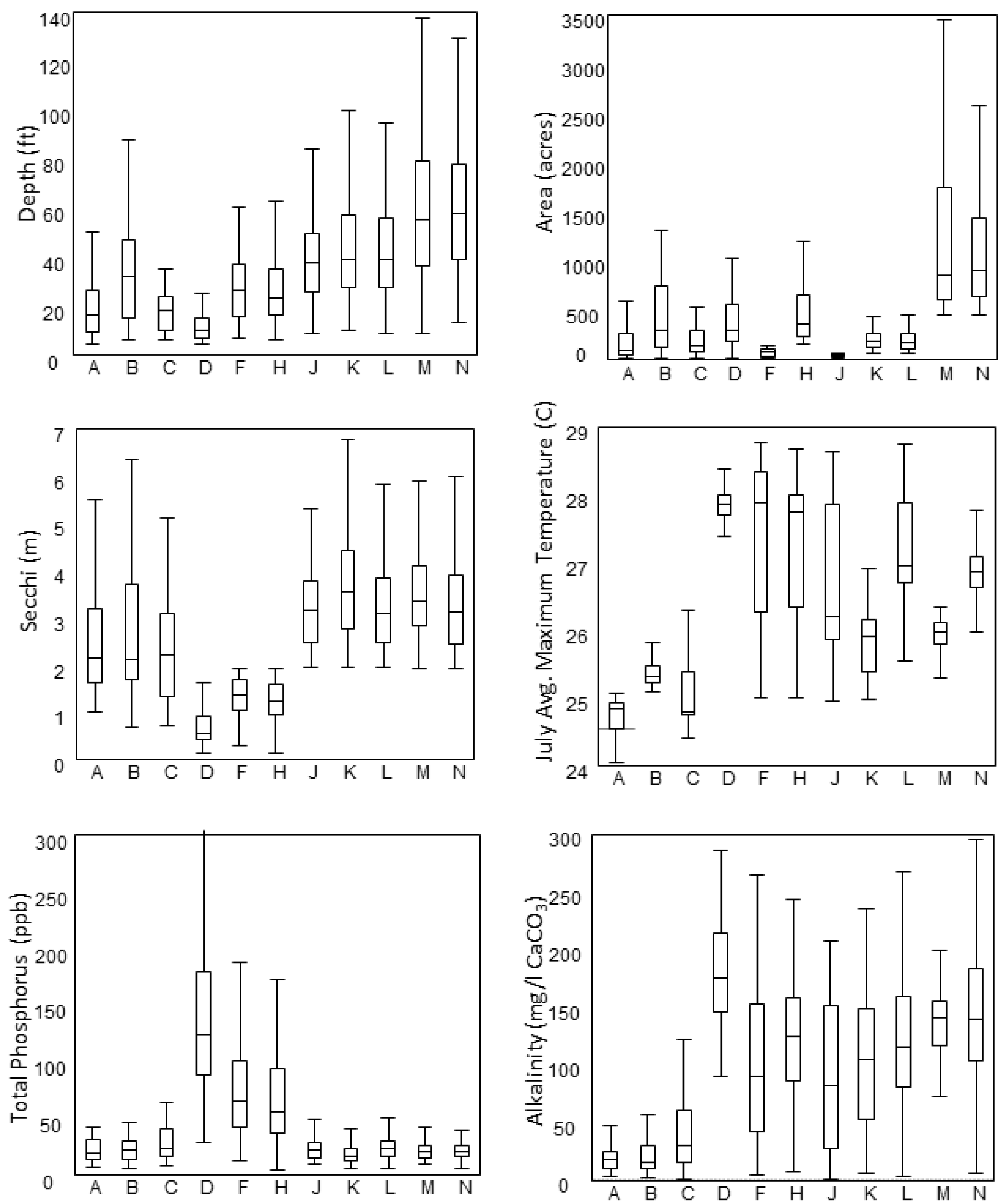


FIGURE 5. Box plots of the distribution of selected large-scale habitat variables in lake classes. The line within the box represents the median, the ends of the box are first and third quartile, and the whiskers represent a computed outlier range.

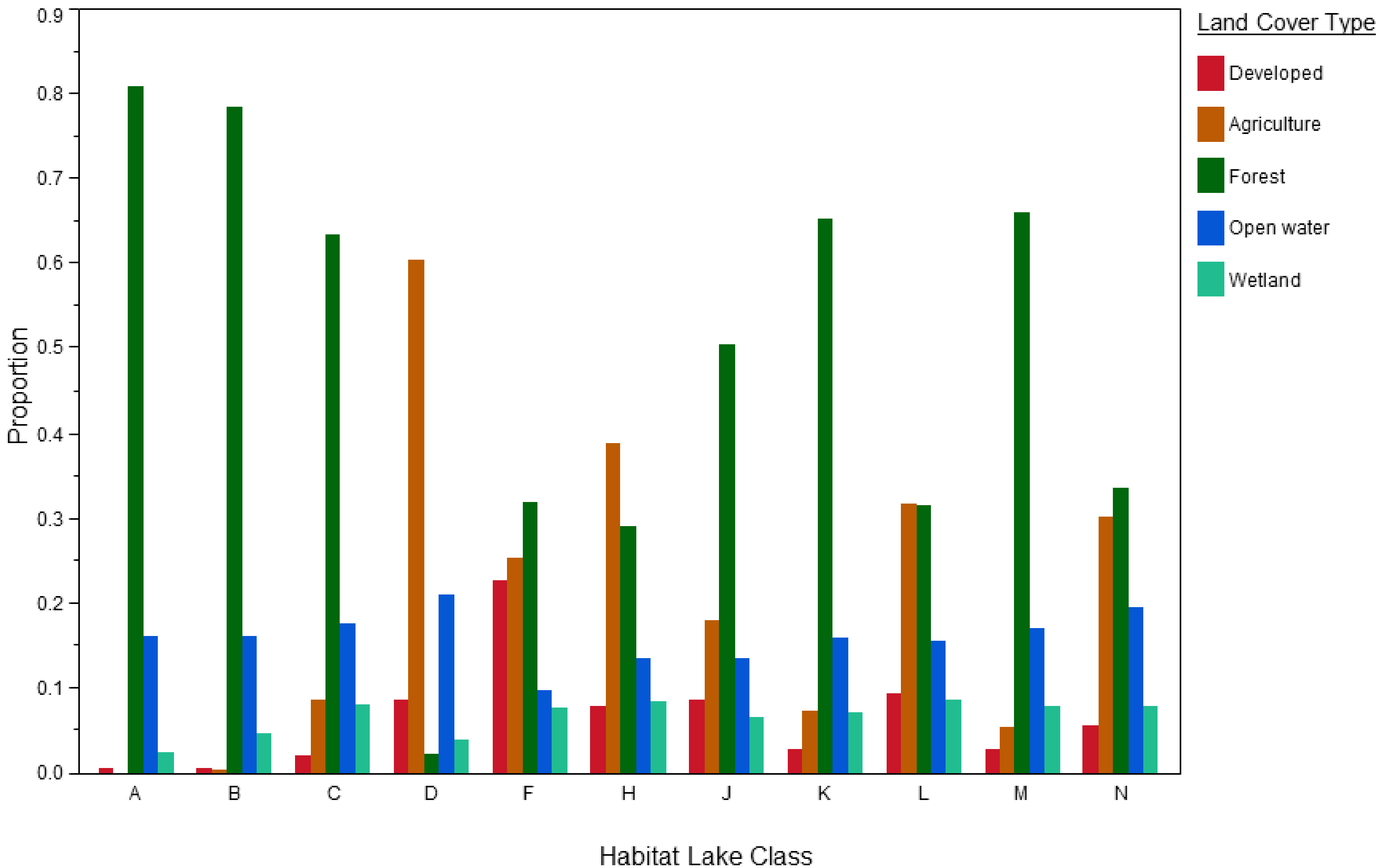


FIGURE 6. Average proportion of lake watersheds in five major land cover categories by lake habitat class.

Lake classes with lower Secchi have more agriculture and less forest cover. Interestingly, differences between the northern NLF classes (M,K) and the southern NLF classes (N,L) correspond to differences in Secchi that also correspond to land cover differences (Figure 6). In fact, both land cover (ratio of AG to FOR) and Secchi follow a latitudinal gradient that is also evident with Secchi except that the CSSH lakes have slightly lower Secchi values than northern NLF lakes.

Large-scale habitat lake classes explained a large amount of variation in the CPE of individual fish species among lakes. Depending on the species, one-way analysis of variance using LSH lake classes as the independent variable explained 4 to 41% of the variation in CPE (Table 8). Additional subdivision of the LSH class into 14 classes (H14) improved R^2 values slightly, but the added complexity of additional classes was not justified based on interpretability and reduction in BIC (Table 8). By comparison, the widely used Minnesota Ecological Lake Classification (ELC, Schupp 1992) provided a poorer fit for most species (Table 8). However, ELC provided a better fit for cool and cold-water species (WAE, RKB, SMB, and TLC). Differences in fish abundances corresponding to LSH classification appear consistent with existing knowledge of habitat requirements. For example, WAE were shown to be more abundant in LSH classes characterized by larger lake area and higher fertility and BLG most abundant in smaller and less eutrophic classes located at lower latitudes (Figure 7).

TABLE 8. Fit of one way analysis of variance models for estimating gill and trap net CPE of selected fish species using Ecological Lake Class (ELC), Large-Scale Habitat Lake Class with 14 classes (H14), and Large-Scale Habitat Lake Class with 11 classes (LSH). Model fit estimated with R^2, AIC, and BIC criteria and R^2 is totaled at the bottom of each column to represent combined fit for all species.

	Model Lakes (N=1533)									Verification Lakes (N=380)					
	R^2			AIC			BIC			R^2		AIC		BIC	
Species	ELC	H14	LSH	ELC	H14	LSH	ELC	H14	LSH	ELC	LSH	ELC	LSH	ELC	LSH
LMB	0.14	0.22	0.22	-1002	-1172	-1172	-770	-1092	-1108	0.05	0.14	-55	-127	93	-81
NOP	0.16	0.23	0.20	822	672	723	1054	752	787	0.13	0.17	232	180	380	227
SMB	0.14	0.09	0.09	-2805	-2744	-2750	-2573	-2664	-2686	0.10	0.16	-632	-690	-484	-643
TLC	0.30	0.16	0.16	-206	49	44	26	129	107	0.25	0.12	-41	-18	107	29
WAE	0.35	0.29	0.28	919	1029	1032	1151	1109	1096	0.31	0.29	289	264	437	310
WTS	0.24	0.26	0.25	744	670	692	976	750	756	0.26	0.24	214	186	362	232
YEP	0.13	0.20	0.19	2454	2293	2314	2687	2372	2378	0.10	0.18	692	624	840	670
BLB	0.45	0.44	0.41	1928	1925	2001	2160	2005	2065	0.47	0.39	537	553	685	600
BLC	0.24	0.27	0.27	1259	1159	1154	1491	1239	1218	0.24	0.26	362	551	510	597
BLG	0.38	0.39	0.38	2255	2158	2165	2487	2238	2229	0.34	0.32	573	551	721	597
BOF	0.13	0.17	0.16	-1784	-1881	-1875	-1552	-1802	-1811	0.05	0.13	-441	-508	-293	-462
CAP	0.36	0.45	0.36	-696	-957	-731	-464	-877	-667	0.36	0.34	-132	-153	16	-107
GSF	0.03	0.04	0.04	-1257	-1307	-1303	-1024	-1227	-1239	0.08	0.03	-464	-476	-316	-429
PMK	0.20	0.25	0.24	852	709	725	1084	788	789	0.24	0.26	219	176	367	222
RKB	0.35	0.28	0.28	-882	-756	-761	-650	-676	-698	0.40	0.34	-252	-255	-104	-208
YEB	0.20	0.30	0.29	883	645	672	1115	725	736	0.11	0.24	230	136	378	183
Total	3.80	4.04	3.82							3.49	3.61				

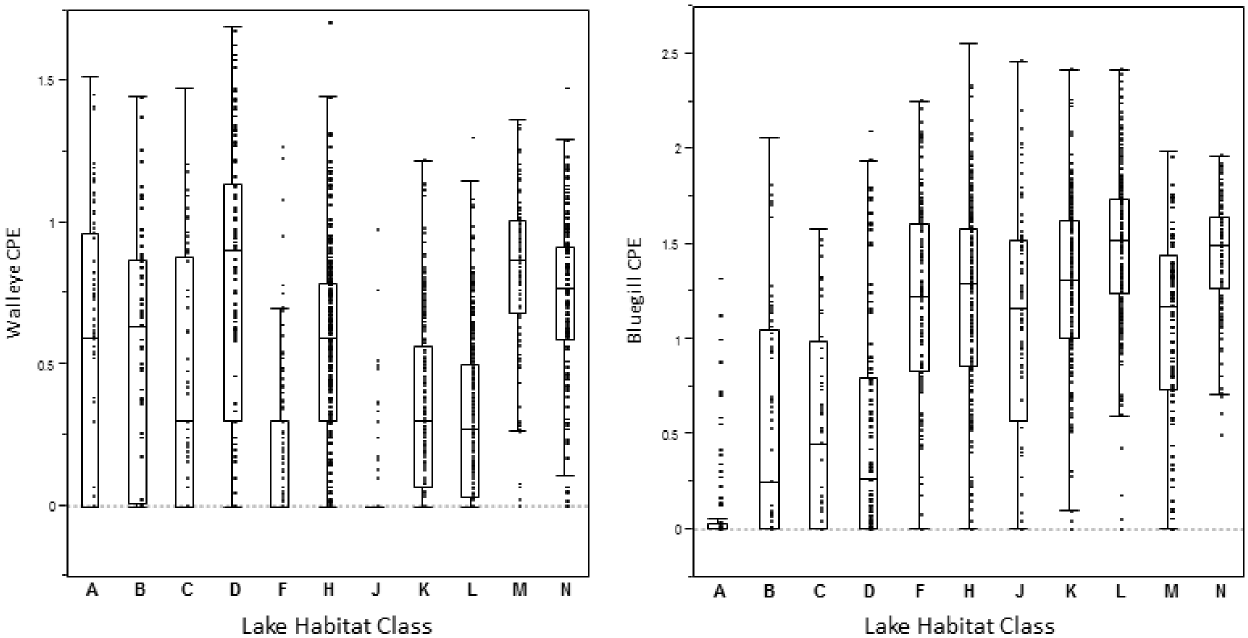


FIGURE 7. Distribution of log transformed mean lake gillnet walleye and trapnet bluegill catch-per-effort (by habitat lake class.

Fish response to habitat factors.–Random forest models provided more detailed information on WAE response to LSH. Statewide variation in WAE CPE using LSH variables revealed lake area having the highest variable importance (Table 9). Log transformed WAE CPE increased linearly with lake area on a log scale to around 4000 acres (Figure 8). However, for some northern lakes (B, C, and M) and prairie lakes (D) the effect on WAE CPE seems to reach a maximum at significantly smaller lake areas (400 to 1000 acres). Also, for lake classes D, K, L, and C it is obvious that WAE abundance is very limited when lake area is < 300 acres (Figure 8). The effect of lake depth on WAE CPE differed even more among LSH classes. Walleye CPE in class A lakes sharply decreases with increases in maximum depth up to 15 feet; whereas, the opposite occurs in class D lakes(Figure 9). This difference perhaps relates to the risk of winter mortality which decreases with increased depth in shallow prairie lakes, whereas Class A lakes are much less susceptible to winterkill so increases in depth are associated with a loss of productive littoral habitats that support higher WAE densities. Walleye RF models for class N and D lakes further demonstrate differences in the influence of habitat among LSH lake classes. SDI appears to be much more influential in Class N model than in either the statewide or Class D models. Also, Secchi was much more influential in Class D lakes than in the statewide or Class N models. Interestingly, despite the high influence of Secchi in the Class D model, disturbed land cover which generally correlates highly with Secchi, had only a small influence. Because virtually all Class D lake watersheds are highly agricultural it is likely the range of disturbance is insufficient to influence the model. Conversely, for class N and statewide models disturbance was a much stronger factor influencing WAE CPE than Secchi.

