Fisheries and Oceans Canada / Pêches et Océans Canada
Science / Sciences

C S A S / **S C C S**

Canadian Science Advisory Secretariat / **Secrétariat canadien de consultation scientifique**

Research Document 2004/087
Document de recherche 2004/087

Not to be cited without
permission of the authors *

Ne pas citer sans
autorisation des auteurs *

Coastal exposure as a first-order predictor of the productive capacity of near shore habitat in the Great Lakes

Le fetch comme variable explicative de premier ordre de la capacité de production d'habitats littoraux des Grands Lacs

R.G. Randall, C.K. Minns and C.M. Brousseau

Fisheries and Oceans Canada
Great Lakes Laboratory for Fisheries and Aquatic Sciences
P.O. Box 5050, 867 Lakeshore Road
Burlington, Ontario L7R 4A6

ISSN 1499-3848 (Printed / Imprimé)

Canada

Abstract

Regression tree analysis with coastal exposure (fetch distance) as a predictor of fish biomass was used to evaluate the productive capacity of near shore habitat in the Great Lakes. Regression tree models were developed using survey data collected at coastal wetlands, harbours and natural shorelines in 1994 (n=100) and validated using data from other areas surveyed in other years (1990 to 1999, n=273). Coastal habitat characteristics that influence fish distribution, including the occurrence and abundance of aquatic macrophytes, water temperature and substrate characteristics, were related to maximum fetch distance in a consistent manner in the model and validation data sets. Three classes of macrophyte density (absent, moderate and dense cover), were predicted from substrate size and fetch distance: plant cover was highest where the predominant particle size was fine (silt or smaller) and maximum fetch was < 12.6 km. Fetch was a significant predictor of the biomass of three species (*Lepomis gibbosus, Perca flavescens, and Alosa pseudoharengus*), each with different habitat preferences, and two fish community indices (Index of Biotic Integrity [IBI], and the Habitat Productivity Index [HPI]). IBI and HPI were used as measures of the diversity and production components of habitat productive capacity, respectively. For all fish response variables, classification was improved if fetch was used together with associated habitat attributes as predictors. The degree of resolution of habitat classification (number of classes that were distinct) was limited to 2 to 4 classes, depending on the fish response variable. Proportional reduction in error for the regression trees ranged between 0.30 and 0.76. Four classes of *L. gibbosus* habitat were determined and validated, but the number of habitat classes for *P. flavescens* and *A. pseudoharengus* was less. For the whole fish assemblage, four habitat classes were identified using IBI and HPI together in a two-axes approach for evaluating productive capacity, along with fetch and water temperature as predictors. Knowledge of site exposure and the associated habitat covariates can be used to determine and map first-order estimates of coastal habitat productive capacity in the Great Lakes.

Résumé

Pour évaluer la capacité de production des habitats littoraux des Grands Lacs, nous nous sommes servis de la technique d'arbre de régression avec l'exposition du littoral au vent (fetch) comme variable explicative de la biomasse des poissons. Nous avons élaboré des modèles d'arbres de régression à l'aide de données recueillies en 1994 dans des terres humides littorales, des zones portuaires et le long de berges naturelles (n = 100), et nous avons validé ces modèles au moyen de données recueillies dans d'autres secteurs ayant fait l'objet de relevés durant d'autres années (de 1990 à 1999, n = 273). Dans le modèle et les jeux de données de validation, les caractéristiques des habitats littoraux qui influent sur la répartition des poissons, notamment la présence et l'abondance de macrophytes aquatiques, la température de l'eau et les caractéristiques du substrat, présentaient une relation systématique avec le fetch maximal. Trois classes de densité des macrophytes (couvert absent, modéré et dense) ont été prédites à partir de la taille du substrat et du fetch : le couvert végétal était le plus dense là où le substrat était dominé par des particules fines (limon ou plus petit) et le fetch maximal était inférieur à 12,6 km. Le fetch a permis de prédire significativement la biomasse de trois espèces (*Lepomis gibbosus, Perca flavescens et Alosa pseudoharengus*), chacune ayant différentes préférences en matière d'habitat, ainsi que deux indices de communauté de poissons, soit l'indice d'intégrité biotique (IIB) et l'indice de productivité de l'habitat (IPH). L'IIB et l'IPH ont respectivement servi de mesures des composantes diversité et production de la capacité de production de l'habitat. Pour toutes les variables dépendantes concernant les poissons, la classification était meilleure lorsque les variables indépendantes comprenaient des attributs de l'habitat en plus du fetch. Selon la variable dépendante concernant les poissons, la résolution de la classification des habitats (nombre de classes distinctes) n'était que de 2 à 4 classes. La réduction proportionnelle de l'erreur associée aux arbres de régression variait de 0,30 à 0,76. Nous avons déterminé et validé quatre classes d'habitats pour *L. gibbosus*, mais moins pour *P. flavescens* et *A. pseudoharengus*. Pour tout l'assemblage de poissons, nous avons établi quatre classes d'habitats en utilisant à la fois l'IBI et l'IPH dans une méthode à deux axes pour évaluer la capacité de production, ainsi que le fetch et la température de l'eau comme variables explicatives. La connaissance de l'exposition du site et des covariables associées décrivant l'habitat peut servir à déterminer et à cartographier des estimations de premier ordre de la capacité de production des habitats littoraux des Grands Lacs.

Introduction

Coastal exposure refers to the extent that a shoreline habitat and its fauna are subjected to the physical stress of wave energy. Exposure and morphometry affects water retention time and ecological function in coastal marine areas (Håkanson et al 1986). In large lakes, exposure influences the nature of the near shore habitat, affecting aquatic plant growth, substrate characteristics and water temperature. Wave energy affects the occurrence and density of aquatic macrophytes both directly by the physical action of waves and indirectly by its influence on the sediment characteristics (Chambers 1987). Sediment transport and deposition based on wave energy and water currents were modeled in the Great Lakes by W.F. Baird and Associates (1996). Thermal habitat in the nearshore will also be linked to wind generated currents and to offshore-onshore wave action. Although the associations are complex, coastal exposure is a proxy for these thermal and structural habitat attributes, all of which affect the occurrence and distribution of fishes that use shallow coastal water during all or part of their life history.

Fetch distance as a measure of coastal exposure has been shown to be a significant predictor of physical habitat conditions and fish abundance in the Great Lakes (Randall et al. 1996; Randall et al. 1998). For whole fish assemblages, fish abundance decreased but average fish size increased with increases in exposure (Randall et al. 1996). Fetch, together with related habitat attributes was a significant predictor of the occurrence of individual species in Severn Sound, Georgian Bay (Randall et al. 1998). The spatial resolution for discerning fish-habitat linkages was limited to two or three habitat classes (i.e., the power of resolution of classes that could be shown to be statistically different was 2 or 3). Nevertheless, Minns and Moore (2003) showed that although fish habitat associations are often uncertain, robust management decisions can be made with limited habitat classifications.

The specific objectives of this study were to use a large near shore fish-habitat database to: 1) quantify and model the relationship between site exposure (measured as fetch distance) and the habitat features of surface water temperature, macrophyte abundance, and substrate size; 2) develop and validate empirical models to assess habitat productive capacity, and to compare the efficacy of fetch alone and together with other habitat attributes as predictors of habitat productive capacity; and 3) compare the accuracy of predicting individual species biomass versus fish assemblage indices of productive capacity. We define habitat productive capacity as the sum of production of all co-habiting species in a particular area, with the additional qualifier that the fish community should reflect the natural biodiversity for that area (Minns 1995; Minns 1997; Randall 2003). Productive capacity is measured by linking biological indices of fish production and diversity (biomass, P/B, species composition and richness) to habitat surrogates or classes with different capacities (Randall 2003). The general goal was to determine the spatial resolution of habitat classes that was achievable using regression tree analysis. The expectation was that, although limited in number, habitat classes would be spatially robust, and that the relationship between habitat features and fish abundance is likely more evident for individual species than for fish assemblages. The regression tree models will be used in future to determine and map first order estimates of coastal habitat productive capacity from knowledge of site exposure and the habitat covariates.

Methods

***Database*:** Data collected during electrofishing surveys from 1990 to 1999 were used to determine associations between shallow coastal habitat and fish catches. Data were collected from three Great Lakes' Areas of Concern (Hamilton Harbour, Bay of Quinte and Severn Sound; Minns et al. 1994; Randall et al. 1996), coastal habitats in the vicinity of harbours and coastal wetlands in lakes Erie and Ontario (Randall and Minns 2002), and from Prince Edward County, Lake Ontario (Randall; ongoing project). During this time period, a total of 373 transects were surveyed on a seasonal basis (Fig. 1 and Table 1). As explained below (Statistical Methods), the dataset was divided into two subsets, where the first subset was used for model building, and a second subset was used for validation. The number of surveys at each transect varied from 2 to 4 per year, usually at monthly intervals beginning in early summer and ending in early autumn. Survey objectives differed over the years but a standard and consistent field protocol was used to collect the fish and habitat data for all surveys (below). Although data were collected during a number of years from Hamilton Harbour as part of an ongoing monitoring project (Smokorowski et al. 1998), only two years of data were used for this study (1990 and 1997). Species richness was lower at Hamilton than elsewhere, probably because of degraded habitat and water quality at this Great Lakes' Area of Concern (Minns et al. 1994; Randall et al. 1993). A remedial action plan was initiated in Hamilton Harbour in 1992 to restore habitat at six shoreline areas (Smokorowski et al. 1998). Two years of data were used in this study to measure fish-habitat associations in the Harbour before (1990) and after (1997) habitat restoration.

***Fish data*:** Fish were captured by boat electrofishing along 100 metre line transects at 1.5 metre water depth. Electrofishing settings were: pulsed DC, 120 pps, and an output of about 8 amperes. Information was recorded on species richness (number of species per transect), and the number of fish and total biomass (g wet weight). For each transect sample, each fish was identified and measured (weight in g, fork length in mm) up to a total of 20 specimens per species. If the catch for an individual species exceeded 20, remaining fish were counted and weighed in batches. Further details of the survey and sampling protocol are given by Valere (1996).

For this study, fish-habitat associations were examined for both individual species and for fish assemblages. Catches (biomass) of three species (*Perca flavescens*, *Lepomis gibbosus*, and *Alosa pseudoharengus*) were used separately in the analysis; these three species were the most abundant and ubiquitous species captured at the coastal habitats (Randall et al. 1996 and 1998) and they had contrasting physical and thermal habitat preferences (Table 2). For the whole fish assemblage, both an Index of Biotic Integrity (IBI) and a Habitat Productivity Index (HPI) were used as multi-metric measures of composition and size structure (Randall and Minns 2002). An IBI score (Minns et al. 1994) was calculated for each electrofishing sample for each date, using the software described by Stoneman (1998). IBI scores were based on three categories and 12 fish metrics. The categories and metrics were: species richness – 5 metrics, including numbers of native, centrarchid, intolerant, nonindigenous and cyprinid species; trophic structure - three metrics, including percent (biomass) of species which were piscivores, generalists and specialists; and abundance and condition – 4 metrics, including the number of native individual fish, biomass of native fish, the percent nonindigenous numbers and the percent nonindigenous biomass. Intolerant species were species that were intolerant of high turbidity (Minns et al. 1994). Piscivore species consumed fish prey as adults, generalists had multitrophic, adaptable diets (omnivores), and specialists had specialized diets (insectivore, planktivore). IBI metrics for each species are listed in Minns et al. (1994). Four of the 12 metrics negatively affected the final IBI score (3 metrics dealing with nonidigenous species and the percent generalists), while the remaining 8 metrics were positively related to the final score. IBI metrics were standardized and summed to produce an IBI score that ranged between 0 and 100. Further details of the individual metrics and the algorithms used to determine the IBI value are provided by Minns et al. (1994). A final annual transect IBI was calculated as the arithmetic mean of the individual IBI scores for each transect at each location for each year (number of samples per year ranged from 2 to 4; Table 1).

Fish data from seasonal samples per transect (n = 2 to 4) were pooled to calculate a Habitat Productivity Index (HPI). For each sample, biomass density was estimated by assuming a survey width of 10 m, and a capture efficiency of 0.3 (Randall et al. 1993). For each species, an average biomass (kg ha^{-1}) was calculated for the 2-4 seasonal samples. A production index for each species was calculated as the product of the average biomass ($B_{average}$) and a species P/B ratio (y^{-1}) (Randall and Minns 2000). The species P/B ratio was calculated as $P/B = 2.64\ W^{-0.35}$ (Randall and Minns 2000), where W was the average weight (g) of each species captured at the transect (total biomass/ total number of fish). A final HPI (kg ha^{-1} y^{-1}) for each transect was calculated as the sum of the production indices of all species captured at the transect (Randall and Minns 2000).

In the Great Lakes dataset, HPI was strongly correlated with fish biomass, and IBI was correlated with species richness (Fig. 2). HPI and IBI were weakly but significantly correlated ($P<0.01$).

Habitat: Water temperature (^{o}C) was recorded at the time of survey (mid transect at 0.5 m depth). For each transect, an average seasonal water temperature was calculated (n = 2 to 4).

Transect locations were recorded using a Global Positioning System (GPS). Using GIS generated maps of the lake shoreline and the transect locations, effective fetch was calculated for each of 16 compass directions as: (Sum(Cos(*a*)*Fetch(*a*))/(Sum(Cos(*a*)) for a range of angles (*a*) +-45° of the compass direction (Scheffer et al. 1992). Fetch was the distance from the transect to the shoreline at angle *a* (km). The maximum effective fetch value (km), the maximum of the 16 effective fetch values, was used to measure site exposure. Maximum effective fetch rather than effective fetch in the direction of prevailing wind was used because single wind events from any direction can influence site conditions (W.F. Baird and Associates 1996).

Macrophyte density was measured as percent bottom cover during the year of survey. Visual estimates (Minns et al. 1993) or echogram records (Randall et al. 1998; Randall and Minns 2002) were used to estimate bottom cover. Observations of macrophyte cover were made in late summer or autumn. For the 1999 survey, transects were assigned one of four categories of macrophyte density (0 = 0%; 1 = 1 to 19%; 2 = 20 to 80%; and 3 = > 80%) based on visual observation at two locations at each transect; the percent cover of each of the 4 categories was assigned 0%, 10%, 50%, and 90%, respectively.

Substrate size was determined by visual observation (gravel or coarser) or by Ekman or grab samples (sand or finer substrate). During 1990, substrate samples were collected at 6 locations at a subset of the transects (Randall et al. 1996); during subsequent years only two samples per transect were collected. Substrate samples were usually recorded during the year of survey, except for 1990 and 1998 when substrate was measured in the year following the fish survey. Substrate was assigned to one of seven categories based on particle size: mud 1; clay/silt 2; sand 3; gravel/pebble 4; cobble 5; boulder/armour stone 6; and bedrock 7. For each transect, the two (or more) substrate samples were averaged to determine an overall transect substrate score.

***Statistical methods*:** Before statistical analyses, the biomass of the three individual species and HPI were transformed [$\log_{10}$ (catch +1)] to normalize the distribution of the catch data. The Index of Biotic Integrity was arcsine square root transformed. In addition, both maximum fetch and macrophyte percent cover were transformed ($\log_{10}$ and arcsine square root, respectively).

Initially, relationships between fish catches and the coastal habitat features (fetch, water temperature, macrophytes density, and substrate) were illustrated using scatterplots and trend lines generated using a locally weighted robust regression (LOWESS procedure, SYSTAT 2000) procedure (tension = 0.8). LOWESS smoothing reveals potential functional relationships between variables without prejudging the shape of the relationship. Principal Components Analysis (PCA) was used to identify interrelationships among the four habitat variables and to compare the data structure in the model and the validation data sets (see below).

Linkages between fish catches and habitat attributes were quantified using regression tree analysis (TREES procedure; SYSTAT 2000). Regression tree analysis does not assume a linear relationship between the predictor variables and the response variable (Brieman et al. 1984), and preliminary examination of the habitat-fish scatterplots indicated that the correlations were non-linear. For the tree models, IBI, HPI and individual species biomass were used as response variables, and fetch by itself and then together with the three other habitat features were used as predictor variables. In addition to separate models for IBI and HPI, a 2-axis response approach was used (Randall and Minns 2002), where the two indices were combined into one model with HPI, temperature and fetch as predictors of IBI.