TABLE 9. Importance values (MSE % = percent included mean square error and Purity = included node purity) for large-scale habitat (LSH) predictors (ranked top to bottom in decreasing importance) in Random forest models of mean lake Walleye catch-per effort for all 1916 Minnesota lakes (statewide; R^2 = 44.2), lake habitat class N lakes (N = 130; R^2 = 37.8, and lake habitat class D lakes (N=99; R^2 =6.4). The percent watershed area with agriculture or development land cover is abbreviated as Pdist and day of the year fish sampled as Sday. All other predictors are previously used (LSH) variables.

Statewide			Class N			Class D		
Variables	MSE %	Purity	Variables	MSE %	Purity	Variables	MSE %	Purity
L-area	76.4	56.86	L-area	20.15	1.65	L-area	11.8	2.41
Sday	28.1	16.08	Pdist	17.2	0.96	Secchi	7.9	2.37
W-area	27.6	19.69	SDI	10.0	0.76	Water	4.9	1.19
Pdist	27.4	13.13	Water	9.0	0.59	W-area	4.6	1.52
Jtemp	26.3	12.11	Jtemp	5.5	0.68	Depth	2.5	1.57
Bflow	17.8	11.53	Sday	2.1	0.43	Bflow	0.1	1.00
Water	17.7	11.35	W-area	1.7	0.38	SDI	0.0	1.08
SDI	14.5	10.77	Secchi	1.4	0.37	Jtemp	-0.5	0.94
Depth	9.4	10.72	Bflow	1.1	0.23	Pdist	-1.1	0.89
Secchi	9.4	11.46	Depth	0.8	0.46	Sday	-1.7	0.73

Walleye RT models for class D and N lakes depict how LSH factors interact uniquely within lake classes to affect CPE (Figure 10). Regression tree models for class D resulted in seven subclasses explaining 45% of the variation in WAE CPE among lakes, and for class N lakes ten subclasses explaining 60% of variation in WAE CPE. Classes D and N both have high WAE CPE compared to other MN lakes (Figure 7). As in RF models, lake area was the most influential RT habitat factor affecting WAE, but the effect on class N lakes extends to 4000 acres, whereas the effect on class D lakes extends to < 500 acres. In larger class D lakes (> 340 acres), maximum depth appears to be most influential on WAE CPE; in the smaller lakes (< 340 acres), water clarity is most influential (Figure 10). The R2 fit of RT cross validation and RF models of class D WAE CPE were 0.30 and 0.09, respectively. This indicates the RT model likely overfits the dataset and has limited reliability as a predictive model perhaps indicative of their high level of disturbance (land use, winterkill ...). A better and more reliable fit was found with the class N RT model. For this class, RT cross validation and RF model R^2 were 0.49 and 0.38, respectively. Large class N lakes (> 1778 acres) had the highest WAE CPE. In mid-sized lakes (727 to 1778 acres) WAE were predominately influenced by lake shape (SDI) with higher CPE favored by a more circular perimeter. Finally, for small class N lakes (< 1778 acres) land cover disturbance was most influential with WAE CPE highest in lakes with more than 48% land cover disturbance (Figure 10).

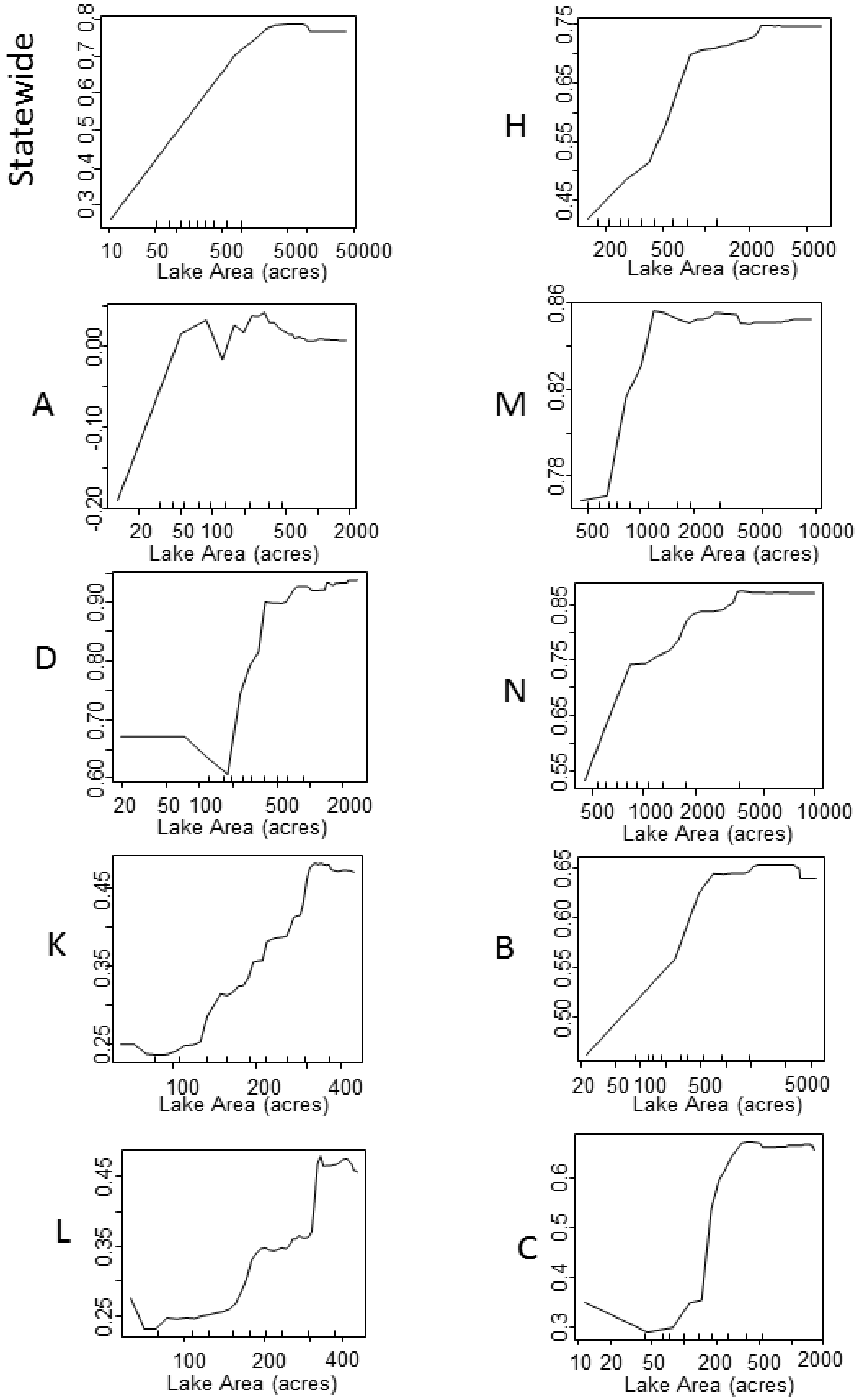


FIGURE 8. Response of log transformed walleye CPE to lake surface area (acres) in the statewide random forest model and in random forest models for selected lake classes. The x-axis scale reflects the distribution of lakes in each model (10% quantiles shown on x-axis).

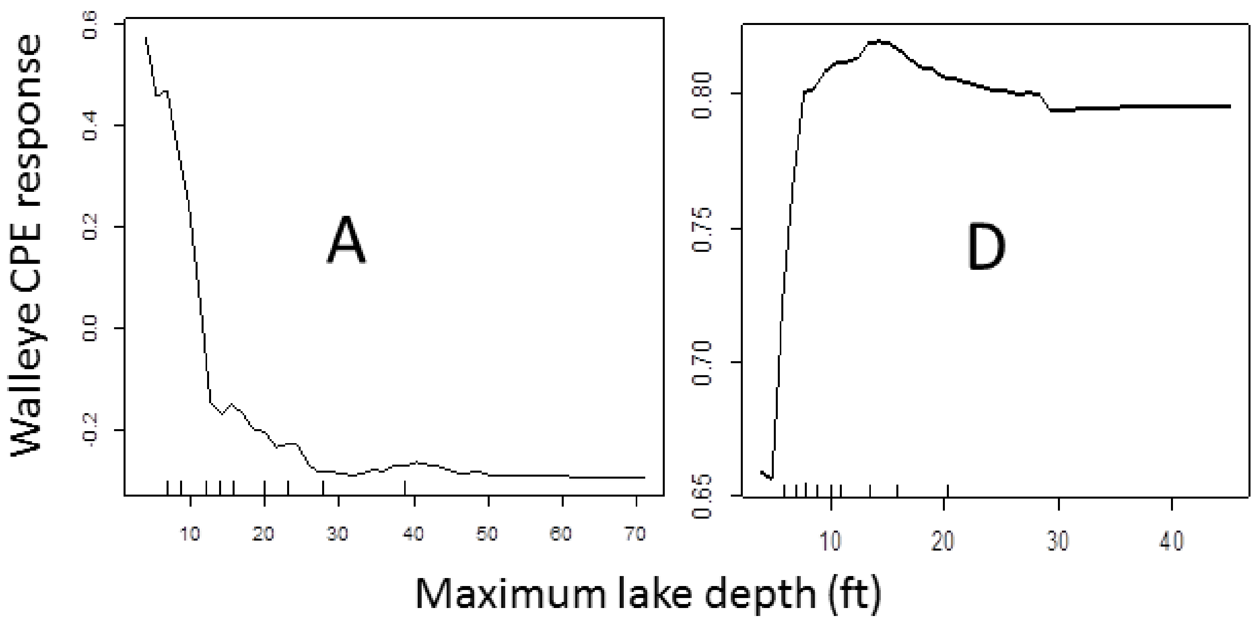


FIGURE 9. Response of walleye log transformed mean lake gill net catch-per-effort to maximum lake depth in lake habitat class A and D lakes.

Combining individual RT models of WAE CPE for all 11 LSH classes resulted in 62 classes explaining 57% of the variation in WAE CPE (Table 10). The amount of nonrandom spatial autocorrelation of residuals was also reduced using RT subclasses, thus improving the models accuracy (Table 10). Despite improvements spatial distribution of model residuals was still clumped with underestimates for lakes in central Minnesota and overestimates to the west (Appendix III and IV). Spatially autocorrelated residual hotspots are indicative missing explanatory variables; however, missing variables may not relate to a habitat factor.

Largemouth Bass RF models revealed that water clarity (Secchi) was the most influential LSH factor. Other influential LSH variables in order of importance in the statewide RF model were temperature, baseflow, and land cover disturbance (Table 11). These three factors relate strongly to geographic and ecoregional factors defining differences among LSH classes. Although the influence of Secchi on LMB was similar among LSH classes (mostly increasing between 3 to 5 m) many other differences in LMB response to habitat factors among LSH classes were evident (Figure 11; Table 11). For example, class M lakes were characterized with relatively high LMB CPE responding negatively to watershed area and lake area (above 900 acres), but the influence of these two factors is much less in statewide or class N models.

Differences between LSH influences in class N and class M lakes are also evident in LMB RT models (Figure 12). Both class N and M are comprised of larger mesotrophic lakes (lake area > 450 acres and Secchi > 1.9 m) located in the NLF ecoregion. However, the class N RT model explaining 39% of the variation in LMB CPE showed increased CPE with increased Secchi. The class M RT model explained 43% of the variability in LMB CPE among lakes with the highest CPE occurring in lakes between 626 and 910 surface acres with watershed areas < 16,522 acres (a ten-fold greater LMB CPE than in lakes with large watersheds located in areas of low baseflow).

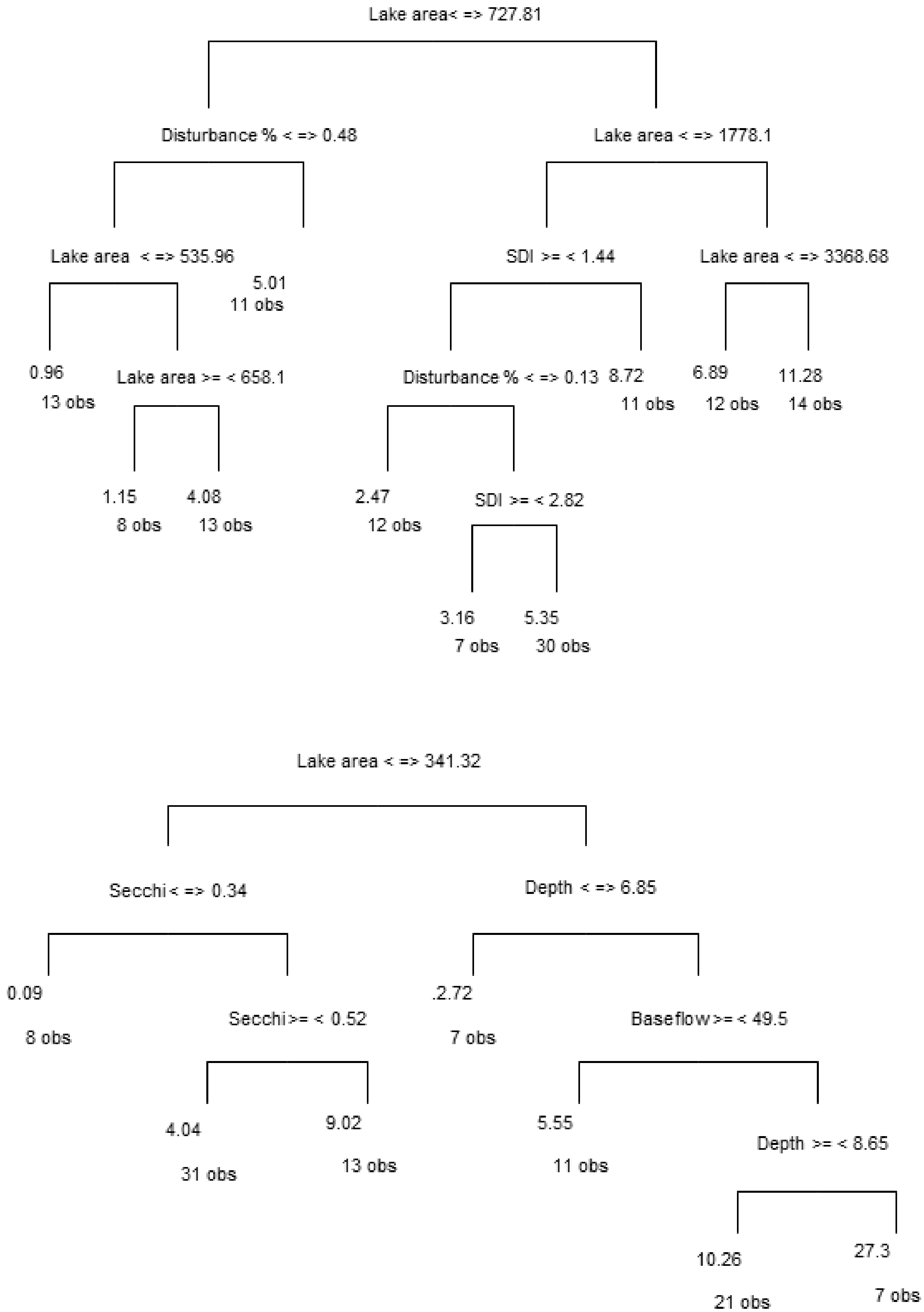


FIGURE 10. Regression tree of large-scale habitat predictors of Walleye log transformed mean lake gill net catch-per-effort for habitat class N (top; N=131, R^2=0.60, 5-fold cross validation R^2=0.49) and habitat class D (bottom; N=99, R^2 = 0.45, 5-fold cross-validation R^2=0.30).

TABLE 10. Fit statistics for models predicting walleye catch-per-effort (CPE) in Minnesota lakes. The R^2 statistic was used to explain the amount of variation explained by the model and the Morans Index was used to assess the spatial distribution of the variation to see if the pattern is clustered, dispersed or random. The lower the Morans index and corresponding z score the more random the distribution.

Model	R^2	Morans index	z score
Multiple regregession	0.41	0.116	19.323
Decision Tree	0.29	0.093	11.789
Decision Tree -RTA	0.57	0.056	7.113
Ecological Classification	0.36	0.095	11.944

TABLE 11. Importance values (MSE % = percent included mean square error and Purity = included node purity) for large-scale habitat (LSH) predictors (ranked top to bottom in decreasing importance) in random forest models of mean lake Largemouth Bass catch-per effort for all 1916 Minnesota lakes (statewide; R^2 = 30.6), lake habitat class N lakes (N = 130; R^2 = 8.6, and lake habitat class M lakes (N=93; R^2 = 16.2). The percent watershed area with agriculture or development land cover is abbreviated as Pdist and day of the year fish sampled as Sday. All other predictors are previously used (LSH) variables.

Statewide			Class N			Class M		
Variables	MSE %	Purity	Variables	MSE %	Purity	Variables	MSE %	Purity
Secchi	41.6	7.56	Secchi	11.2	0.39	W-area	11.2	0.27
Jtemp	28.9	4.35	Bflow	8.6	0.28	L-area	9.7	0.23
Bflow	23.7	4.25	Jtemp	5.8	0.20	Secchi	4.7	0.20
Pdist	18.6	3.23	W-area	5.3	0.26	Water	3.7	0.16
Depth	18.2	3.31	Pdist	3.3	0.17	Pdist	2.5	0.12
Sday	17.7	3.62	Sday	2.9	0.26	SDI	1.9	0.10
Water	16.1	3.78	Water	1.6	0.24	Jtemp	1.4	0.14
W-area	12.1	3.29	L-area	0.4	0.19	Depth	1.3	0.12
L-area	12.1	3.07	Depth	-2.5	0.22	Bflow	0.1	0.08
SDI	2.9	2.12	SDI	-2.7	0.14	Sday	-0.5	0.09

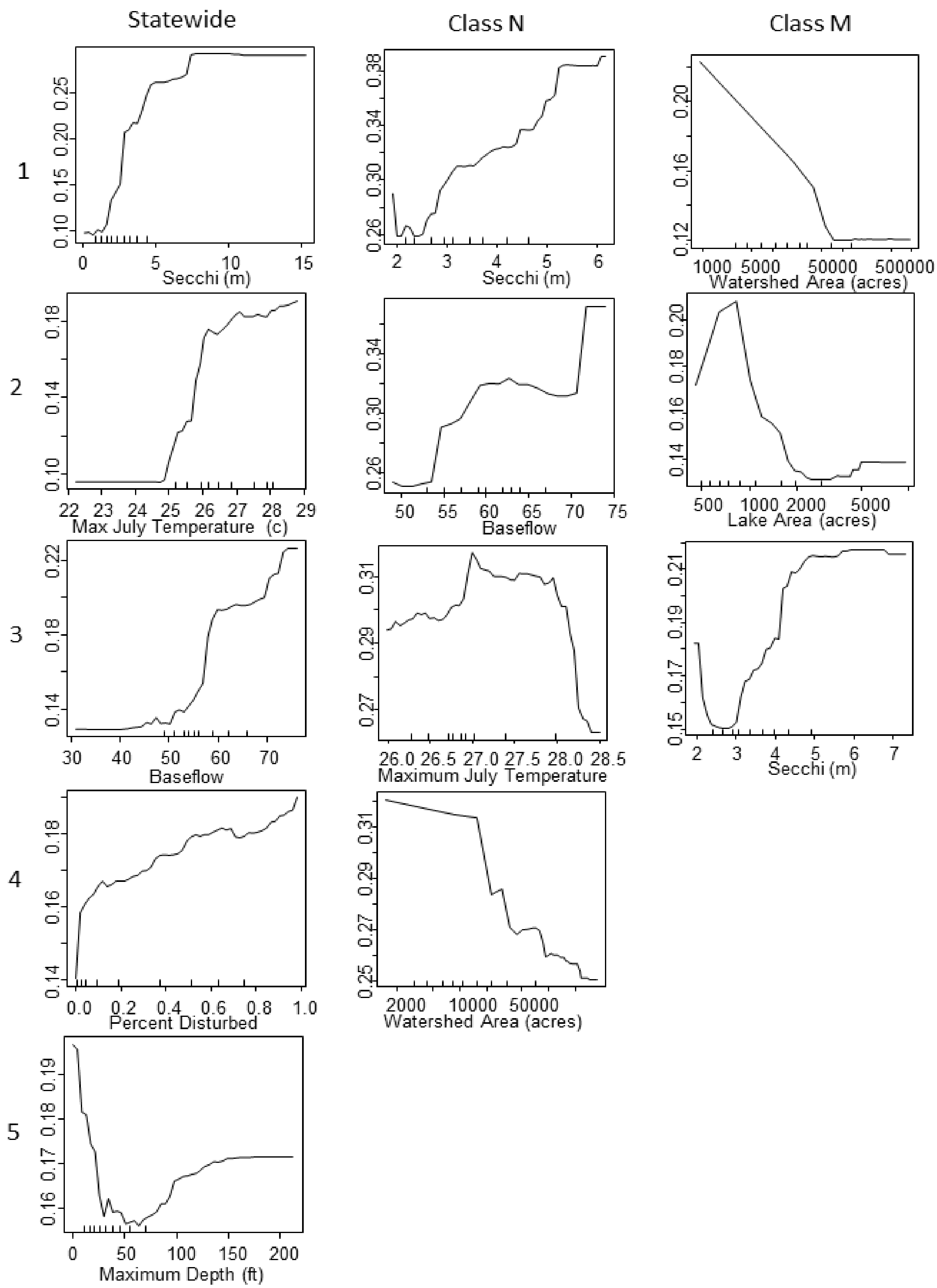


FIGURE 11. Response of Largemouth Bass log transformed mean lake gill net catch-per-effort to large-scale habitat variables in random forest models for all Minnesota lakes (Statewide n=1916), lake habitat class N (N=131), and lake habitat M (N= 123). The x-axis scale reflects the distribution of lakes in each model (10% quantiles shown on x-axis).

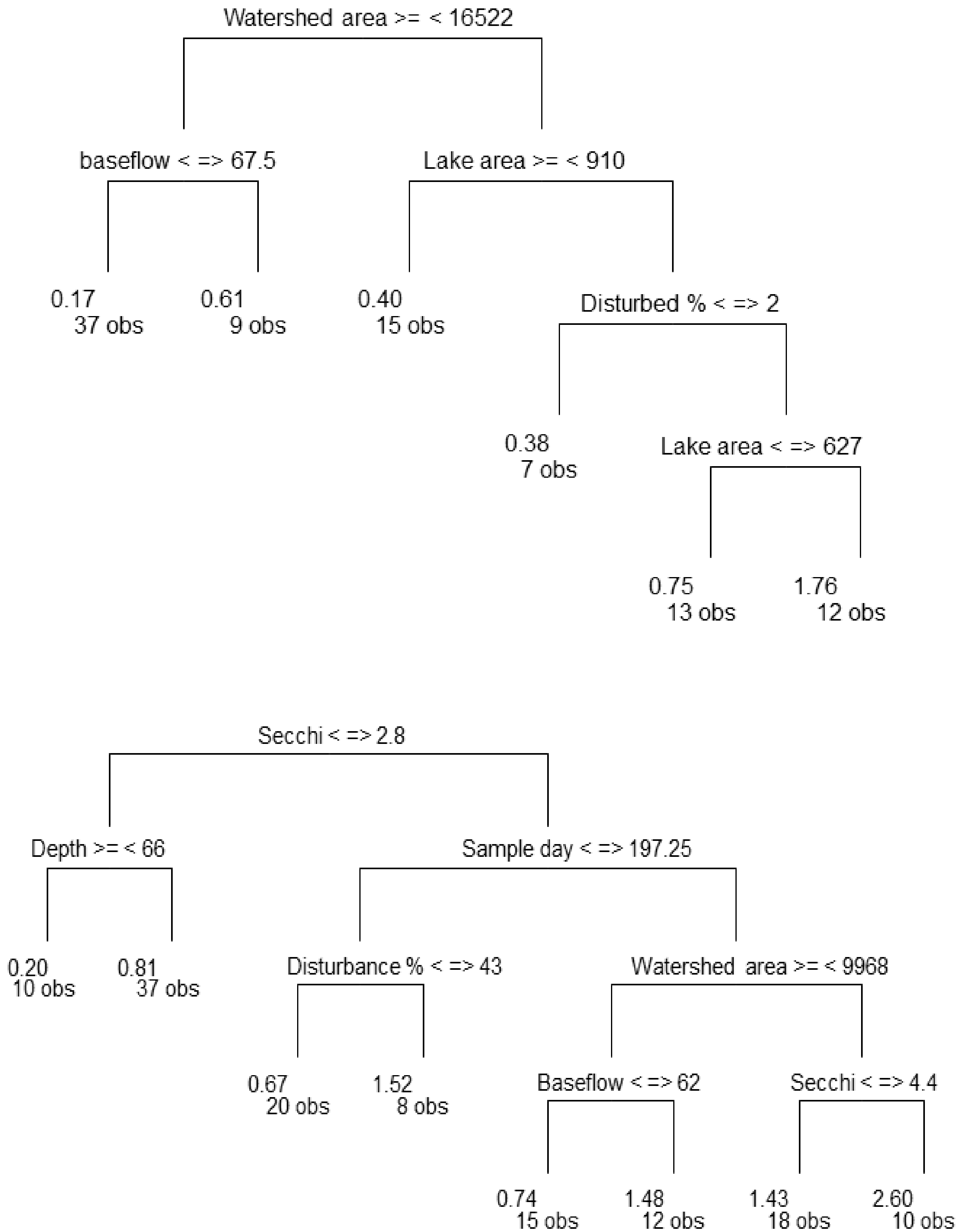


FIGURE 12. Regression tree of large-scale habitat predictors of Largemouth Bass gill net catch-per-effort for habitat class M (top; R^2 = 0.43, 5-fold cross-validation R^2=0.33) and habitat class N (bottom; R^2=0.39, 5-fold cross validation R^2=0.30).

DISCUSSION

Large-scale habitat factors explained a significant amount of variation in baseline (18-y average) fish abundances among Minnesota lakes. Hierarchal decision tree classification of lakes by LSH attributes identified primary habitat influences on the type and abundance of fish in Minnesota lakes. This classification facilitated further comparisons among lakes because it identified classes of lakes with similar LSH influences that were relatively robust to interactions and nonrandom spatial variability affecting fish abundance. Models developed from this study should be useful for setting goals for fish population maintenance, restoration, or enhancement actions based on habitat and for troubleshooting and development of new hypotheses when goals are not met.

Data mining and GIS analysis of an updated (1993-2011) dataset of standardized MNDNR lake surveys reinforced previous studies by Moyle (1956) and Schupp (1992) showing lake morphology, fertility, and temperature influencing the species and abundance of fish present in Minnesota lakes. GIS provided spatial analytical capabilities to use with ecoregion, watershed, and other geographical attributes for obtaining a more quantitative description of factors contributing to lake fish community differences. Moyle (1956) and Schupp (1992) concluded that key fish habitat factors follow a general geographical gradient from northeast to southwest in Minnesota which is largely responsible for differences in fish communities among Minnesota lakes. This study showed that much of this geographical variation is quantified by differences among aquatic ecoregions which have an additional advantage of over other variables because they are widely recognized with concisely mapped geographical boundaries. Aquatic ecoregions identify geographical areas with LSH attributes in common, meaning that differences in geomorphology, (including some lake morphology), climate, soils, and plant cover are relatively uniform within areas defined by ecoregions (Omernik et al. 2004). The Level IV EPA ecoregion scheme used in this study characterizes influences on water body chemistry and places emphasis on hydrology (Omernik et al. 2000). Differences in fertility among Minnesota lakes correspond closely with ecoregion units (Heiskary and Wilson 2008; Cross and Jacobson 2013). Lake temperatures were not specifically measured in this study, but much can be inferred with surrogate measures used in this study such as mean July maximum air temperature, latitude, and ecoregion. Once again these findings confirmed basic observations made by Moyle (1956) and Schupp (1992) while providing more spatial resolution. While temperature was an influential variable for describing differences in fish communities among Minnesota lakes, it appeared less influential relative to other habitat variables used by Wehrly et al. (2012) to analyze Michigan lakes. Also, Wehrly et al. (2012) used modeled lake temperatures as opposed to surrogate values and described fish community differences in terms of species presence rather than abundance.

An 11-class hierarchical decision tree using geographical location (ecoregion and latitude), lake morphology (surface area), and trophic status (Secchi) successfully explained a substantial amount of among lake variation in fish abundance. For most fish species, more variation was explained by the 11 LSH classes than with the 43 lake classes developed by Schupp (1992). Schupp (1992) did not use fish community gradients to guide or supervise lake classification and it represents a purer grouping of lakes with respect to commonalities in physico-chemical attributes. However, hierarchical decision tree methodology yielded a highly interpretable LSH lake classification guided by dominant fish community gradients assimilating species interactions. Among others, Robinson and Tonn (1989) demonstrated the importance of acknowledging species interactions in small lakes where predation was more important than habitat factors in structuring the composition of fish communities. Still, using fish community gradients simultaneously with LSH factors to classify lakes could complicate interpretation of habitat factors if communities are altered by management practices (stocking, fishing, etc.) or habitat degradation (shoreline development, agriculture, etc.). However, post hoc analysis can be used to address these factors. Validation of the LSH lake classification scheme was demonstrated with the fit to WAE CPE which was also congruent with habitat relationships reported in the literature for that species (Bozek et al. 2011). Despite being widely stocked, relationships between WAE abundance and lake size and productivity conform to general descriptions of habitat provided by Scott and

Crossman (1973) as well as scientific reports on North American lakes studied by Kitchell et al. (1977), Johnson et al. (1977), Lester et al. (2004), as well as those specific to Minnesota lakes (Schupp 1992; Jacobson and Anderson 2007). Most prominent is the habitat need of WAE for larger lakes consistent with a “Lebensraum requirement” stated by other investigators (Johnson et al. 1977; Jacobson and Anderson 2007) which was shown in Figure 8.

Although the use of fish communities to supervise lake classification resulted in good R^2 values for CPE of most fish species, the relative paucity of fish survey data for northeastern Minnesota lakes (CSSH ecoregion) limited the analysis. Consequently, analysis of lake habitats in northeast Minnesota was limited, likely contributing to a poorer fit of the LSH classes relative to ELC classes for cooler water species such as TLC, WAE, and RKB. Nineteen lake classes are devoted to lakes in the three northeastern counties in the ELC system where only three LSH classes occur in those same counties.

Large-scale habitat lake classes were modeled using fish community criteria so fish species interactions were implicit in the classification. Similar correlations with fish community PC scores were observed for both WAE and YEP at every split except 2A1 and 2A11 (northern ecoregions; Table 6, Figure 4); whereas dissimilar correlations were observed for WAE and NOP with PC scores at every split except 2B1 and 2B21 (lake area; Table 6, Figure 4). General associations of WAE and YEP are well documented in the literature as well as interactions with NOP predation (Colby et al. 1987; Jacobson and Anderson 2007). However, few actual experiments have ever been done to prove the existence of species interactions. Another species interaction of concern to fisheries managers is that between WAE and BLC. Colby et al. (1987) indicated an inverse relationship between these species was strongest in moderately deep clear central Minnesota lakes and speculated that it was due to greater macrophyte densities that favored BLC in these lakes. Indeed, a strong inverse WAE-BLC interaction based on Secchi depth in NLF lakes was observed for the PC defining split 2B2 (Table 6; Figure 4). However, there was also a strong inverse WAE-BLC relationship indicated with PC splits 2B22 and 2B212 that did not correspond with Secchi, but instead described a latitudinal split within the NLF ecoregion lakes suggesting conditions favoring BLC over WAE in more southerly lakes(Tables 6 and 7; Figure 4). In general, more southerly latitudes and associated warmer water temperatures are thought to favor Crappie over WAE (Hokanson 1977; Kitchell et al. 1977).