For the regression tree method, a least squares loss function was used to estimate the PRE statistic (proportional reduction in error) attributable to the predictive model. The PRE value is a goodness of fit statistic equivalent to multiple R^2 in regression models (SYSTAT 2000). TREES produced graphical trees beginning with one root node (entire sample) and branching (splitting) to two or more terminal nodes (subsets), each with similar habitat attributes. The TREES process is binary because each node is split into only two subsamples. Each split was determined by one habitat predictor (split variable) which divides the nodes using a split value. For example, as will be seen later, temperature was a split variable for *L. gibbosus* biomass, with a critical split value of 19.7^{o}C; *L. gibbosus* biomass was high if the water temperature was > 19.7^{o}C. At the end of each branch is a terminal node box showing the average and SD of the response variable and the number of observations (transect samples) for that node. Nodes were split using a forward stepwise procedure and algorithms from Brieman et al. (1984). For the tree stopping criteria, a minimum proportion reduction in error (PRE) allowed at any split was set at 0.05, and the minimum number of cases for the terminal nodes was set at 10. Branches stopped splitting if these criteria were not achieved. Stopping criteria are needed to avoid large tree structures with many terminal nodes. The goal of this study was to develop robust predictor models that could potentially be applied to a large geographic scale in the Great Lakes. Learning from past studies (Randall et al. 1998), our strategy was to minimize the number of nodes to a practical level of 2 to 4 terminal nodes; i.e., subsets of habitat-related transects that can be demonstrated to have biological significance, as discussed later.

Regression trees were cross validated by dividing the dataset (n = 373) into model and validation subsets. To generate preliminary predictor models, data from the 1994 survey (n = 100 transects) were used (henceforth termed the ‘model’ dataset). The 1994 data were used for modeling because a comprehensive survey was conducted in that year which covered a broad range of habitat and exposure conditions in both Lake Ontario and Lake Erie (Randall and Minns 2002). The SYSTAT TREES procedure generated a BASIC program for classifying (coding) new observations. Models were cross-validated by classifying transect samples from other years and locations (total n = 273 transects; henceforth called the ‘validation’ dataset). Classified assemblage data (HPI and IBI) were tested for significance using analysis of variance (ANOVA) and the Bonferroni post-hoc test. Although fetch and water temperature data were available from most transects, substrate and macrophyte data were available from only a subset of the validation transects. However, the validation sample size was always at least 130, and was often higher.

For individual species, the proportion of samples with zero catch of a particular species was significant (i.e., the distribution of the species catch data was positively skewed). The classification of biomass of individual species was therefore cross-validated using a two-step approach. First, a Chi-square analysis was used to test if the occurrence of 0 catches differed among the tree nodes. Second, if the Chi-square was significant, differences in average biomass of the species where present (i.e, 0 catches excluded) was tested using ANOVA.

Results

Comparison of model and validation data

Results of the PCA analysis indicated both similarities and dissimilarities in the structure of the model and validation habitat data. The model data resulted in only one component with an eigenvalue > 1 while the validation data had two components (Table 3). Coefficient loadings on the first component were similar for both data sets, showing a contrast in loadings between fetch and substrate (both negative) and temperature and cover (both positive). Water temperature was the highest coefficient for the second factor of the validation data. For both data sets, water temperature and macrophyte cover were negatively associated with exposure, and substrate size was positively associated with exposure (Fig. 3 and 4). The relationships between fetch and habitat were non-linear. The component loadings explained 58% and 86% of the variance in the model and validation data, respectively. For the model data, the percent variance explained by the PCA analysis was higher (80%) if the harbour breakwalls were excluded from the analysis (Table 1). Although PCA indicated redundancy and covariance in the habitat data, habitat variables were used separately as predictors in the tree analysis.

Habitat differed significantly among the survey areas. Macrophyte cover was highest at the coastal wetlands and within the Severn and Quinte Areas' of Concern, and was lowest at the shore sites (Fig. 5). Differences in percent cover were significant (ANOVA $F_{7,282} = 36.3$, $P < 0.01$), and were higher at the coastal wetlands in the model data than elsewhere (Bonferroni post hoc $P < 0.05$). Substrate size varied among transects and among survey areas, particularly in the validation data, but generally increased with coastal exposure (Fig. 3). Substrates were coarser at the shore and harbour transects than elsewhere (ANOVA $F_{7,242} = 40.9$, $P< 0.01$). Within the coastal wetlands and AOC areas, the average substrate category was 3 or less (sand or finer). Coastal exposure varied among survey areas (ANOVA $F_{7,272} = 30.6$; $P < 0.01$), being highest at the shore sites (Fig. 5), but with no difference in average fetch between the model and validation shore sites (Bonferroni $P > 0.05$). Although the range in exposure conditions was roughly similar in the model and validation data sets, a higher proportion of validation transects were in protected bays with low or moderate exposure than the model transects (Fig. 3). Average seasonal water temperature varied among survey locations from 16.4 to 22.7 ^{o}C (ANOVA $F_{7,354} = 38.9$, $P < 0.01$) and was lowest at the shore and harbour sites, and highest at the coastal wetlands and the three Areas of Concern. Generally, the range in habitat conditions in the validation data equaled or was less than the range for the model data.

Both average HPI and IBI differed by survey area (Fig. 6). HPI was lowest at the shore sites, and highest at the wetlands, harbours and the three AOC areas. Average IBI was also low at the shore sites, but in contrast to HPI, was relatively low at the harbours and Hamilton Harbour. Generally, the range in HPI and IBI values was similar for both the model and validation data (Fig. 6). A list of fish species captured during the electrofishing surveys is provided by Randall and Minns (2002).

Classification

Habitat: Macrophyte density is an important determinant of fish occurrence at inshore areas of the Great Lakes and a habitat model to predict the presence and density of macrophytes would be useful for developing habitat maps of plant cover in the littoral zone. In regression tree analysis, percent cover of macrophytes was related to effective fetch and substrate size for both the model and validation data sets (PRE = 0.62, N = 92, Fig. 7). Preliminary analysis using fetch alone as a predictor resulted in a model that could not be properly validated; the fetch cut value was 28 km, and this cut value was exceeded in only 6 validation transects. The regression tree model using all habitat predictors was more useful than the model with fetch by itself; three categories of macrophyte cover (low, medium and high) were predicted with confidence (Fig. 7). Applied to the validation data, the model effectively classified new transects and there was good agreement between the predicted and observed average density of macrophytes (ANOVA $F_{2,118} = 49.8$, $P<0.01$). Macrophyte percent cover was highest where substrate was fine (< 2.7) and fetch was low (< 12.6 km). Average percent cover in each of the three terminal nodes was 0%, 30% and 78%, respectively (Fig. 7). This regression tree model can be used to predict and classify habitats of varying macrophyte abundance in the lower Great Lakes if the substrate and fetch conditions are known.

Fish: For each of the fish response variables below (3 fish species, HPI and IBI), results are presented for two regression tree models: 1) using fetch alone as a predictor and 2) using fetch together with the habitat attributes of percent cover, temperature and substrate size as predictors. Results are summarized in Table 5.

***Lepomis gibbosus*:** Using fetch as the predictor, 3 nodes or habitat classes with different pumpkinseed biomasses were produced by TREE analysis (PRE = 0.44; n = 98). Two fetch cut values resulted in an increasing biomass of pumpkinseeds with decreases in fetch. Chi-square analysis confirmed that the low sunfish node had a higher proportion of zero catches than the medium and high nodes ($\chi^2_2 = 44.0$, $P < 0.01$; subscript denotes the degrees of freedom in the Chi-square analysis). For the validation data set, although *L. gibbosus* were captured at a high proportion of the transects (Table 4), only two samples (2 of 255) were classified in the low node with high fetch, effectively leaving only two groups. Chi-square analysis confirmed that the occurrence data was different among the remaining 2 nodes ($\chi^2_1 = 16.1$, $P < 0.01$), but most of the transects were classified into the high group (235 of 253). At transects where pumpkinseeds occurred, there was no significant difference in the biomass between the two groups (ANOVA $F_{1,217} = 1.9$). Fetch alone was a limited predictor of pumpkinseed biomass.

Three habitat features were predictors of pumpkinseed biomass in the TREE analysis using both habitat and fetch as predictors: macrophyte cover, water temperature and fetch (PRE = 0.76; n = 92; Fig. 8). Pumpkinseed biomass was high at sites where cover was high (> 10.5% cover), water temperature was warm (> 19.7^{o} C) and fetch was low (< 1.7 km). Presence-absence data indicated a significantly greater proportion of zero catches in the low category compared to the other 3 groups ($\chi^2_3 = 38.2$, $P < 0.01$). At sites where pumpkinseeds were present, average biomass differed among the four nodes (ANOVA $F_{3,148} = 13.6$, $P < 0.01$), although the two lower biomass groups were not significantly different (Bonferonni $P > 0.05$). Evidently, habitat predictors can be used to identify at least three habitat classes with different biomasses of pumpkinseeds (Table 5).

***Perca flavescens*:** Regression tree analysis to predict yellow perch biomass from fetch resulted in a four node model (PRE = 0.55, n = 98). However, this model was not useful for predicting the occurrence or biomass of perch for the validation data set. The first node for the model tree had a high fetch cut value of > 31 km, which included most transects with nil catches of perch. Only a small number of transects (6) in the validation set had fetch values exceeding this cut value. A large number of the remaining transects had a catch of zero as well. When the first two nodes were pooled in the validation data set to avoid low frequencies, there was no significant difference in the occurrence of 0 catches in the remaining three nodes ($\chi^2_2 = 0.13$, $P = 0.94$).

For the second model with all habitat predictors, macrophyte density was the primary habitat factor affecting the catch of yellow perch. Tree analysis of the model data resulted in three nodes of perch biomass (PRE = 0.73; n = 92), that corresponded to low (<10.5 percent cover), moderate (11-70%) and high (>71%) percent cover. Chi-square analysis of both the model and validation data sets indicated that the proportion of zero catches was significantly higher in the low node and lower in the moderate and high biomass nodes than expected ($\chi^2_2 = 51.7$ and 38.4, respectively, both $P < 0.01$). For transects where yellow perch were captured, average biomass was significantly different among the predicted tree nodes ($F_{2,153} = 5.89$, $P < 0.01$), although the post hoc Bonferroni test indicated a significant difference between the low and high nodes only. Three classes of habitat were identified with different yellow perch biomasses, although the moderate and high nodes could be combined.

***Alosa pseudoharengus*:** Based on the model with fetch as the only predictor, three nodes of alewife biomass (low, medium, high) were identified in the regression tree analysis (PRE = 0.36, n = 98). High abundance of alewife occurred in areas of high fetch, in contrast to yellow perch and pumpkinseed which occurred less frequently or were absent at sites with high exposure. Chi-square analysis of the presence - absence data was significant for both the model ($\chi^2_2 = 16.3$, $P<0.01$) and the validation data sets ($\chi^2_2 = 8.3$, $P<0.05$), indicating a higher proportion of zero catches in the low node than in the moderate or high nodes. The latter test was weak as the high node had a low frequency of samples (n = 6). When the moderate and high nodes were pooled, the chi-square test remained significant ($\chi^2_1 = 12.6$, $P<0.01$). Analysis of samples where alewife were present showed significant differences in mean biomass among groups (ANOVA $F_{2,136} = 4.1$, $P < 0.05$), but the difference was significantly different between the low and high nodes only (Bonferroni $P< 0.01$). Fetch was of limited use for predicting alewife biomass in the validation data set, as only 6 of 255 samples were in the high node (Table 5).

Only two nodes of alewife biomass were produced from the tree analysis that included all habitat predictors, with water temperature being the only significant predictor (PRE = 0.71, n = 92). For the model data set, alewife biomass was high at transects where the water temperature was less than 16.7^{o}C. The occurrence of 0 catches was significantly higher in the low node for both the model ($\chi^2_1 = 24.8$, $P<0.01$) and validation data set ($\chi^2_1 = 5.6$, $P<0.05$). However, the results were similar to the fetch model above, in that the high node had a low frequency of samples (n = 10 of 244 samples). For transects where alewife were present, there was no significant difference in the average biomass of alewife between the two nodes ($P > 0.05$). Possibly this result was related to the schooling behaviour of alewife, suggesting that alewife occurrence may best be best predicted with presence/absence data.

Habitat Productivity Index: Tree analysis of the Habitat Productivity Index with fetch as the predictor produced four nodes with the model data (PRE = 0.38, n = 98), but only two nodes were significantly different in the validation data (ANOVA $F_{1,153} = 33.1$; P = 0.01; Fig. 9). A fetch cut value of 4 km separated transects into groups of low (mean 26.0 kg ha yr^{-1}) or high (48.2 kg ha yr^{-1}) productivity. Average HPI was about 1.9 times greater in the low fetch node (Table 5).

For the model with all habitat predictors, macrophyte density, substrate and fetch were significant predictors of productivity, resulting in four nodes (PRE = 0.59, n=98). However, only two nodes were significantly different in the validation data set (Bonferonni post hoc analysis, $P<0.05$). Consequently the data were pooled into two significant nodes based on the first tree cut value of > 7 percent macrophtyte cover (ANOVA $F_{1,119} = 9.5$; $P < 0.01$). The average HPI in the low and high nodes was 24.7 and 40.6 kg ha yr^{-1}, respectively, similar to the values above. Fetch and macrophyte density yielded similar results as classification variables and resulted in only two classes of habitat productivity.

Index of Biotic Integrity: With IBI as the response variable and fetch as the predictor variable, two groups were identified in the model data, with a fetch cut value of 31 km (PRE = 0.30, n = 98). Using this fetch criterion, two significant nodes were produced in the validation data as well (ANOVA $F_{1,253} = 15.7$, $P < 0.01$). However, only 6 of 255 transects in the validation data had a fetch value exceeding the cut value and were classified in the low IBI group. The large proportion of the samples in the high IBI group could not be classified further using fetch.

The habitat predictor variables of percent cover and water temperature produced 3 IBI groups (PRE = 0.50, n = 92). The high IBI group resulted from a percent cover cut value of > 19%, and the low IBI group occurred at transects where the water temperature was < 19.7 ^{o}C (Fig. 10). The resulting TREES model successfully classified three groups in the validation data (ANOVA $F_{2,187} = 32.8$) (Table 5); however, the low and moderate groups were not significantly different (Bonferroni $P > 0.05$).

Two-axes Index of Habitat Capacity: The above results indicated that both HPI and IBI were related in a non-linear but consistent manner to fetch (Fig. 11) and percent cover (Fig. 12) in the model and validation datasets. There was considerable variability in both indices along the two habitat gradients, and the two indices were to some extent independent, as shown below.

For the 2-axes approach, IBI was predicted from fetch, thermal conditions and fish biomass. Water temperature, fetch and HPI were significant grouping variables for identifying areas of high, moderate and low IBI for both the model data (Fig. 13; PRE = 0.46) and the validation data ($F_{3,240}$ = 18.2; n = 244; P < 0.01). Post hoc testing of the validation data indicated that most group averages were significantly different (Bonferroni P < 0.01), with the exception of the two moderate groups which were only marginally different (P = 0.08). Based on these results, all data were pooled (model plus validation), and all four groups significantly different from one another (ANOVA $F_{3,338}$ = 43.2; n = 342; P < 0.01; group means different at P < 0.01).

Average IBI scores in the 2-axes groups increased significantly from about 36 to 67 from the low to the high group (Fig. 13; Table 5). Of special interest was the moderate-high IBI group: although IBI was relatively high, HPI was relatively low in this group. For individual species, the average biomass of yellow perch and pumpkinseed increased with the IBI groups as expected, and the biomass of alewife declined. P/B ratios of perch and alewife were significantly higher in the medium-high or high IBI group. The low 2-axes group included shore and a few harbour sites, the low-medium group included harbours, shore and many Hamilton sites, the medium-high group included some wetlands and Severn Sound sites, and the high group was primarily wetlands and high productivity areas of Quinte, Severn Sound and a few Hamilton sites (Table 6). For the pooled data, the medium-high and high IBI groups could be separated on the basis of cover; average percent cover was 40% and 58%, for the medium-high and high groups, respectively.

Discussion

The near shore habitat features of macrophyte cover, water temperature, and substrate size were related to coastal exposure in the Great Lakes, and this resulted in a discernible link between fetch and the catch of fish at the coastal areas. All three hypotheses addressed in this study were supported to a varying extent with the field data: (1) the association between fetch and habitat characteristics was quantifiable; (2) fetch, directly or together with habitat attributes, was used to predict fish catch metrics; and (3) individual species abundance was predicted more accurately than assemblage indices, although the results were species-dependent and preliminary (only 3 species were examined). Each of these hypotheses are discussed individually below, following an evaluation of the regression tree method for classifying coastal habitat. Results of the habitat classification models can be used to map the productive capacity of coastal habitat in the Great Lakes.