Modeling of species abundances within individual LSH lake classes described more precise relationships than explained using the statewide classification alone. Although ecoregion membership served as useful criteria for LSH lake classification, they are based on spatially correlated landscape/climate variables so the contributions of individual variables were less apparent. However, WAE and LMB responses to specific habitat variables modeled within LSH lake classes using RF yielded response curves more precisely described influences of individual variables (i.e. land use disturbance, groundwater, and temperature). Modeling within individual LSH lake classes helped eliminate confounding influences of LSH factors of little relevance for a specific class. For example, models developed for Minnesota prairie lakes (LSH class D) were less confounded by lake depth > 30 feet and baseflow > 54% conditions not applicable to that class. The stark contrast between class D and A lakes in WAE CPE response to lake depth (Figure 9) illustrated the importance of modeling habitat responses for each lake class separately. Individual lake class specific models also showed contrasting WAE response to lake size (Figure 8). Lake class specific species–habitat relationships provided additional insight into fish community interactions. This was illustrated by the negative influence of watershed size and lake area on Largemouth Bass CPE in LSH lake class M (northerly NLF lakes > 450 acres and < 1.9 m Secchi) in contrast to the positive influence both these habitat factors have on competing predator species (NOP and WAE). Lake class specific models provided more specific information describing habitat relationships than that provided using statewide classifications alone.

Habitat models developed in this study were subject to data and methodological limitations that potentially lead to poor interpretations if not fully examined. Fish data were subject to bias inherent with the nonrandom lake selection process used to establish the survey database and sampling methods. Rather than random sampling, lakes represented in the MNDNR lake survey database are sampled in proportion to management priorities influenced by demographic and lake quality factors. Sampling gear (trap net and gill net) efficiencies differ for different species and sizes of fish and do not necessarily accurately reflect true population densities in lakes. Furthermore, catch efficiencies can differ with sampling date (Grant et al. 2004). Sampling date was included in lake class specific modeling to control for seasonal variation, but not in the LSH lake classification. Also, fish data used in classification and modeling were restricted to an 18 year average or baseline CPE and did not include factors describing biomass and mean individual size (though generally correlated with CPE). Analysis of average conditions do not reflect temporal variation, so lakes suspected to have high annual variation (e.g. shallow or disturbed lakes) could have fish communities different from those predicted by habitat models if the number of surveys in the study period were few. Shallow lakes are more likely to have regime shifts and are vulnerable to differences in temperature, DO, and water levels (Scheffer 2001). Furthermore, all lakes are subject to changes in climate and watershed land use factors that could contribute to a shifting baseline condition.

Regression tree classification of lakes was predicated on the ability of LSH variables describing differences in fish communities quantified by PCA. Therefore, it is important to recognize that PCA performs best on datasets with abundant species and linear correlations. Also, outliers can potentially have a large influence on correlations (McCune and Grace 2002). Regression tree models are transparent and easy to interpret, but are known to be susceptible to overfitting and can sometimes change with small differences in the data (Prasad et al 2006). However, RT models were verified with RF models that are generally more accurate, do not overfit the data, and provided variable importance indices and response curves (Prasad et al. 2006; Cutler et al. 2007). Furthermore, both linear and nonlinear responses to habitat factors were shown with RF variable response curves which can be useful for evaluating what if scenarios, an invaluable tool for diagnosing past performance problems or projecting future responses. In a review of similar studies, Hawkins et al. (2010) found "*substantial variation in biota associated with natural environmental gradients even after modeling or adjusting for the effects of natural environmental variation*". In support of this observation, they cited the recent USA National Wadeable Stream Assessment (USEPA 2006) that used RF models to show that up to 30% more variation in biological index variables for individual ecoregions were associated with natural environmental gradients. Machine learning algorithms such as RF models provide a relatively objective method for showing fish responses to habitat. That being said, the accuracy of the models is only as good as the data supplied them which heavily relies on the appropriate selection of explanatory variables as discussed in the previous paragraph. Key habitat components are not always known and oftentimes habitat variables are correlated, both of which can affect model outcomes. These factors were addressed with preliminary exploratory analysis using correlation matrices to identify relatively independent variables and mapping of model residuals to look for spatial clumping that can often identify environmental factors absent in the model.

Many earnest attempts to "improve" habitat conditions for fish have failed due to a lack of understanding of the effects of LSH factors on fish communities (Ziemer 1997). This lack of understanding could be partially addressed with relationships revealed with models such as those demonstrated in this study or subsequent efforts to advance them. Still, LSH models are only helpful if applied correctly in recognition that they are most applicable to lakes within the range of habitat conditions used as input taking into account factors not explicitly modeled, including various human influences and species interactions. Deviation outside the naturally occurring normal range of habitat conditions for a particular water body elevates the risk of unanticipated consequences. That being said, anthropogenic stresses can push habitat conditions outside the normal range forcing a need to address these situations despite the risk.

Management Implications and Needs for Additional Research.–Habitat management issues confronting authorities are daunting and continually require decisions to protect, regulate, restore, or enhance to attain fish community and sport angling goals. Lake classification based on LSH can be used by managers to identify lakes at a statewide scale with relatively homogeneous habitat influences from which to draw comparisons. Like the ecological lake classification (Schupp 1992), the LSH lake classification could be used to identify management units of similar lakes for targeting management actions, for stratifying lakes to gain statistical power, and for extrapolating results from more intensive or mechanistic investigations on specific lakes to a broader population of lakes. According to Bakelaar et al. (2004) GIS classification, analysis, and habitat modeling should play a pivotal role in the assessment of conservation planning issues. For Minnesota lakes such work can augment ongoing monitoring efforts aimed at predicting and planning for potential changes in sport fish populations resulting from climate and landscape changes that affect land cover and connectivity of aquatic habitats. Also, LSH lake classification could be used by sampling programs seeking to optimize the distribution of effort by stratifying lakes with similar habitats.

Habitat models specific to individual lake classes and fish species can be used to objectively identify and diagnose habitat related problems. Modeling within lake classes removes much of the variation associated with spatial autocorrelation at the statewide scale allowing for more direct habitat specific answers. Managers should find the use of open source machine learning tools, such as those contained in the Rattle package implemented in the R platform, useful to facilitate the analysis. They provide a means to perform the analysis with the transparency and robustness helpful for dealing with the type of data, interactions, and nonlinear relationships common to ecological systems.

Additional research directed at habitat classification and modeling could broaden the scope and applicability of this study including specific efforts directed at different spatial and temporal scales along with additional habitat factors. Fish habitat relationships differ across spatial scales and this study only addressed variation at the statewide and level IV ecoregion scales. Additional knowledge regarding larger scale factors can be obtained from integration with multi-state and international models and finer scale factors from spatial units smaller than level IV ecoregions. Factors such as climate, land cover, geomorphological, and lake productivity factors that have a similar pattern of spatial correlation within Minnesota could be discriminated better at larger spatial scales. Documentation of large-scale spatially robust relationships to habitat factors are potentially helpful because it infers that a knowledge gained from one lake can be more easily transferred to another lessening the need for costly additional data and evaluation. Conversely, more detailed relationships influenced by more local interactions unique to smaller spatial units are described better at a smaller scale. This includes the integration of more casual and mechanistic modeling of habitat such as those involving aquatic plant communities, phosphorus dynamics, and dissolved oxygen demand, and winterkill-summerkill risks moderated by climate.

There are still numerous information gaps that need to be addressed with further exploration of habitat classification and models of fish habitat relationships. This study was limited to fish communities described by larger ubiquitous fish species sampled with trap and gill nets, so sampling with a broader array of gear reflecting all fish species occurring in Minnesota lakes could add much needed insight comparable to modeling done by Wehrly et al. (2012). Also this additional analysis could better account for habitat needs of sensitive species and not just the more common habitat generalists analyzed in this study.

Additional studies should be done to relate LSH to physical substrates and aquatic plant cover (e.g. emergent, fine-leaf, broad-leaf, invasive classes identified by Cross and McInerny 2006 or Reschke et al. 2005) and to determine the extent that Secchi acts as surrogate variable for the amount/depth of plant cover. Currently, there are few quantitative descriptions available showing how fish communities are affected by inshore substrates. Furthermore, there is no spatial analysis of substrate occurrence among Minnesota lakes corresponding with spatial variation in fish and plant abundance. Mechanistic models of the relationship between wind-wave energy (lake morphometry, topography, wind) and mud boundaries or deposition zone (e.g. Cooley and Franzin 2008) would be useful applied to Minnesota lakes. Finally, knowledge of habitat interactions with stocking and exploitation would be useful for identifying specific effects of habitat on fish communities in addition to retrospective analysis showing the effect on fish communities from historical changes in lake habitat conditions.

REFERENCES

Bakelaar, C. N., P. Brunette, P. M. Cooley, S. E. Doka, E. S. Millard, C. K. Minns, and H. A. Morrison. 2004. Geographic information systems applications in lake fisheries. Pages113-152 *in* W. L. Fisher and F. J. Rahel, editors. Geographic information systems in fisheries. American Fisheries Society, Bethesda.

Blann, K., and M. Cornett. 2008. Identifying lake conservation priorities for The Nature Conservancy in Minnesota, North Dakota and South Dakota. The Nature Conservancy. Arlington, VA.

Bozek, M. A., T. J. Haxton, and J. K. Raabe. 2011. Walleye and sauger habitat. Pages 133-197 *in* B. A. Barton, editor. Biology, management, and culture of walleye and sauger. American Fisheries Society, Bethesda, Maryland.

Colby, P. J., P. A. Ryan, D. H. Schupp, and S. L. Serns. 1987. Interactions in north-temperate lake fish communities. Canadian Journal of Fisheries and Aquatic Science 44:104-128.

Cooley, P. M., and W. G. Franzin. 2008. Predicting the spatial mud energy and mud deposition boundary depth in a small boreal reservoir before and after draw down. Lake and Reservoir Management 24:261-272.

Cross, T. K., and P. C. Jacobson. 2013. Landscape factors influencing lake phosphorus concentrations across Minnesota. Lake and Reservoir Management 29:1-12.

Cross, T. K., and M. C. McInerny. 2005. Spatial habitat dynamics affecting bluegill abundance in Minnesota bass-panfish lakes. North American Journal of Fisheries Management 25:1051-1066.

Cross, T. K., and M. C. McInerny. 2006. Relationships between aquatic plant cover and fish populations based on Minnesota lake survey data. Minnesota Department of Natural Resources, Division of Fisheries, Investigational Report Number 537. St. Paul.

Cutler, R. D., T C Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Ecology. 88:2783-2792.

Emmons, E. E., M. J. Jennings, and C. Edwards. 1999. An alternative classification method for northern Wisconsin lakes. Canadian Journal of Fisheries and Aquatic Sciences 56: 661-669.

Esri. 2012. ArcMap10. Esri. Redlands, CA. C. Daly, M. Halbleib, J. I. Smith, W. P. Gibson, M K. Doggett, G. H. Taylor, J. Curtis, and P. A. Pasteris. 2008. Physiographically-sensitive mapping of temperature and precipitation across the conterminous United States. International Journal of Climatology, 28: 2031-2064.

Grant, G. C., Y. Schwartz, S. Weisberg, and D. H. Schupp. 2004. Trends in abundance and mean size of fish captured in gill nets from Minnesota lakes, 1993-1997. North American Journal of Fisheries Management 24.417-428.

Hawkins, C. P., R. R. Olson, and R. A. Hill. 2010. The reference condition: Predicting benchmarks for ecological and water-quality assessments. Journal of the North American Benthological Society 29:312-343.

Heiskary, S. A., and C. B. Wilson. 2008. Minnesota's approach to lake nutrient criteria development. Lake and Reservoir Management 24:282-297.

Hokanson, K. E. F. 1977. Temperature requirements of some percids and adaptations to the seasonal temperature cycle. Journal of the Fisheries Research Board of Canada 34:1524-1550.

Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan. 2004. Development of a 2001 national land-cover database for the United States. Photogrammetric Engineering and Remote Sensing 70:829-840.

Jackson, D. A., P. R. Peres-Neto, and J. D. Olden. 2001. What controls who is where in freshwater fish communities - the roles of biotic, abiotic, and spatial factors. Canadian Journal of Fisheries and Aquatic Sciences 58:157-170.

Jacobson, P. C., and C. S. Anderson. 2007. Optimal stocking densities of walleye fingerlings in Minnesota lakes. North American Journal of Fisheries Management 27:650-658.

Johnson, M. G., and J. H. Leach, C. K. Minns, and C.H. Olver. 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). Journal of the Fisheries Research Board of Canada 34:1592-1601.

Kitchell, J. F, M. G. Johnson, C. K. Minns, K. H. Loftus, L. Greig, and C. H. Olver. 1977. Percid habitat: The river analogy. Journal of the Fisheries Research Board of Canada 34:1936- 1940.

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

McCune, B., and J. B. Grace. 2002. Analysis of ecological communities. MjM Software Design. Gleneden Beach, Oregon.

Minnesota Department of Natural Resources (MNDNR). 1993. Manual of instructions for lake survey. Minnesota Department of Natural Resources Section of Fisheries Special Publication 147, St. Paul.

Minnesota Department of Natural Resources (MNDNR). 2012. DNR Watersheds - DNR Level 09 - DNR AutoCatchments, Originator: Minnesota DNR - Fisheries Publication Date: 9/13/2012.

Minnesota Department of Natural Resources (MNDNR). 2013a. Fish habitat plan: A strategic guidance document. Minnesota Department of Natural Resources, Section of Fisheries. St. Paul.

Minnesota Department of Natural Resources (MNDNR) 2013b. Streams with Strahler stream order. Available online at http://deli.dnr.state.mn.us/metadata.html?id=L390005970202. Last accessed on April 1, 2013.

Moyle, J. B. 1945. Some chemical factors influencing the distribution of aquatic plants in Minnesota. American Midland Naturalist 34:402-420.

Moyle, J. B. 1956. Relationships between the chemistry of Minnesota surface waters and wildlife management. Journal of Wildlife Management 20:303-320.

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

Omernik J. M., S. S. Chapman, R. A. Lillie, R. T. Dumke. 2000. Ecoregions of Wisconsin. The Wisconsin Academy of Sciences 88:77-103.

Omernik, J. M. 2004. Perspectives on the nature and definition of ecological regions. Environmental Management 34:S27-S38.

Osgood, R.A., P.L. Brezonik, and L. Hatch. 2002. Methods for classifying lakes based on measures of development impacts. University of Minnesota Water Resources Center Technical Report 143.

Pittman S. J., and K. A. Brown. 2011. Multi-scale approach for predicting fish species distributions across coral reef seascapes. PLoS ONE 6(5): e20583.

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

R Development Core Team. 2013. R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria. Available from http://www.R-project.org/.

Rahel, F. J., and D. A. Jackson. 2007. Watershed level approaches. Pages 887-946 *in* C. S. Guy and M. L. Brown, editors. Analysis and interpretation of fisheries data. American Fisheries Society, Bethesda.

Reschke, C., G. E. Host, and L. C. Johnson. 2005. Evaluation of DNR aquatic vegetation surveys: Data summaries and comparative analysis. Minnesota Department of Natural Resources CFMS Contract Number A61156, St. Paul.

Robinson, C. L. K., and W. M. Tonn. 1989. Influence of environmental factors and piscivory in structuring fish assemblages of small Alberta lakes. Canadian Journal of Fisheries and Aquatic Science 46:81-89.

Schupp, D. H. 1992. An ecological classification of Minnesota lakes with associated fish communities. Minnesota Department of Natural Resources, Section of Fisheries Investigational Report Number 417, St. Paul.

Scheffer, M. 2001. Ecology of shallow lakes. Kluwer Academic Publishers. Boston.

Scott, W.B., and E.J. Crossman. 1973. Freshwater fishes of Canada. Fisheries Research Board of Canada, Ottawa.

Soranno, P. A., K. E. Webster, K. S. Cheruvelil, and M. T. Bremigan. 2009. The lake landscape-context framework: linking aquatic connections, terrestrial features and human effects at multiple spatial scales. Verh. Internat. Verein. Limnol. 30:695-700.

SAS Institute Inc. 2012. JMP 10 modeling and multivariate methods. SAS Institute Inc. Cary, NC.

Strahler, A. N. 1957. Quantitative analysis of watershed geomorphology. Transactions of the American Geophysical Union 38:913-920.

United States Geological Survey (USGS). 2012. Federal standards and procedures for the National Watershed Boundary Dataset (WBD). U.S. Geological Survey, Reston.

Wehrly, K. E., J. E. Breck, L. Wang, L. Szabo-Kraft. 2012. A landscape-based classification of fish assemblages in sampled and unsampled lakes. Transactions of the American Fisheries Society 141:414-425.

Williams, G. J. 2009. Rattle: A data mining GUI for R. The R Journal 1: 45-55.

Williams, G. J. 2011. Data mining with Rattle and R: The art of excavating data for knowledge discover. Springer, New York.

Williams, J. E., C. A. Wood, and M. P. Dombeck, editors. 1997. Watershed restoration: Principles and practices. American Fisheries Society, Bethesda.

Wolock, D. M. 2003. Base-flow index grid for the conterminous United States. U.S. Geological Survey Open-File Reprt Issue Identifiation: 03-263. Reston. Accessed March 15, 2003 http://water.usgs.gov/lookup/getspatial?bfi48grd.

Zeimer, R. R. 1997. Temporal and spatial scales. Pages 80-95 *in* J. E. Williams, C. A. Wood, and M. P. Dombeck, editors. Watershed restoration: principles and practices. American Fisheries Society, Bethesda.

APPENDIX I. A matrix of Pearson correlation coefficients for primary species sampled by standardized gill and trap net assessments in 1916 Minnesota lakes.

	WTS	BLB	CAP	YEP	BLC	GSF	COLD	SMB	WAE	BOF	RKB	NOP	YEB	PMK	BLG	LMB
WTS	1.00	-0.07	0.05	0.27	-0.13	-0.09	0.09	0.11	0.35	-0.09	0.13	-0.08	-0.17	-0.15	-0.28	-0.19
BLB	-0.07	1.00	0.52	0.20	0.33	0.17	-0.21	-0.12	0.07	-0.09	-0.27	-0.19	-0.09	-0.17	-0.15	-0.17
CAP	0.05	0.52	1.00	0.17	0.40	0.06	-0.14	-0.06	0.20	0.03	-0.19	-0.14	-0.03	-0.26	-0.13	-0.17
YEP	0.27	0.20	0.17	1.00	0.10	0.00	-0.06	0.01	0.35	0.01	0.07	0.00	-0.07	0.06	-0.07	-0.13
BLC	-0.13	0.33	0.40	0.10	1.00	0.05	-0.23	-0.10	-0.02	0.06	-0.24	-0.08	0.02	0.01	0.22	-0.08
GSF	-0.09	0.17	0.06	0.00	0.05	1.00	0.02	-0.04	-0.04	-0.03	-0.04	-0.11	0.02	0.04	0.09	-0.04
COLD	0.09	-0.21	-0.14	-0.06	-0.23	0.02	1.00	0.10	0.05	0.09	0.40	-0.08	-0.01	0.05	-0.02	-0.03
SMB	0.11	-0.12	-0.06	0.01	-0.10	-0.04	0.10	1.00	0.18	-0.06	0.26	-0.11	-0.13	-0.10	-0.07	-0.03
WAE	0.35	0.07	0.20	0.35	-0.02	-0.04	0.05	0.18	1.00	-0.01	0.25	-0.12	-0.03	-0.11	-0.10	0.00
BOF	-0.09	-0.09	0.03	0.01	0.06	-0.03	0.09	-0.06	-0.01	1.00	0.05	0.25	0.36	0.22	0.24	0.04
RKB	0.13	-0.27	-0.19	0.07	-0.24	-0.04	0.40	0.26	0.25	0.05	1.00	0.09	0.02	0.19	0.15	0.14
NOP	-0.08	-0.19	-0.14	0.00	-0.08	-0.11	-0.08	-0.11	-0.12	0.25	0.09	1.00	0.35	0.40	0.32	0.22
YEB	-0.17	-0.09	-0.03	-0.07	0.02	0.02	-0.01	-0.13	-0.03	0.36	0.02	0.35	1.00	0.37	0.43	0.24
PMK	-0.15	-0.17	-0.26	0.06	0.01	0.04	0.05	-0.10	-0.11	0.22	0.19	0.40	0.37	1.00	0.50	0.24
BLG	-0.28	-0.15	-0.13	-0.07	0.22	0.09	-0.02	-0.07	-0.10	0.24	0.15	0.32	0.43	0.50	1.00	0.34
LMB	-0.19	-0.17	-0.17	-0.13	-0.08	-0.04	-0.03	-0.03	0.00	0.04	0.14	0.22	0.24	0.24	0.34	1.00

APPENDIX II. Mean, median, first and third quantiles of catch-per-effort for primary fish species captured in standardized MNDNR lake survey gill and trap-nets (1993-2011) by large-scale habitat (LSH) lake class.

Gillnet

		LMB				NOP				WAE				SMB			
CLASS	N	Mean	Median	Q25	Q75	Mean	Median	Q25	Q75	Mean	Median	Q25	Q75	Mean	Median	Q25	Q75
A	107	0.07	0.00	0.00	0.00	2.33	2.75	0.00	5.42	2.51	3.00	0.00	7.83	0.18	0.00	0.00	0.00
B	85	0.10	0.00	0.00	0.04	3.30	3.47	1.82	5.75	2.55	2.87	0.00	6.53	0.34	0.08	0.00	0.67
C	71	0.13	0.00	0.00	0.17	3.83	4.39	2.50	8.00	1.75	1.00	0.00	4.50	0.09	0.00	0.00	0.00
D	131	0.14	0.00	0.00	0.03	1.64	1.13	0.11	4.96	5.33	7.00	1.00	15.42	0.01	0.00	0.00	0.00
F	187	0.19	0.00	0.00	0.25	3.49	4.00	1.50	8.00	0.48	0.00	0.00	1.00	0.00	0.00	0.00	0.00
H	311	0.25	0.12	0.00	0.36	5.08	5.38	3.12	8.65	2.40	2.63	0.83	5.00	0.02	0.00	0.00	0.00
J	119	0.46	0.25	0.00	0.75	3.81	5.00	1.75	8.50	0.20	0.00	0.00	0.00	0.01	0.00	0.00	0.00
K	306	0.68	0.50	0.12	1.25	6.74	7.50	4.50	10.42	1.25	1.00	0.17	2.88	0.06	0.00	0.00	0.00
L	334	0.71	0.52	0.24	1.25	7.24	8.00	5.32	11.00	1.09	0.95	0.11	2.18	0.01	0.00	0.00	0.00
M	116	0.48	0.37	0.04	0.93	6.43	7.43	3.72	9.64	5.57	6.35	3.78	8.99	0.20	0.00	0.00	0.18
N	164	1.03	0.94	0.48	1.76	6.88	7.19	4.71	11.32	4.43	5.02	2.70	7.28	0.15	0.00	0.00	0.02

		YEP				TLC				WTS			
CLASS	N	Mean	Median	Q25	Q75	Mean	Median	Q25	Q75	Mean	Median	Q25	Q75
A	107	5.05	4.87	1.83	12.75	0.03	0.00	0.00	0.00	6.27	7.39	2.28	14.40
B	85	4.75	5.00	0.91	11.05	0.56	0.00	0.00	0.68	2.53	3.04	0.51	5.60
C	71	6.37	6.77	2.30	15.96	0.002	0.00	0.00	0.00	2.84	3.25	0.86	7.25
D	131	15.84	16.39	7.78	38.67	0.00	0.00	0.00	0.00	0.94	0.17	0.00	2.33
F	187	2.89	2.00	0.00	9.17	0.00	0.00	0.00	0.00	0.64	0.33	0.00	1.25
H	311	12.61	13.40	5.14	28.90	0.08	0.00	0.00	0.00	1.35	0.83	0.17	3.22
J	119	2.46	1.00	0.00	7.75	0.07	0.00	0.00	0.00	0.31	0.00	0.00	0.25
K	306	5.72	5.20	1.60	14.19	0.45	0.00	0.00	0.35	0.96	0.67	0.08	2.17
L	334	2.80	1.97	0.29	7.28	0.31	0.00	0.00	0.00	0.51	0.29	0.00	1.00
M	116	20.50	22.40	11.15	41.87	1.53	1.44	0.01	4.07	2.11	2.27	1.16	3.84
N	164	5.622	5.65	1.51	15.48	0.64	0.13	0.00	1.30	1.09	1.05	0.34	1.84

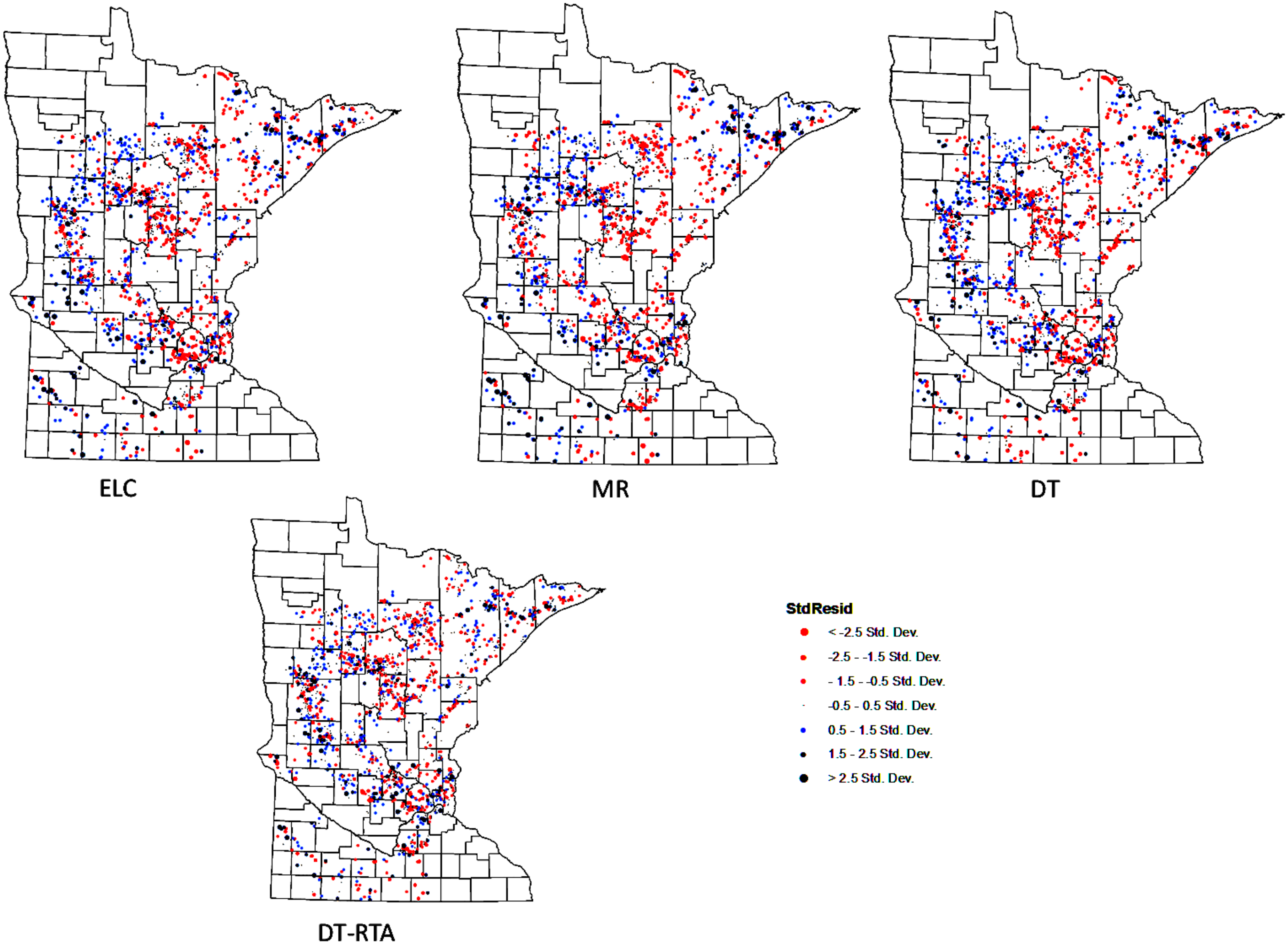
APPENDIX II continued.

Trapnet

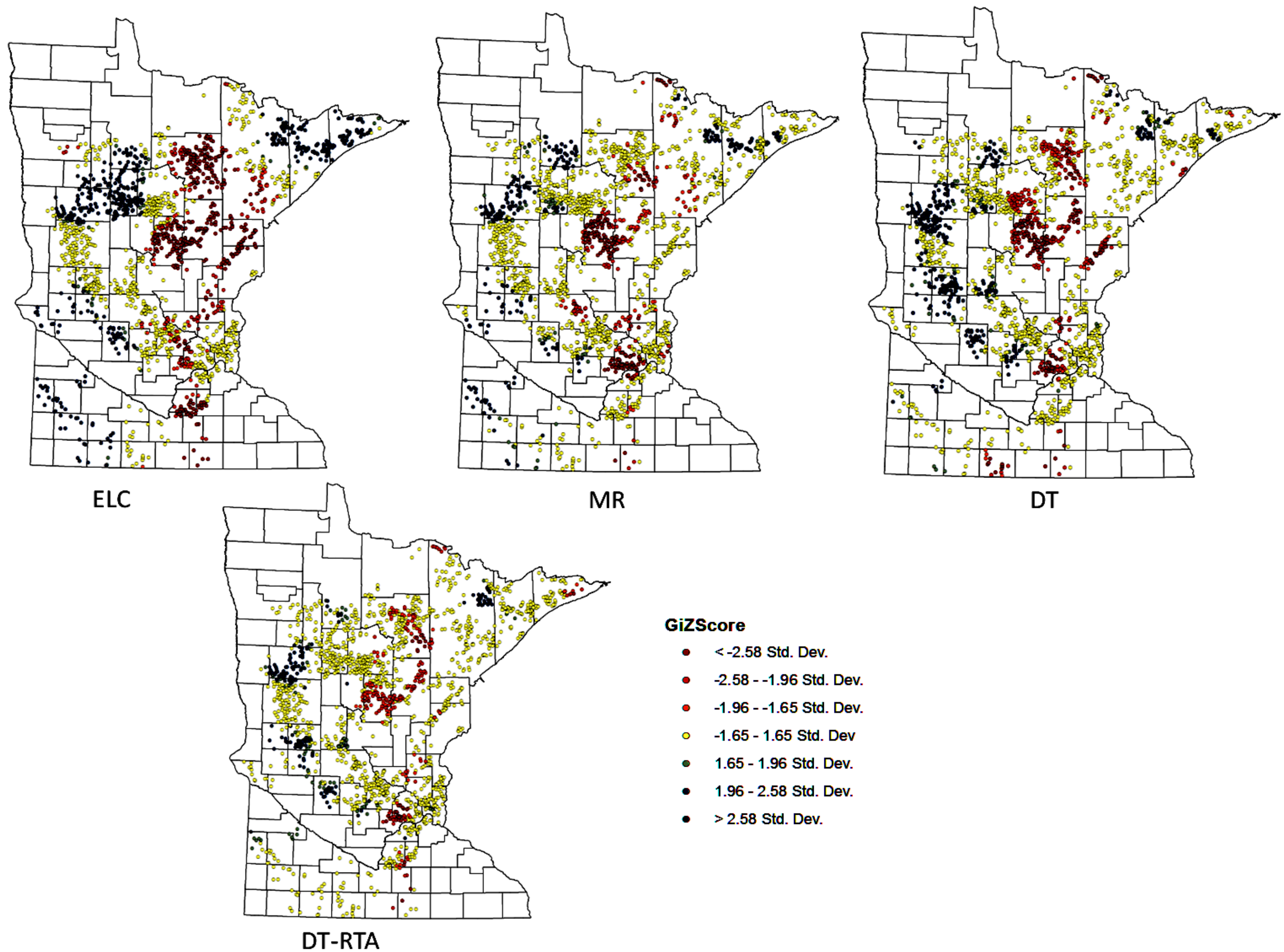
CLASS	N	BLC Mean	BLC Median	BLC Q25	BLC Q75	BLG Mean	BLG Median	BLG Q25	BLG Q75	GSF Mean	GSF Median	GSF Q25	GSF Q75	PMK Mean	PMK Median	PMK Q25	PMK Q75
A	107	0.07	0.00	0.00	0.00	0.28	0.00	0.00	0.09	0.02	0.00	0.00	0.00	0.23	0.00	0.00	0.03
B	85	1.04	0.56	0.00	2.99	2.66	1.62	0.00	9.54	0.03	0.00	0.00	0.00	0.37	0.00	0.00	0.57
C	71	1.05	0.67	0.00	2.74	2.66	2.29	0.00	12.72	0.01	0.00	0.00	0.00	1.12	0.84	0.00	2.74
D	131	5.38	5.55	0.63	17.13	2.20	0.80	0.02	5.19	0.27	0.03	0.00	0.20	0.16	0.00	0.00	0.02
F	187	4.60	4.67	1.50	9.46	14.07	16.50	6.11	39.50	0.22	0.00	0.00	0.11	1.30	1.22	0.17	2.67
H	311	4.45	3.94	1.61	9.55	15.08	18.75	6.79	36.30	0.11	0.00	0.00	0.07	1.64	1.42	0.38	3.29
J	119	1.45	1.06	0.00	3.11	10.12	15.67	1.67	33.00	0.11	0.00	0.00	0.00	1.86	1.67	0.00	4.75
K	306	1.28	1.06	0.44	2.22	16.62	18.19	8.79	39.17	0.06	0.00	0.00	0.00	2.36	2.12	0.93	4.44
L	334	1.50	1.17	0.56	2.71	27.64	29.37	15.66	52.68	0.22	0.00	0.00	0.22	3.03	2.83	1.50	5.55
M	116	0.75	0.58	0.23	1.04	10.34	11.95	4.49	23.76	0.05	0.00	0.00	0.00	2.71	2.72	1.53	3.98
N	164	1.18	1.00	0.51	1.81	26.72	30.45	17.96	41.81	0.12	0.03	0.00	0.15	2.86	2.90	1.76	4.39