The relationship between fish biomass and fetch was spatially robust as the survey data used to develop and validate the models were temporally and spatially extensive, covering both impacted and natural shorelines that varied in the nature of the physical habitat and the extent of exposure. Using the 1994 survey data to develop regression tree models, which were then applied to and tested on the remaining survey data, had both advantages and disadvantages. An advantage was that the model data included a range of habitat conditions that equaled or exceeded conditions at the other sites which allowed validation of the models without extrapolating to new conditions. The model data were collected from both lakes Erie and Ontario, and previous analysis indicated that the fish-habitat associations were similar at each lake (Randall and Minns 2002). In addition, survey data from harbour breakwalls in the model dataset provided information from areas with altered habitat. The harbour breakwalls were man-made structures that provided an 'experimental' situation for examining the effect of reduced fetch on fish catches at exposed shoreline areas. A disadvantage, however, was that break-wall transects were rare in the validation data set, and there was a much higher proportion of low or moderate fetch sites in the validation data than in the model data. Much of the validation data came from wind-protected bays. Generally, however, the models generated from the extreme conditions of the shore-harbour-wetland data collected in 1994 provided robust and generic guidelines on the factors that influence fish distribution.

Regression tree analysis, although related to other statistical methods (discriminate function analysis, regression), was useful for classifying the fish-habitat data as it does not assume a linear or even monotonic relationship between the predictor variables and the response variable (Brieman et al. 1984; Systat 2000). Regression tree analysis is non-parametric and it is robust to outliers (LeBlanc 2002). For the Great Lakes dataset, scatterplots of fish density or diversity showed non-linear and highly variable patterns along the gradients in habitat conditions. Regression tree procedures provided an objective method of quantifying cut points, the threshold habitat values that significantly influenced the response variable. Threshold values are critical for developing fish-habitat linkage models, but importantly, the cut values depended on the dataset being modeled. For this reason, validation of the test models using new data was a key component of this study. The number of validated classes was often less than the number generated by the model data. Also, variability within tree nodes was high and although average response variables were significantly different among the nodes, classification of individual sites remained uncertain.

Linkages between coastal exposure and habitat attributes were quantifiable. As expected, substrate particle size increased and percent cover and water temperature decreased with increased fetch distance at the sites. Substrate size and percent cover were probably functionally related to fetch through the effect of wave energy, but identifying whether the relationship was functional or simply correlative was beyond the objective of this study. PCA analysis confirmed that the fetch-habitat relationships were consistent between the model and validation data. The occurrence and density of aquatic macrophytes is a primary factor affecting the distribution and abundance of fishes in the littoral zone of the Great Lakes (Jude and Pappas 1992), and aquatic plants are a useful indicator of habitat productive capacity (Randall et al. 1996). Although the associations between fetch, substrate and percent cover were variable, three macrophyte classes of low, moderate or high plant density were successfully predicted and validated with new data. Plant occurrence and maximum fetch were inversely related, but the predictive model was improved significantly when both substrate and fetch were used together as predictors, as macrophyte abundance and substrate size were also related. Fetch distance below a threshold of 12.6 km increased the likelihood of plant occurrence but fine substrate was also critical for plant growth. The resulting regression tree model can be used to quantify and map the occurrence and abundance of aquatic macrophytes at 1.5 metre water depth from knowledge of coastal exposure and substrate conditions. In future, this plant-habitat model can be improved and the application extended by incorporating measures of water clarity, as irradiance determines the maximum depth of aquatic plant colonization in the littoral zone of lakes (Chambers 1987; Chambers and Kalff 1985).

Fish species occurrence was significantly related to fetch distance, but as was the case with macrophytes, fetch alone was a limited predictor of fish biomass, resulting in only two classes of low and high biomass. The degree of resolution of habitat classes increased if additional habitat attributes were used as predictors. In some cases, both habitat and fetch were significant predictors (e.g., *L. gibbosus* habitat was best classified with cover, temperature and fetch), in other cases, habitat attributes classified the response variable more effectively than fetch (cover and water temperature for *P. flavescens* and *A. pseudoharengus*, respectively). Proportional reduction in error (PRE) for the habitat models was higher (0.50 to 0.76) than for the fetch models (0.36 to 0.55). Also, the number of validated habitat classes was greater if habitat attributes (2 to 4 classes) rather than fetch (usually 2 classes) was used to predict fish abundance or diversity. Validated fetch-based models were often rated as poor because, despite the large validation sample size, the high fetch node had few cases (< 10), and fetch was not a good discriminator or predictor variable for fish biomass within the protected bays with low to intermediate fetch. *L. gibbosus*, a species that resides solely in warm shallow areas, was classified more successfully using habitat information than *A. pseudoharengus*, an offshore pelagic species. Generally, the PRE values for individual species were higher than for the assemblage response variables, but the results were species-dependent and preliminary (only 3 species were examined).

As a first order estimate of habitat productive capacity, fish community measures are more generally applicable for habitat management than individual species metrics. Using IBI and HPI indices together in a two-axes approach for determining productive capacity has potential merit. Productive capacity by definition has both a fish production and a biodiversity component (Minns 1997), and IBI and HPI account for both the diversity and production components, respectively (Randall and Minns 2002). Although both indices responded to fetch and cover in a consistent manner, cut values were index-dependent; that is, threshold values for both fetch and cover tended to be higher for IBI than HPI (Fig. 11). If biomass and species richness are used as indicators of productive capacity, the highest capacity occured in areas with high IBI and high HPI. Areas with high HPI (fish biomass) but only moderate diversity as measured by IBI were ranked lower than areas with moderate HPI but high IBI (Randall and Minns 2002). The 2-axis regression tree method produced four habitat classes, ranging from low IBI in areas of cold water and high maximum fetch to high IBI in low exposure areas with warm water and high HPI. Interestingly, the second highest IBI group occurred where IBI was high, but HPI was only moderate. Coastal wetland sites were classified into one of these two groups, with the high group occurring in areas with more macrophytes than the medium-high group. The two-axes tree model illustrated in Fig. 13 is a potentially useful first-order estimate of the productive capacity of inshore habitat in the Great Lakes.

Although the degree of resolution of habitat classes was limited, the ability to determine productive capacity from habitat attributes has potential utility and application for habitat management. In addressing the role of uncertainty and complexity in habitat management decision making, Minns and Moore (2003) concluded that ‘despite the broad uncertainty surrounding many fish-habitat associations, simple habitat classifications involving as few as three or four levels of productive capacity can provide a basis for robust decisions’. The first-order model of productive capacity described in this study provides such a habitat classification scheme. Fetch measures can be determined using GIS information for the Great Lakes. Substrate characteristics are known for some areas, or it may be predicted from erosion-transport deposition (ETD; Franzin et al. 2001) or other coastal process models. Macrophyte occurrence and density is known for Areas of Concern (Minns et al.1999; Project Quinte 2001), or cover can be predicted using the habitat model from this study. Thermal habitat in shallow water coastal areas can possibly be predicted from air temperature data, as is done in streams (Stoneman and Jones 1996), together with other governing factors such as surface water currents (C. Chu, personal communication). All of the habitat characteristics used in the regression tree predictive model are known or can be generated for large areas of coastal habitat.

Evaluation of habitat productive capacity can be viewed as a two-stage task, where first-order evaluations (this study) could be followed by more detailed site-specific evaluations (Defensible Methods; Minns 1997) if warranted by the nature of the project. For both stages, significant evaluations can be made on the productive capacity of Great Lakes shoreline habitat from knowledge of the coastal exposure and the associated habitat covariates.

Acknowledgements

Fish and habitat data were collected by field teams led by Doug Geiling (1990), Brent Valere (1991 to 1998) and Christine Brousseau (1998 to present). The line-transect survey design was developed by K. Minns and V. Cairns for evaluating the status of the near shore fish communities at Great Lakes' Areas of Concern. T.C. Pratt, R. S. Gregory and J.R.M. Kelso provided helpful comments on an earlier draft. This study was funded by Fisheries and Oceans Canada (Environmental Sciences Strategic Research Program).

References

Brieman, L. , J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. Classification and regression trees. Wadsworth and Brooks/Cole, Monterey, California.