CLASS	N	BLB Mean	BLB Median	BLB Q25	BLB Q75	YEB Mean	YEB Median	YEB Q25	YEB Q75	RKB Mean	RKB Median	RKB Q25	RKB Q75	CAP Mean	CAP Median	CAP Q25	CAP Q75
A	107	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B	85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.74	0.27	0.00	1.50	0.00	0.00	0.00	0.00
C	71	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.14	0.00	0.00	0.00	0.00
D	131	9.92	47.00	10.67	99.00	0.00	0.04	0.00	0.85	0.00	0.00	0.00	0.00	0.28	2.61	0.31	0.00
F	187	1.12	0.78	0.00	9.00	0.01	0.17	0.00	1.48	0.03	0.00	0.00	0.00	0.04	0.00	0.00	0.00
H	311	1.23	1.54	0.16	9.00	0.11	1.04	0.11	2.81	0.07	0.00	0.00	0.00	0.09	0.26	0.00	0.00
J	119	0.19	0.00	0.00	0.00	0.08	0.00	0.00	1.24	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00
K	306	0.10	0.00	0.00	0.00	0.08	0.11	0.00	1.33	0.64	0.43	0.00	1.17	0.00	0.00	0.00	0.00
L	334	0.19	0.10	0.00	0.00	0.37	2.75	0.68	6.11	0.24	0.00	0.00	0.25	0.00	0.00	0.00	0.00
M	116	0.02	0.00	0.00	0.00	0.04	0.39	0.05	1.60	1.64	1.44	0.59	3.12	0.00	0.00	0.00	0.00
N	164	0.06	0.11	0.02	0.00	0.15	2.45	0.97	4.44	0.90	0.78	0.07	1.81	0.00	0.02	0.00	0.00

APPENDIX III. Spatial distribution of residuals of from four models prediction walleye CPE in Minnesota lakes (Ecological Lake Classification-ELC, Multiple regression-MR, Decision Tree-DT, and Decision Tree with Regression Tree Analysis-DT-RTA). A high positive standard deviation indicates lakes with observed CPE much higher than that predicted by the model.



APPENDIX IV. Results of ArcMap hot spot analysis performed on residuals of four models (Ecological Lake Classification-ELC, Multiple regression-MR, Decision Tree-DT, and Decision Tree with Regression Tree Analysis-DT-RTA) using Getis-Ord Gi* statistic Z score For quantifying lakes differing from a geographic random distribution pattern. A high standard deviation indicates more intense clustering of lakes with high residual values.



STUDY ADDENDUM

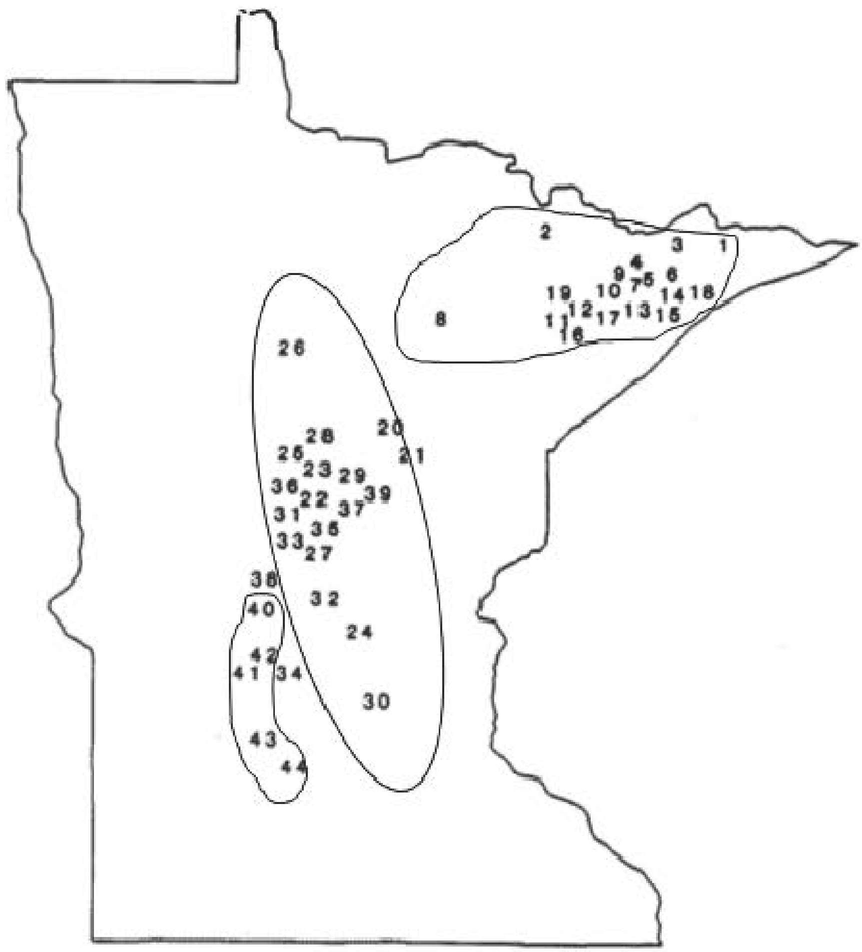
A geographical lake classification (7 classes) of Minnesota lakes derived from large-scale habitat classification (11 classes).

Analysis of large-scale factors that structure fish communities in lakes across Minnesota prompted a more intuitive geographically based lake classification developed as a tool for analyzing and communicating intrinsic habitat related differences in fish populations. There is considerable merit in technically defined classifications that separate lakes into larger numbers of highly defined lake classes; however, potential applications can also be confounded in these systems due to small class sizes and interclass differences that may not be relevant or significant to the application. For example, the effects of changing climate conditions generally differ geographically and a large number of lakes can be required to achieve the statistical power to overcome the high amount of natural and seasonal variation associated with detecting change in ecological indicators (Sondergaard et al. 2016). The concept for a geographic-based classification follows pioneering work by Moyle (1956, Figure A1), Schupp (1992, Figure A1) and others (Peterson 1974) that describe influence on fish populations in lakes using empirically derived geographical regions. This concept is still widely used by the Minnesota DNR, PCA, and other conservation organizations to characterize differences in fish communities among lakes. Modern advances in geographic descriptions of land and water characteristic and improved assessments of fish populations have provided a means to update this concept with a discrete classification using widely adopted ecoregion boundaries as related to fish communities.

APPROACH

We sought to simplify the 11-class large-scale habitat (LSH) classification of Minnesota lakes developed in this Investigational Report, while retaining most of the predictive/explanatory power describing differences in fish communities among lakes quantified with principle component analysis (PCA). The LSH classification was derived as a classification tree that could be easily simplified by pruning the lower branches explaining the least amount of fish community variation among lakes. We began with eliminating two terminal splits of lakes in the Northern Lakes and Forest ecoregion defined by latitude. Latitude explained only a minor amount of the overall variation in fish communities among lakes and likely served primarily as a proxy for water temperature. A similar rational was used to prune the terminal split of lakes in the LSH classification Canadian Shield and Superior Highlands ecoregion defined by air temperature. Finally, because the LSH classified Northern Lakes ecoregion had low membership and was spatially disjunct, the split was eliminated to effectively combine lakes located in EPA Level IV Toimi Drumlins ecoregion (50p) with the Canadian Shield and Superior Highlands ecoregion (EPA level IV 50n and 50t) to cover most of the northeast Minnesota arrowhead region. This modified ecoregion was termed Northeast Forest (Figure A2). Also, the few Northern Lakes ecoregion lakes located outside of the Toimi Drumlins were combined with LSH classified Northern Lake and Forest lakes, which were more aligned and in close proximity. This modified ecoregion was termed the Northcentral Forest (Figure A2). Together these changes resulted in a more intuitive and coherent spatial distribution with minimal loss in the amount of explained variation in fish communities. The three regions based on EPA level IV assignments and the final pruned geographic classification are depicted in Figures A2 and A3 respectively.

Using the set of data previously described for LSH analyses (this Investigational Report) we calculated descriptive statistics summarizing large-scale habitat variables and net catches for each geographical lake class. Median values were calculated for selected large-scale habitat variables and median values, along with quantiles, were calculated for fish species commonly assessed in gill and trap nets. Also, random forest modeling implemented in R (The R Development Core Team 2013) similar to that described for LSH analyses (this Investigational Report). For each geographic lake class we modeled gill net CPE of Northern Pike, Walleye, Yellow Perch, and Black Crappie and trap net CPE of Bluegill using 10 predictor habitat variables (lake surface area, maximum depth, shoreline development/geometry, alkalinity, phosphorus, Secchi, average July maximum air temperature, watershed area:lake area ratio, watershed disturbance, and baseflow). The percent of variation in fish catches explained by each model were used to assess model fitness and importance values (scaled average of prediction accuracy) for each predictor variable were used to identify variable influence in the model.



Approximate geographic centers of 44 lake classes.

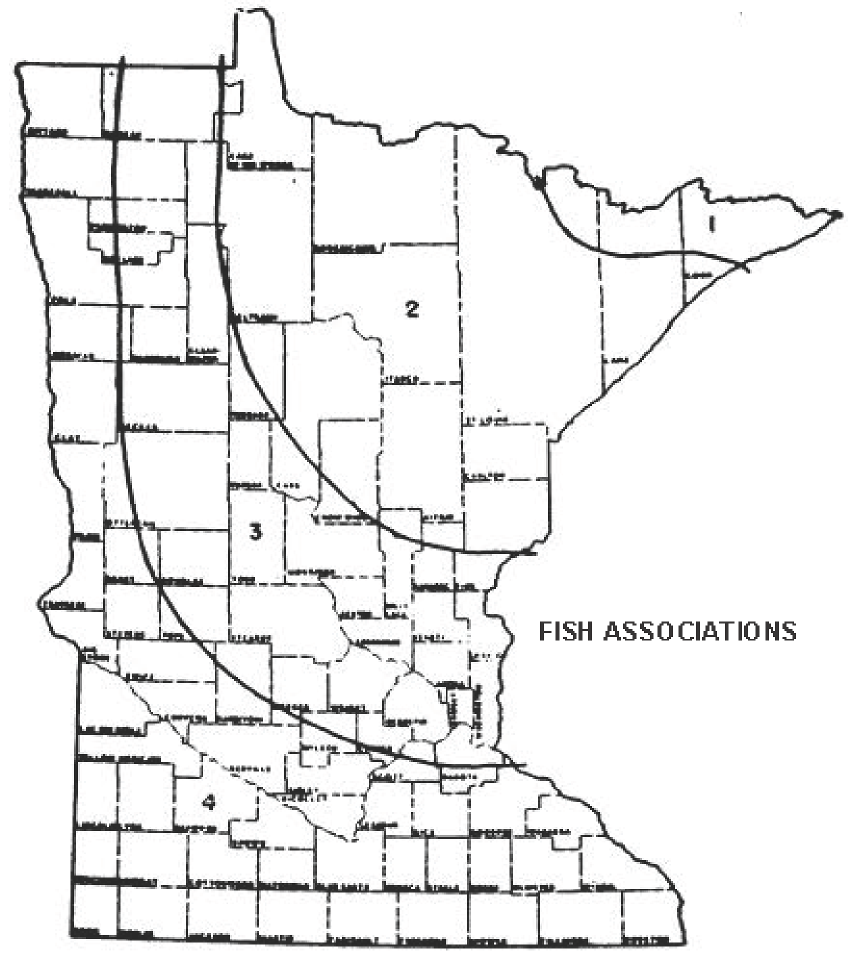


FIGURE A1. Maps characterizing the geographic distribution of lake types defined by fish associations (on right; Moyle, 1956) and physical-chemical criteria (on left; Schupp 1992).

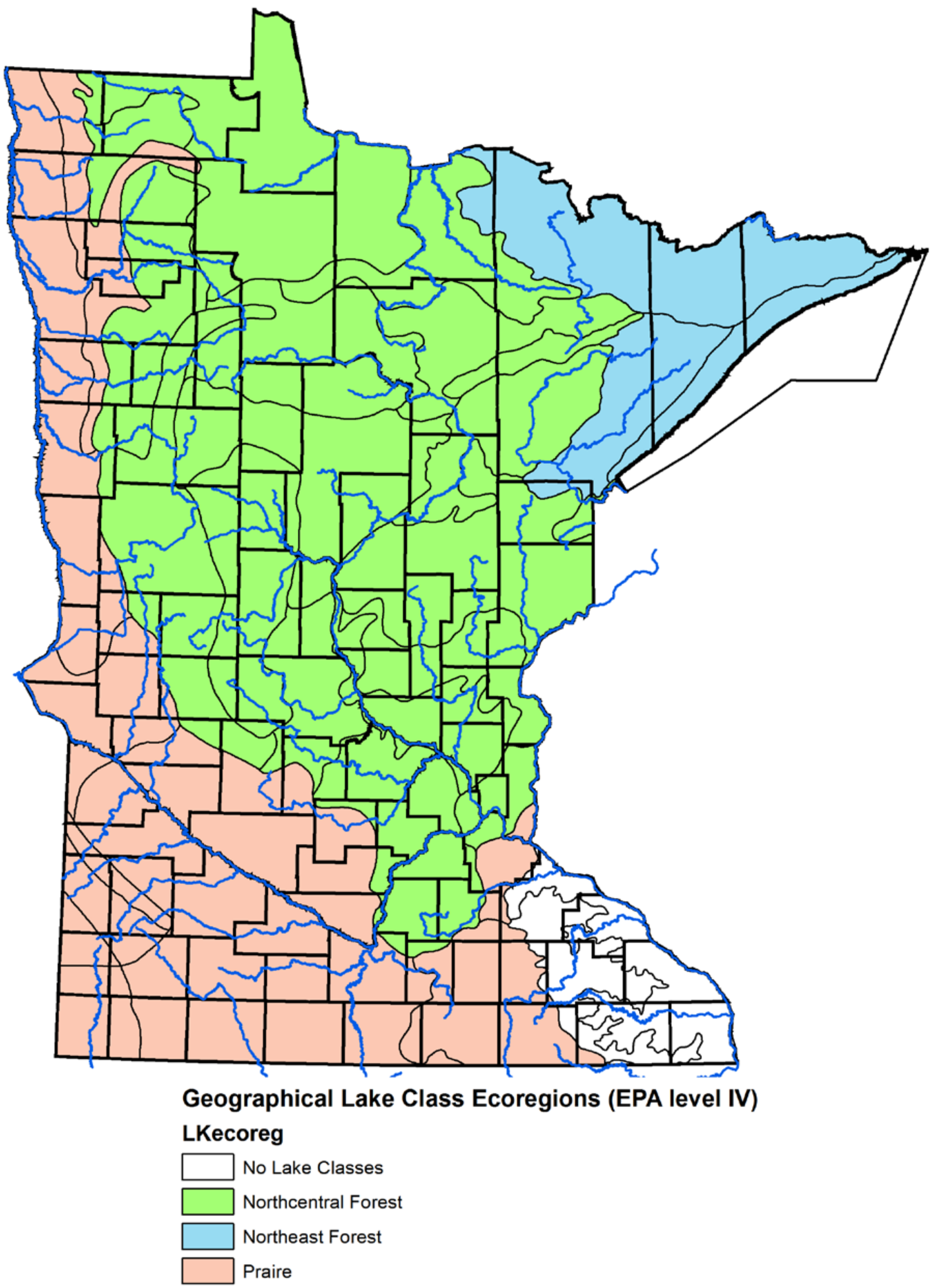


FIGURE A2. Map of lake ecoregions used to define geographic lake classes. Fine black lines show EPA level IV ecoregion boundaries, bold black lines are county boundaries, and blue lines are outlines of major rivers.