Chambers, P.A. Nearshore occurrence of submersed aquatic macrophytes in relation to wave action. Can. J. Fish. Aquat. Sci. 44:1666-1669.

Chambers, P.A. and J. Kalff. 1985. Depth distribution and biomass of submersed aquatic macrophyte communities in relation to Secchi depth. Can. J. Fish. Aquat. Sci. 42:701-709.

Coker, G.A., C.B. Portt and C.K. Minns. 2001. Morphological and ecological characteristics of Canadian freshwater fishes. Can. MS Rpt. Fish. Aquat. Sci. 2554:iv + 86 p.

Franzin, W.G., P.A. Nelson and P.M. Cooley. 2001. An introduction to ETD concepts in understanding fish habitat, pg. 90-99 *In* Randall, R.G. (ed). 2001. Field Measurement of the Productive Capacity of Fish Habitat – Proceedings of the Second Project Workshop, November 2000. Fisheries and Oceans Canada, Burlington, Ontario.

Håkanson, L., H. Kvarnås and B. Karlsson. 1986. Coastal morphometry as regulator of water exchange – a Swedish example. Estuarine, Coastal and Shelf Science 23:873-887.

Jude, D.J. and J. Pappas. 1992. Fish utilization of Great Lakes coastal wetlands. J. Great Lakes Res. 18:651-672.

LeBlanc, M. 2002. Regression Trees, pg 1734-1738 *In*: Encyclopedia of Environmetrics. A.H. El-Shaarawi and W.W. Piegorsch. 2002. John Wiley & Sons, Ltd, West Sussex, England.

Minns, C.K. 1995. Calculating net change of productivity of fish habitats. Can. Manusc. Rep. Fish. Aquat. Sci. No. 2282.

Minns, C.K. 1997. Quantifying 'No Net Loss' of productivity of fish habitats. Can. J. Fish. Aquat. Sci. 54:2463-2473.

Minns, C.K. and J.E. Moore. 2003. Assessment of net change of productive capacity of fish habitats: the role of uncertainty and complexity in decision making. Can. J. Fish. Aquat. Sci. 60:100-116.

Minns, C.K., P. Brunette, M. Stoneman, K. Sherman, R. Craig, C. Portt, and R.G. Randall. 1999. Development of a fish habitat classification model for littoral areas of Severn Sound, Georgian Bay, a Great Lakes' Area of Concern. Can. Manusc. Rep. Fish. Aquat. Sci. No. 2490, 88 p.

Minns, C.K., Kelso, J.R.M. and Randall, R.G. 1996. Detecting the response of fish to habitat alterations in freshwater ecosystems. Can. J. Fish. Aquat. Sci. 53 (Suppl. 1):403-414.

Minns, C.K., Cairns, V.W., Randall, R.G., and Moore, J.E. 1994. An Index of Biotic Integrity (IBI) for fish assemblages in the littoral zone of Great Lakes' Areas of Concern. Can. J. Fish. Aquat. Sci. 51:1804-1822.

Portt, C.B. and King, S.W. 1991. Fish community structure, abundance, and biomass along three southern Ontario breakwaters and in adjacent reference habitats. Manuscript Report prepared for Small Craft and Harbours Branch, Dept. of Fisheries and Oceans, Burlington, Ontario. 45 p.

Project Quinte. 2001. Project Quinte Annual Report, 1999-2000. Bay of Quinte Remedial Action Plan Monitoring Report No. 11, 142 p. Kingston, Ontario.

Randall, R.G. 2003. Fish productivity and habitat productive capacity: definitions, indices, units of field measurement, and a need for standardized terminology. Canadian Science Advisory Secretariat Research Document 2003/061.

Randall, R.G. and Minns, C.K. 2002. Comparison of a Habitat Productivity Index (HPI) and an Index of Biotic Integrity (IBI) for measuring the productive capacity of fish habitat in nearshore areas of the Great Lakes. J. Great Lakes Research 28:240-255.

Randall, R.G. and Minns, C.K. 2000. Use of fish production to biomass ratios for measuring the productive capacity of fish habitats. Can. J. Fish. Aquat. Sci. 57:1657-1667.

Randall, R.G., Minns, C.K., Cairns, V.W. and Moore, J.E. 1998. Habitat predictors of fish species occurrence and abundance in nearshore areas of Severn Sound. Can. Manuscript Rep. Fish. Aquat. Sci. No. 2440, 30 p.

Randall, R.G., Minns, C.K., Cairns, V.W. and Moore, J.E. 1996. The relationship between an index of fish production and submerged macrophytes and other habitat features at three littoral areas in the Great Lakes. Can. J. Fish. Aquat. Sci. 53 (Suppl. 1):35-44.

Randall, R.G., C.K. Minns, V.W. Cairns, and J.E. Moore. 1993. Effect of habitat degradation on the species composition and biomass of fish in Great Lakes Areas of Concern. Can. Tech. Rept. Fish. Aquat. Sci. 1941, 37 p.

Scheffer, M., Redelijkhied, M. R. de, and Noppert, F. 1992. Distribution and dynamics of submerged vegetation in a chain of shallow eutrophic lakes. Aquatic Botany 42:199-216.

Smokorowski, K.E., M.G. Stoneman, V.W. Cairns, C.K. Minns, R.G. Randall and B. Valere. 1998. Trends in the nearshore fish community of Hamilton Harbour, 1988 to 1997, as measured using an Index of Biotic Integrity. Can. Tech. Rep. Fish. Aquat. Sci. 2230: xii+97 p.

Stoneman, C.L. and M.L. Jones. 1996. A simple method to classify stream thermal stability using single observations of daily maximum water and air temperature. North American Journal of Fisheries Management 16:728-737.

Stoneman, M. 1998. A program to calculate Index of Biotic Integrity scores for Great Lakes' Areas of Concern: summary and operating intructions, pg. 77-86. In: Smokorowski, K.E., M.G. Stoneman, V.W. Cairns, C.K. Minns, R.G. Randall, and B. Valere. 1998. Trends in the nearshore fish community of Hamilton Harbour, 1988 to 1997, as measured using an Index of Biotic Integrity. Can. Tech. Rept. Fish. Aquat. Sci. No. 2230.

Systat 10. 2000. SYSTAT 10 Statistics I. SPSS Inc. Chicago, IL .

Valere, B. G. 1996. Productive capacity of littoral habitats in the Great Lakes: field sampling procedures (1988-1995). Can. Manuscript Rep. Fish. Aquat. Sci. No. 2384, 50 p.

W.F. Baird & Associates. 1996. Defensible methods of assessing fish habitat: physical habitat assessment and modeling of the coastal areas of the lower Great Lakes. Can. Manuscript Rep. Fish. Aquat. Sci. No. 2370.

Table 1. Summary of survey areas for the model and validation databases. The number of transects (N) in indicated by year.

Database	Area	Year	N	Notes
Model	Harbour	1994	51	Surveyed exposed outside and protected inside of harbour breakwalls (armour stone); Port Dover and Port Colbourne, Lake Erie; Port Dalhousie, and Bronte Harbour, Lake Ontario
	Shore	1994	30	Exposed shores to each side of harbours
	Wetland	1994	19	Long Point, Lake Erie and Presqu'ile, Lake Ontario
		Total model	100	[1]The number of samples per year (n=2, 3 or 4) was 2, 98 and 0
Validate	Wetland	1998	11	Black River bay, West Lake
		1999	12	Black River bay, West Lake
	Shore	1998	12	McMahon Bluff, Little Bluff, Bronte shore
		1999	12	McMahon Bluff, Little Bluff, Bronte shore
	Severn Sound (AOC)	1990	55	AOC monitoring; protected bays of Penetang, Hog and Matchedash
		1992	20	Exposed outer areas of Penetang and Hog bays; Sturgeon Bay
		1995	18	Green Island (Canadian Shield) and Hog Bay
	Bay of Quinte (AOC)	1990	20	AOC monitoring at Trenton, Belleville and Big Island
		1992	23	Hay and Carnachan bays, Conway
		1999	26	AOC monitoring at Trenton, Belleville and Big Island
	Hamilton Harbour (AOC)	1990	20	AOC monitoring, protected or moderately exposed shoreline
		1997	44	AOC monitoring as above, after habitat restoration
		Total validate	273	[1]Seasonal samples per year (n=2, 3 or 4): 101, 107 and 65

[1] number of surveys per transect per year ranged from 2 (summer) to 4 (late spring, summer and early autumn) depending on the location, year and survey objectives (Valere 1995).