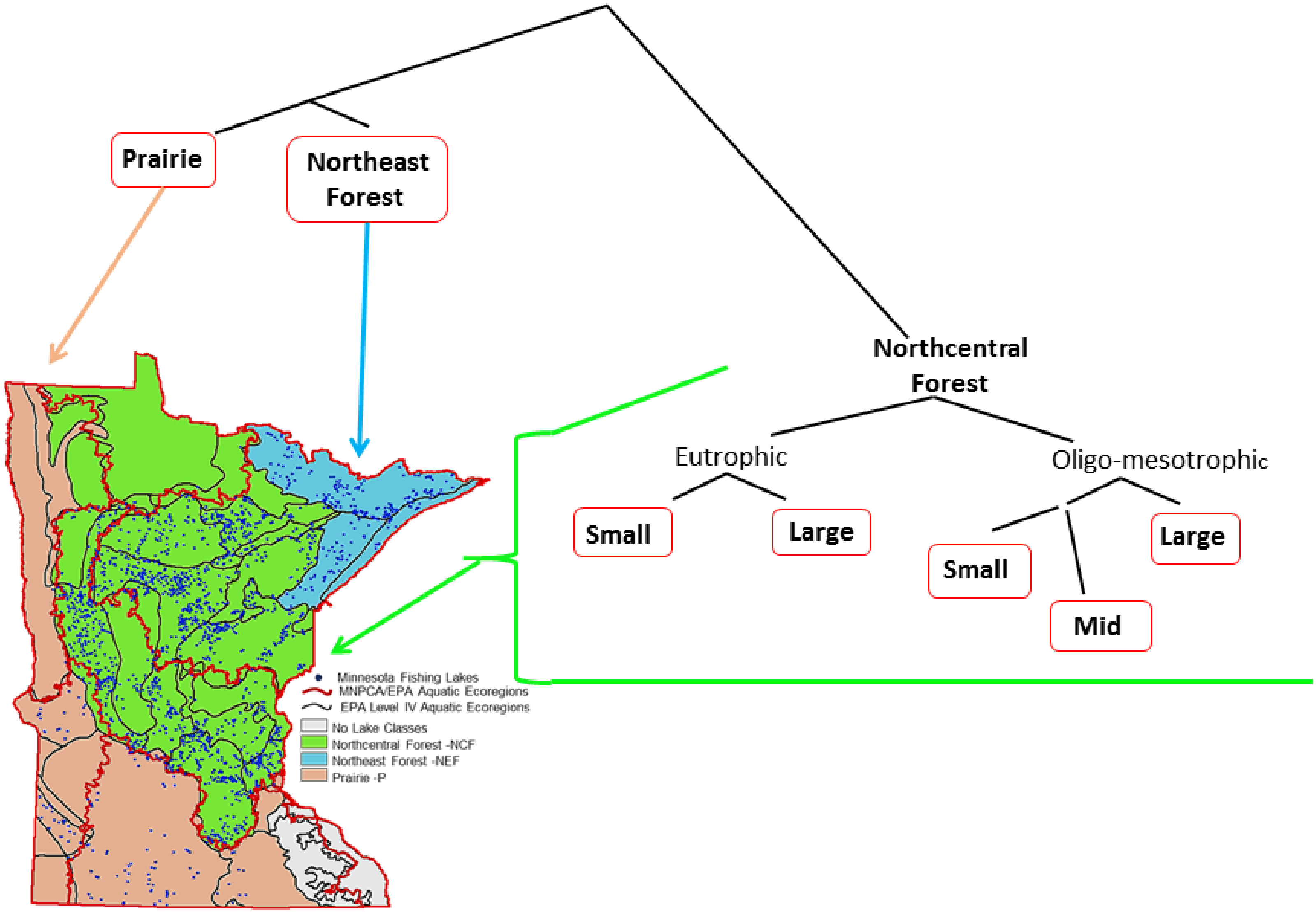


FIGURE A3. Classification tree used to define seven geographic lake classes of Minnesota lakes using lake ecoregion (from Figure 2), trophic status (Mesotrophic Secchi < 1.9m and Eutrophic $\geq$ 1.9m), and lake area (eutrophic - < and $\geq$ 150 acres; oligo-mesotrophic - < 63, 63-450, and $\geq$ 450 acres).

RESULTS AND DISCUSSION

Variation in both large-scale physical habitat and corresponding fish communities in Minnesota lakes follows a geographical gradient from southwest to northeast defined by classification tree analysis. A terminal node of lakes defined by the EPA IV prairie ecoregions (Prairie;P) located in southwest Minnesota are characteristically shallow, turbid, alkaline, and have watersheds that are highly agricultural (71% median row crop land cover in watershed; Table A1). Fish communities in these lakes have abundant Black Bullhead and Common Carp, but can also have high abundances of Walleye, Black Crappie, and Yellow Perch (Table A2). Another geographic lake class was defined at the opposite end of the geographic gradient with EPA level IV ecoregions confined to the Arrowhead region of northeast Minnesota (Northeast Forest). These lakes are often bog stained and characteristically very low alkalinity with a median air temperature (July average maximum) 3°C cooler than lakes in the Prairie class (Table A1). Lakes in the Northeast Forest class (NEF) also have watersheds with much less human disturbance and higher proportion of forested land cover than is typical of other Minnesota geographic lake classes (Table A1). Northeast Forest lakes characteristically contain fewer species and lower abundances with the exception of White Sucker (Table A2). Many lakes in the region contain coldwater fish species (notably Lake Trout). However, coldwater species had little influence on the classification due the low frequency of occurrence in the net types used in our analyses. Lakes in the Northeast Forest were also recognized as being distinct from other Minnesota lakes by Schupp (1993) who consequently classified them with a separate analysis. A similar approach would likely provide a greater number of defined lake classes in the Northeast Forest. A presence-absence analysis may be more appropriate than using PCA on continuously distributed CPE data for rarely occurring species.

The large, center of the geographic southwest to northeast Minnesota geographical lake gradient was identified by a group of EPA Level IV ecoregions and collectively termed Northcentral Forest. This geographically defined area contains a majority of Minnesota lakes, many of which rank as the most popular and highly used lakes in the state (Keeler et al. 2015). Corresponding to differences in fish assemblages assessed with PCA, five different geographical lake classes were identified in the Northcentral Forest by classification tree analysis. Northcentral Forest lakes were separated based on differences in surface area and Secchi summarized in Figure A4. The 1.9 m Secchi breakpoint corresponds with a Carlson Trophic State Indicator (TSI) value characterizing the break between mesotrophic and eutrophic conditions previously associated with changes in fish species composition in MNDNR fish surveys (Schupp and Wilson 1993). The three clear Northcentral Forest classes (Large Clear - LC, Medium Clear - MC, Small Clear - SC) have lakes located farther north, in areas of cooler air temperature, and have watersheds with less disturbance and higher forest cover (Table 1) than the two turbid Northcentral Forest lake classes (Large Turbid - LT, Small Turbid - ST). The three clear lake classes had fish communities with proportionately high abundances of bass, sunfish, Northern Pike, and Yellow Bullhead than the two turbid lake classes (Table A2). Finally, Northcentral Forest lakes classed with larger surface area contained proportionately more Yellow Perch and Walleye than the smaller lakes and also tended to have higher overall abundance of fish across more species (Table A2).

Because a PCA defined fish assemblage gradient was used to identify geographic lake classes it is logical for these classes to account for significant differences in the relative abundance (Catch per Effort; CPE) of individual fish species used in the PCA. The amount of variation in the relative abundance of most fish species (CPE) explained with the 7-class geographic classes was very similar to that explained by the 11-class LSHC and for most species this was comparable to variation explained with the 1993 Ecological Classification (this Investigational Report 562, Table 8). Species specific information on habitat relationships within each geographical lake class was demonstrated using random forest models with habitat variables to model CPE abundances for five selected common species of fish (YEP, WAE, BLC, BLG, and NOP). The percent of variation in abundance accounted by random forest models for each species and variable importance ranking of each habitat variable helped identify the relevance of factors controlling the abundance of each species within each geographic lake class.

TABLE A1. Number of lakes by geographic lake class and median values of selected large-scale habitat variables. N = number of lakes, SDI = shoreline development index, TA = total alkalinity.

	N	Depth (ft)	Area (acres)	Watershed: Lake Area	SDI	Forest	Ag	Disturbance	UTM northing	UTM easting	Secchi (m)	T.A. (mg/l)	Average July Max. Air Temp. (C)
LC	374	54	690.7	11.6	1.9	49%	9%	14%	5183025	379009	3.3	139	26.3
LT	339	23	340.4	16.1	1.7	19%	41%	52%	5058439	435044	1.4	125	27.7
MC	536	40	149.7	11.2	1.7	51%	8%	14%	5177440	408408	3.3	110	26.3
NEF	248	19	133.2	10.3	2.0	80%	0%	0%	5292061	603147	2.2	19	24.9
P	131	10	298.6	7.5	1.6	1%	64%	71%	4906701	331395	0.6	174	27.9
SC	80	37	35.0	12.1	1.4	57%	4%	13%	5174143	445146	3.1	76	26.2
ST	201	25	74.0	26.9	1.4	22%	12%	54%	5026008	468899	1.4	90	27.9

TABLE A2. Median, 25% quantile, and 75% quantile of MNDNR Lake Survey gill net catch per effort (CPE) for selected fish species in Minnesota lakes by geographic class.

GILL NET

Median	BLB	BLC	BLG	BOF	CAP	CCF	LMB	MUE	NOP	PMK	RKB	SMB	TLC	WAE	WHC	WTS	YEB	YEP
LC	0.07	0.81	3.67	0.00	0.00	0.00	0.75	0.00	7.47	1.25	0.97	0.00	0.21	4.68	0.00	1.33	2.06	9.13
LT	7.03	4.54	2.17	0.00	0.06	0.00	0.12	0.00	5.67	0.14	0.00	0.00	0.00	2.68	0.00	0.83	0.37	12.33
MC	0.04	1.00	3.00	0.00	0.00	0.00	0.50	0.00	7.75	0.33	0.00	0.00	0.00	0.50	0.00	0.33	0.50	3.00
NEF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.30	0.00	0.00	0.00	0.00	2.61	0.00	4.35	0.00	5.16
P	52.33	1.95	0.00	0.00	2.67	0.00	0.00	0.00	1.12	0.00	0.00	0.00	0.00	7.00	0.00	0.17	0.00	16.39
SC	0.00	0.33	1.37	0.00	0.00	0.00	0.00	0.00	4.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.70
ST	4.00	3.50	1.05	0.00	0.00	0.00	0.00	0.00	4.30	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	2.50
25% Quantile																		
LC	0.00	0.33	1.17	0.00	0.00	0.00	0.22	0.00	4.71	0.35	0.10	0.00	0.00	2.10	0.00	0.39	0.39	2.50
LT	0.33	1.72	0.46	0.00	0.00	0.00	0.00	0.00	3.07	0.00	0.00	0.00	0.00	0.83	0.00	0.17	0.00	4.70
MC	0.00	0.25	0.67	0.00	0.00	0.00	0.17	0.00	4.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50
NEF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.26	0.00	0.00	0.00	0.00	0.00	0.00	1.02	0.00	1.72
P	22.29	0.08	0.00	0.00	0.17	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	7.78
SC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ST	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	1.87	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
75% Quantile																		
LC	0.81	1.90	8.31	0.00	0.00	0.00	1.39	0.00	10.28	2.82	2.70	0.02	1.98	7.54	0.00	2.55	6.47	22.62
LT	36.82	12.90	7.65	0.00	0.75	0.00	0.38	0.00	8.72	0.72	0.00	0.00	0.00	4.97	0.00	3.11	2.00	28.14
MC	0.92	2.39	7.33	0.00	0.00	0.00	1.25	0.00	11.00	1.33	0.25	0.00	0.00	1.79	0.00	1.11	3.56	10.03
NEF	0.00	0.33	0.17	0.00	0.00	0.00	0.00	0.00	6.15	0.06	0.15	0.25	0.00	6.47	0.00	10.24	0.00	13.31
P	109.67	10.03	0.46	1.11	9.69	0.31	0.03	0.00	4.96	0.00	0.00	0.00	0.00	15.42	0.07	2.33	0.00	38.67
SC	2.40	2.00	4.31	0.00	0.00	0.00	0.67	0.00	7.73	0.59	0.00	0.00	0.00	0.00	0.00	0.25	0.17	8.28
ST	26.97	10.33	6.94	0.00	0.06	0.00	0.25	0.00	8.39	0.33	0.00	0.00	0.00	1.00	0.00	1.33	0.50	9.97

TABLE A2 continued on next page.

TABLE A2 continued.

TRAP NET

Median	BLB	BLC	BLG	BOF	BRB	CAP	CCF	LMB	NOP	PMK	RKB	SMB	WAE	WTS	YEB	YEP
LC	0.04	0.81	22.56	0.27	0.19	0.00	0.00	0.56	0.61	2.73	0.83	0.00	0.27	0.07	1.52	0.68
LT	1.22	3.62	18.50	0.27	0.17	0.22	0.00	0.18	0.55	1.42	0.00	0.00	0.25	0.11	1.02	0.71
MC	0.00	1.16	24.28	0.00	0.07	0.00	0.00	0.44	0.60	2.72	0.00	0.00	0.04	0.00	0.87	0.33
NEF	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.72	0.00	0.00	0.00	0.33	0.52	0.00	0.78
P	47.00	5.55	0.80	0.00	0.00	2.61	0.00	0.00	0.32	0.00	0.00	0.00	1.25	0.17	0.04	1.55
SC	0.00	0.89	9.41	0.00	0.00	0.00	0.00	0.22	0.22	1.11	0.00	0.00	0.00	0.00	0.00	0.15
ST	0.78	4.65	15.97	0.00	0.00	0.00	0.00	0.11	0.44	1.03	0.00	0.00	0.00	0.00	0.17	0.28
25% Quantile																
LC	0.00	0.41	11.44	0.00	0.04	0.00	0.00	0.23	0.34	1.59	0.22	0.00	0.12	0.00	0.33	0.22
LT	0.11	1.33	6.80	0.00	0.00	0.00	0.00	0.06	0.30	0.44	0.00	0.00	0.03	0.00	0.11	0.22
MC	0.00	0.44	12.33	0.00	0.00	0.00	0.00	0.12	0.33	1.11	0.00	0.00	0.00	0.00	0.00	0.08
NEF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.06	0.00	0.22
P	10.67	0.63	0.02	0.00	0.00	0.31	0.00	0.00	0.01	0.00	0.00	0.00	0.17	0.00	0.00	0.54
SC	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ST	0.00	1.43	6.18	0.00	0.00	0.00	0.00	0.00	0.17	0.17	0.00	0.00	0.00	0.00	0.00	0.00
75% Quantile																
LC	0.18	1.53	37.20	0.72	0.58	0.04	0.00	1.02	0.91	4.25	2.08	0.00	0.49	0.28	3.81	1.65
LT	11.03	8.55	35.47	0.80	0.80	1.11	0.00	0.40	0.91	3.32	0.00	0.00	0.56	0.48	2.89	1.85
MC	0.22	2.53	46.78	0.50	0.44	0.00	0.00	0.89	0.94	5.28	0.67	0.00	0.19	0.06	3.67	1.03
NEF	0.00	1.11	3.15	0.00	0.00	0.00	0.00	0.04	1.25	0.65	0.39	0.04	0.80	1.77	0.00	1.86
P	104.23	17.13	5.19	0.00	0.00	5.81	0.08	0.11	0.91	0.02	0.00	0.00	2.47	0.77	0.85	4.04
SC	1.65	2.83	30.05	0.00	0.11	0.00	0.00	0.67	0.67	4.71	0.00	0.00	0.00	0.06	0.71	1.57
ST	12.00	9.13	39.19	0.27	0.33	0.33	0.00	0.40	0.83	2.67	0.00	0.00	0.17	0.27	1.50	1.15

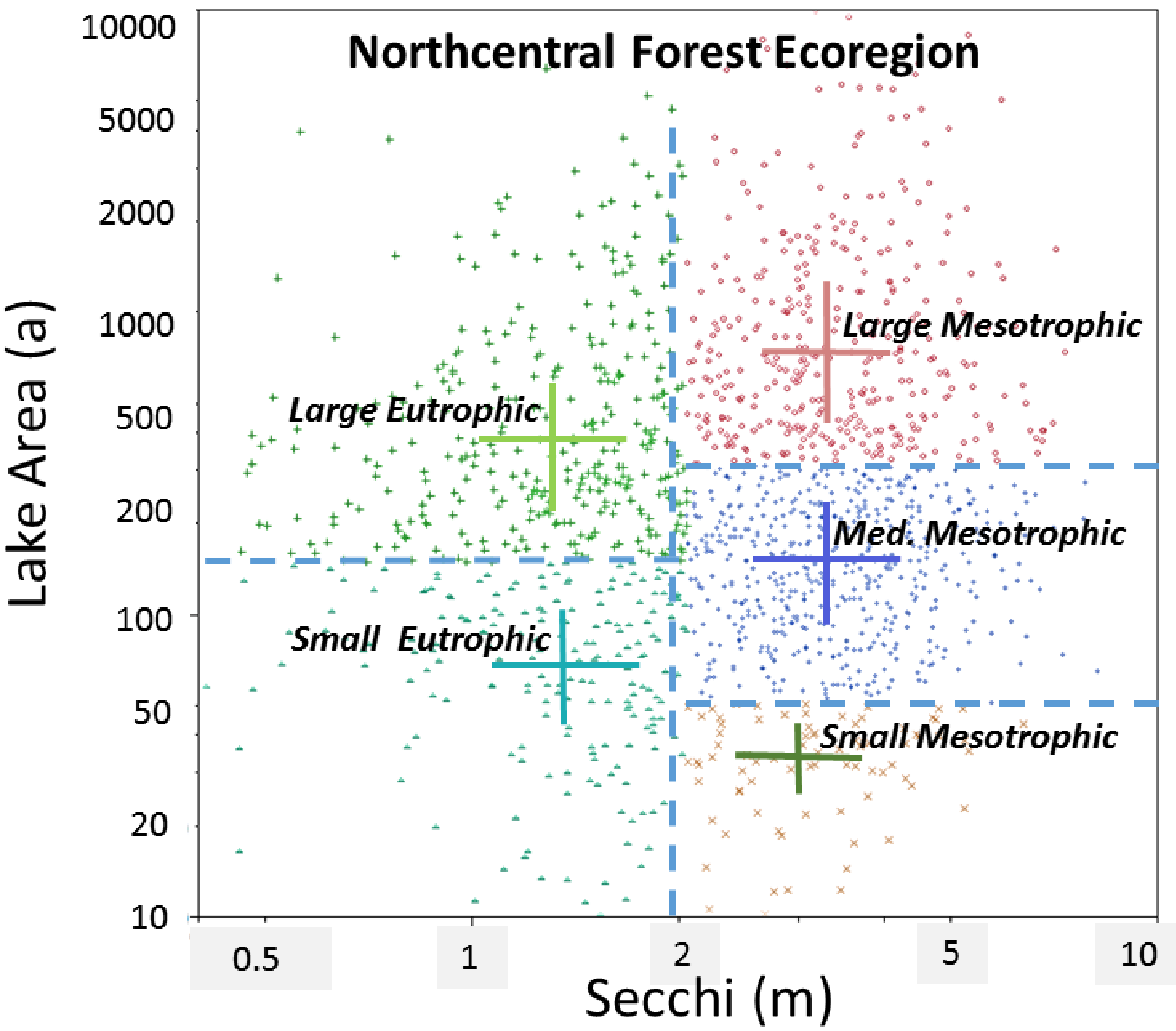


FIGURE A4. Distribution of lakes of the five Northcentral Forest Ecoregion geographic lake classes by lake area (acres) and Secchi depth (meters). The dashed lines indicate lake area and Secchi depth criteria defining membership in each of the five classes.

For NOP we found that large-scale habitat variables could account for little additional variation in CPE within geographic lake classes (Table A3). In contrast, a large amount of variation in Walleye CPE within geographic lake classes was explained with lake area, except in Northcentral Forest ecoregion lake classes defined with small lake area (SC and ST) and the Prairie lake class. Most Prairie lakes are heavily stocked with Walleye, which combined with sporadic winterkill, likely cloud many habitat associations. Variation in bluegill CPE within three Northcentral Forest ecoregion lake classes (LT, LC, and MC) and in the Northeast Forest lake classes primarily corresponded with air temperature differences. Similarly, air temperature was ranked with relatively high importance in random forest models of Black Crappie and Yellow Perch CPE in Northeast Forest lakes and large Northcentral Forest lakes. In addition, in large Northcentral Forest lakes, alkalinity was a factor of high importance for both Black Crappie and Yellow Perch while Secchi also relatively important in Black Crappie models and lake area in YEP models. Lake area was also a variable with high importance explaining variation in YEP abundance in Northeast Forest lakes, along with maximum lake depth. In summary, large-scale factors can account for variation in CPE of some fish species not explained by the geographic lake class, but it is significant to note that these habitat relationships are often not constant across lake classes. Consequently, knowledge of different species - habitat associations within the different geographic lake classes could increase efficiencies in research or management strategies. For example, Yellow Perch and Black Crappie management strategies might be changed based on alkalinity, temperature, and Secchi differences in LT and LC geographic lake classes.

Previous studies addressing habitat factors influencing fish populations in Minnesota lakes have identified strong relationships with lake depth, alkalinity, and temperature (Moyle 1956; Peterson 1974; Schupp 1992; Stefan et al. 1996; and Valley et al. 2004). In general, these variables are strongly associated with "geographic" habitat variables describing variation in MN lake fish communities (ecoregion, lake area, and Secchi). These geographical differences are most obvious between lakes in the Northeast Forest and Prairie ecoregions that strongly correspond to vast differences in alkalinity, temperature, and depth (Moyle 1956, Heiskary and Wilson 2008). Also, differences among Northcentral Forest lake classes defined by Secchi, correspond closely with depth (deeper lakes with deeper Secchi) and temperature (cooler lakes with deeper Secchi). Strong associations between Secchi and depth, land cover, and geo-climatic factors are well documented by Johnston and Shmagin (2006). Although geographically defined lake classes were effective at explaining general differences in fish communities found across Minnesota lakes, the influence of specific large-scale habitat factors on the relative abundance of individual fish species was also important (e.g. Bluegill abundance strongly associated with temperature differences within Northeast Forest and Northcentral Forest large and medium size lakes). While considerable variation in fish community composition can be characterized at a broad geographic scale, any attempt to understand the abundance of individual species must account for more specific habitat influences and interactions with other fish species. The geographic lake classification was based on observational data describing fish communities, so cause and effect habitat relationships cannot be explicitly identified. However, causal inferences can be made with increased confidence when knowledge of geographic lake class differences is combined with other studies (e.g. Jacobson et al. 2017).

In summary, seven geographically defined lake classes explain variation in fish communities more parsimoniously than the 11 classes resulting from previous analyses of key large-scale habitat influences on Minnesota lake fish communities. A distinct advantage of the 7-class geographical scheme is that its descriptor variables are commonly used and widely available to both technical and nontechnical audiences. All three descriptors, location (ecoregion), size, and trophic status (Secchi) are featured in datasets marketed to the public at MNDNR LakeFinder. Although a large amount of variation among lakes is explained with 7 lake classes, relationships expressed on a general geographical bases still contain exceptions as stated by Moyle (1956). Nonetheless, geographic lake classes document a quantitative breakdown of lakes characterizing associations between habitat and fish assemblages. Consequently, geographic classes provide a template for examining differences between lakes relating to perturbation or management activities on a more comparative basis. In essence, because geographical lake classes are based on fish assemblage similarities they are well suited for use as stated by Schupp (1992) for ecological lake classes: "Analyzing results from a holistic viewpoint rather than from the single species approach so common until now should lead to management recommendations that have a higher probability of success."

TABLE A3. Variable importance and percent of variation explained in random forest models using ten large-scale habitat variables to explain variation in CPE of five selected fish species by geographic lake class.

	LC	MC	SC	LT	ST	P	NEF
Yellow Perch							
Lake Area	22.4	2.8	1	9.1	7.9	0.3	20
Maximum Depth	9.7	15.3	0.9	7.4	4.1	0.5	27.4
Shoreline Development	6.2	-0.4	5	7.5	6	-4	1.1
Alkalinity	21.6	23.6	3.9	15.1	5.6	0	8.2
Phosphorus	21.3	3.7	3.2	6.5	1.8	-1.3	2
Secchi	12.1	8.8	0.6	5.6	2.7	-0.9	12.1
Air Temperature	29.3	25.3	11	9.1	12.6	5.4	3.3
Watershed Area	3.4	1.1	-1.2	5.2	-1.9	3.9	-2.7
Watershed Disturbance	11.6	11.9	3.7	7.4	8.1	5.7	0.8
Baseflow	6.4	17.2	4.3	1.6	-1.8	2.4	6.8
% Variation	34.4	19	8.1	13.1	7.7	-10.3	22.9
Bluegill							
Lake Area	9.2	2.2	2.2	4.5	10.7	-0.3	12.7
Maximum Depth	5.6	5	-0.2	13.1	-0.2	13.5	3.6
Shoreline Development	2	-0.3	-1.7	3.2	2.4	-0.4	7.9
Alkalinity	13.2	9	2.2	12.2	9.2	1.2	13.1
Phosphorus	15.8	4.6	-0.4	14	0.4	6.2	5.5
Secchi	8.3	8.9	-4.6	9.4	-0.1	9.4	3.2
Air Temperature	30.2	27.7	-2	33.1	12.3	5	22.7
Watershed Area	9.1	9.5	-3.5	13	3.5	3	2.5
Watershed Disturbance	11.7	12.9	2.7	16.6	17.4	-1.6	11.1
Baseflow	4.2	13.3	-1.5	6.2	1.6	20.3	21.3
% Variation	24.4	18	-11.7	37.3	21.4	30.7	34.4
Black Crappie							
Lake Area	4.8						
Maximum Depth	16	7.9	-0.8	8.1	-1.1	-1.8	6.2
Shoreline Development	2.1	4.5	1.1	2.2	-0.1	2.6	5.7
Alkalinity	18.2	16.7	4	20.4	4.3	0.2	4
Phosphorus	4.8	2	4.6	1	10	0.1	-0.3
Secchi	14.8	13	-3.8	16.8	5.6	0.6	3.9
Air Temperature	14.8	9	1.4	29.2	4.5	-1.5	10.4
Watershed Area	4.8	3	-3	11.1	4.2	-3.2	-1.4
Watershed Disturbance	11.4	5.2	1.5	12.5	7.5	1.2	9.1
Baseflow	10.8	6.6	-2.3	5.8	2.1	7.3	9.6
% Variation	26.9	10	-14.1	36.1	6.8	-10.6	10.1

TABLE A3 continued.

	LC	MC	SC	LT	ST	P	NEF
			Walleye				
Lake Area	34.5	24.4	-2.1	28.5	9.5	7.9	34.6
Maximum Depth	10.2	-0.6	3.4	-1.1	-0.2	4.7	2.7
Shoreline Development	13.5	16.9	-1.2	4.4	1.6	-3.2	8
Alkalinity	23.3	13.1	2.7	9.9	3.1	0.4	3.2
Phosphorus	7	6.7	0.4	5	1.7	4	9.6
Secchi	8.1	5	-0.4	-0.9	-1.2	2.3	1
Air Temperature	13	13.5	10.8	3.4	7.7	-0.5	4.6
Watershed Area	1.4	5.4	1.4	2.8	-4	1	1.3
Watershed Disturbance	11.2	3.8	-1.1	7.3	5.6	3	6.3
Baseflow	6.2	8.7	2.1	3	0.8	0.9	3.8
% Variation	33.7	19	4.8	14.4	-1.5	2.3	29.2
			Northern Pike				
Lake Area	7.3	3.8	0.1	-1.1	24	1	4.2
Maximum Depth	12.8	8.8	-3.1	5.4	-0.6	-0.8	10.4
Shoreline Development	3.1	-2.1	3.1	2.4	0.9	2	0.7
Alkalinity	13.9	4.8	9.6	5.3	1.4	-2.3	11.2
Phosphorus	1.5	3.2	-0.9	10.9	0.1	0	2.3
Secchi	2.5	1.5	3	4.8	0.3	-2.2	9.9
Air Temperature	11.3	4.9	13.5	6.7	3.2	-1.1	0
Watershed Area	4.4	3.3	4.8	1.2	2.1	13.1	2.7
Watershed Disturbance	5.1	5.7	2.1	7.2	5.2	-0.3	2.3
Baseflow	10.7	7.8	1.9	5.3	-6	5.4	7.5
% Variation	8.5	1.8	13.7	1.2	10.9	1.2	8.5

REFERENCES

Heiskary, S. A., and C. B. Wilson. 2008. Minnesota's approach to lake nutrient criteria development. Lake and Reservoir Management 24:282-297.

Jacobson, P. C., G. J. A. Hansen, B. J. Bethke, and T. K. Cross. 2017. Disentangling the effects of a century of eutrophication and climate warming on freshwater lake fish assemblages. PLoS ONE 12(8):e0182667. (https:/doi.org/101371/jounal.pone.0182667)

Johnston, C. A., and B. A. Shmagin. 2006. Scale issues in lake-watershed interactions: assessing shoreline development impacts on water clarity. Pages 297-313 In J. Wu, K. B. Jones, H. Li, and O. L. Loucks (eds) Scaling and uncertainty analysis in ecology: methods and applications. Springer. Netherlands.

Keeler, B. L., S. A. Wood, S. Polasky, C. Kling, C. T. Filstrop, and J. A. Downing. 2015. Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. Frontiers in Ecology and the Environment 13(2):76-81.

Moyle, J. B. 1956. Relationships between the chemistry of Minnesota surface waters and wildlife management. Journal of Wildlife Management 20:303-320.

Peterson, A. R. 1974. Distribution of the larger fishes in Minnesota lakes, 1948-1967. Minnesota Department of Natural Resources, Section of Fisheries Special Publication Number 107, St. Paul.

Schupp, D. H. 1992. An ecological classification of Minnesota lakes with associated fish communities. Minnesota Department of Natural Resources, Section of Fisheries Investigational Report Number 417, St. Paul.

Schupp, D., and B. Wilson. 1993. Developing lake goals for water quality and fisheries. Lakeline 13:18-21.

Sondergaard M., S. E. Larsen, L. S. Johansson, T. L. Lauridsen, and E. Jeppersen. 2016. Ecological classification of lakes: uncertainty and the influence of year to year variability. Ecological Indicators 61:248-257.

Stefan, H. G, M. Hondzo, X. Fang, J. G. Eaton, and J. H. McCormick. 1996. Simulated long-term temperature and dissolved oxygen characteristics of lakes in the north-central United States and associated fish habitat limits. Limnology and Oceanography 41:1124-1135.

The R Development Core Team. 2013. R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria. Available from http://www.R-project.org/

Valley, R. D., T. K. Cross, and P. Radomski. 2006. The role of submersed vegetation as habitat for fish in Minnesota lakes, including the implications of non-native plant invasions and their management. Minnesota Department of Natural Resources, Section of Fisheries Special Publication Number 160, St. Paul.