Table 2. Preferred temperature, habitat preference, and feeding location of the three most abundant fish species captured in near shore areas of the Great Lakes. From Coker et al. (2001) and Lane et al. (1996).

Species	Preferred Temperature (°C)	Habitat Preference (depth, cover and substrate)	Feeding Location	Description
Lepomis gibbosus	26.0 (warm)	0 to 2 m, high preference for macrophyte cover; sand and silt	bottom and pelagic	Near shore
Perca flavescens	21.4 (cool)	0 to >10 m, medium preference for macrophyte cover; sand and silt	bottom and pelagic	Near shore - off shore
Alosa pseudoharengus	18.8 (cold)	0 to > 10m; no cover; rubble, gravel, sand	pelagic and bottom	Off shore

Table 3. PCA analysis of the four habitat variables used in this study after selection of the components with eigenvalues > 1. Results for both the model and validation data sets are compared. For the validation data set, the highest component coefficient for each variable is in bold font. For the model data, results are shown both including and excluding (values in parenthesis) survey transects from the harbour breakwalls (with armour stone).

	Principal Components and Loadings		
Data set	Model	Validate	Validate
Component	1	1	2
Eigenvalue	2.31 (3.19)	2.36	1.08
% variance explained	57.7(79.6)	58.9	26.9
Cumulative		58.9	85.8
Variable coefficients			
Water temperature	0.86 (0.92)	0.45	**-0.87**
Maximum effective fetch	-0.51(-0.90)	**-0.90**	0.15
Substrate category	-0.78(-0.89)	**-0.87**	-0.12
Macrophyte cover	0.84 (0.87)	**0.76**	0.54

Table 4. Presence-absence frequency of *Lepomis gibbosus, Perca flavescens* and *Alosa pseudoharengus* at the survey locations.

Species	Database			
	Model (n=100)		Validate (n=273)	
	Absent	Present	Absent	Present
Lepomis gibbosus	56	44	40	233
Perca flavescens	54	46	55	218
Alosa pseudoharengus	46	54	122	151

Table 5. Summary of regression tree classification of species biomass and fish assemblage measures (response variable) of habitat productive capacity using habitat attributes as predictors. Results of the tree model to predict percent macrophyte cover are also shown.

Response variable[1]	Predictors[1]	Model		Validation[2]		Validation notes[3]
		PRE	Nodes	Nodes	Capacity	
L. gibbosus	1) Fetch	0.44	3	2	1.0\2.6	Poor
	2) Habitat (C,T,F)	0.76	4	3-4	1.0\1.9\2.7\5.4	Good; possibly combine 2 moderate nodes
P. flavescens	1) Fetch	0.55	4	2	1.0\5.4	Poor
	2) Habitat (C)	0.73	3	2-3	1.0\2.3\4.3	Good; possibly combine moderate and high nodes
A. pseudoharengus	1)Fetch	0.36	3	2-3	1.0\1.3\6.5	Poor, even with 2 groups
	2) Habitat (T)	0.71	2	2	1:0\1.4	Poor
HPI	1) Fetch	0.38	4	2	1.0\1.9	Good
	2) Habitat (C,S,F)	0.59	4	2	1.0\1.6	Good
IBI	1) Fetch	0.30	2	2	1.0\2.0	Poor
	2) Habitat(C,T)	0.50	3	3	1.0\1.1\1.5	Good for 2 groups; combine model and validation data?
IBI (2-axes)	T, F, HPI	0.46	4	4 or 3	1.0\1.4\1.6\1.9	Good
Percent cover	F, S	0.62	3	3	0%\30%\78 % cover	Good; cut values: S <2.7 and F<12.6

[1] For each of the first 5 response variables, two predictor models were used: 1) fetch alone and 2) fetch together with other habitat variables. Only significant predictors are shown. Abbreviations: fetch (F), percent cover (C) , temperature (T), and substrate (S). [2] Capacity indicates the magnitude of the increase in the response variable between nodes (for percent cover, average values are shown) [3] For validation notes, poor indicates there are ≤ 10 transects in a node.

Table 6. Survey area, sample size and percent of transects in each of four IBI groups for the model and validation data sets by area. N is the number of transect samples in each area (column) and IBI group (row). IBI groups were determined by regression analysis, with temperature, HPI and fetch as predictor variables (see Fig. 5).

Data set and area	N	IBI Group			
		Low	Low-medium	Medium-high	High
Model					
Harbour	49	20.4	53.1	4.1	22.4
Shore	30	56.7	36.7	6.7	0.0
Wetland	19	0.0	10.5	31.6	57.9
Validate					
Wetland	23	0.0	0.0	34.8	65.2
Shore	18	33.3	44.4	22.2	0.0
Severn	87	0.0	23.0	39.1	37.9
Quinte	69	0.0	1.4	18.8	79.7
Hamilton	47	0.0	59.6	4.3	36.2
N	342	33	96	71	142

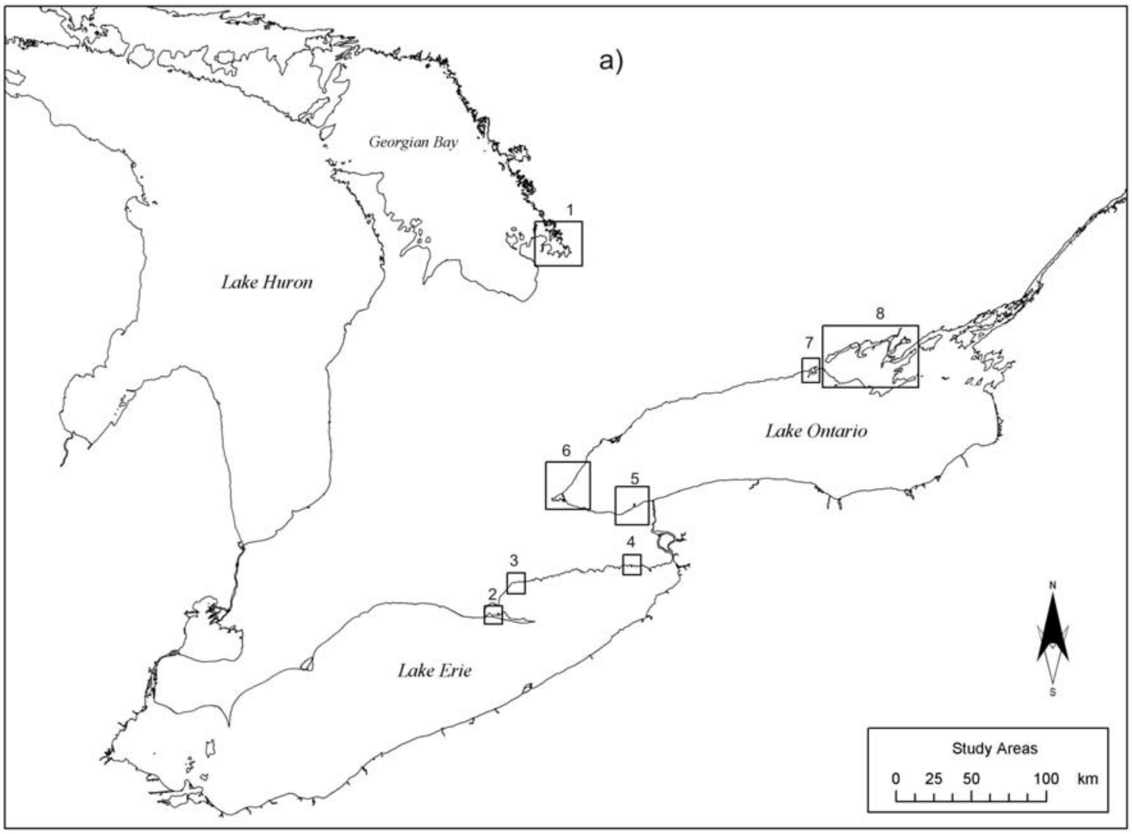
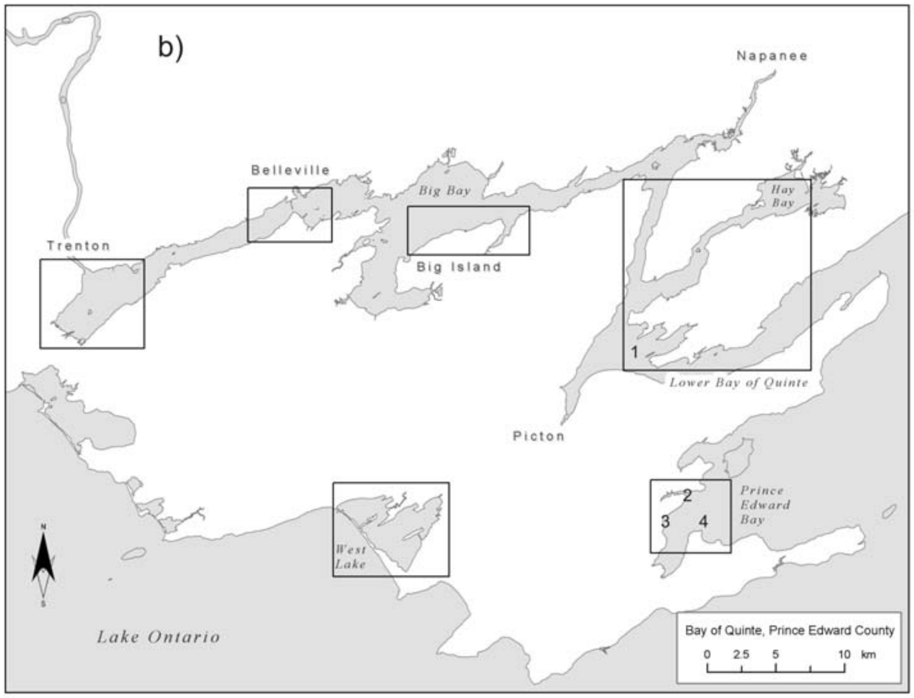


Figure 1a). Map of the lower Great Lakes showing the survey locations (boxes). Legend: 1 – Severn Sound; 2 – Long Point; 3 – Port Dover; 4 – Port Colbourne; 5 – Port Dalhousie; 6 – Hamilton Harbour and Bronte Harbour; 7 – Presqu'ile; 8 – Bay of Quinte. Maps of Bay of Quinte and Severn Sound are enlarged below to show details. The number of electrofishing transects at each location are listed in Table 1.



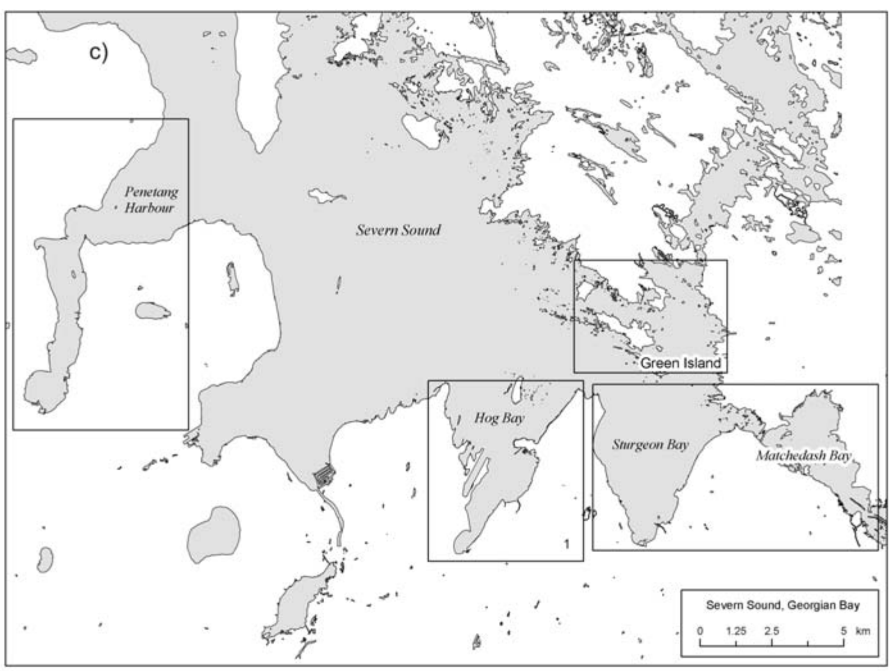


Figure 1b): Enlarged map of Bay of Quinte and Prince Edward County showing the survey locations. Legend: 1 – Carnachan Bay; 2 – Black River mouth; 3 –McMahon Bluff; 4 – Little Bluff. Other locations are labelled. 1c): Enlarged map of Severn Sound showing the survey locations.

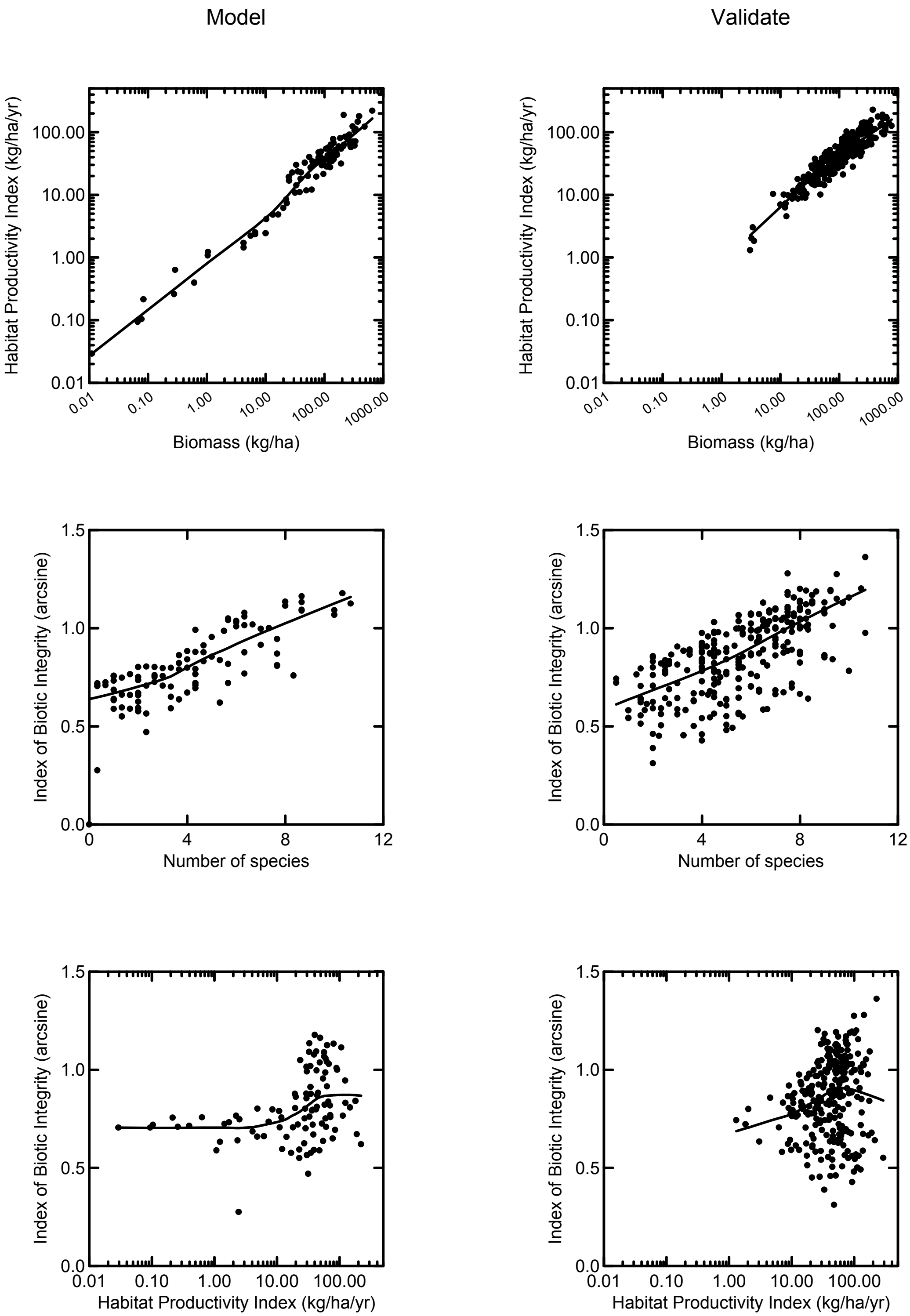


Figure 2. Comparison of average fish biomass and HPI (upper); species richness and IBI (middle), and HPI versus IBI (lower) for both the model and validation data sets..

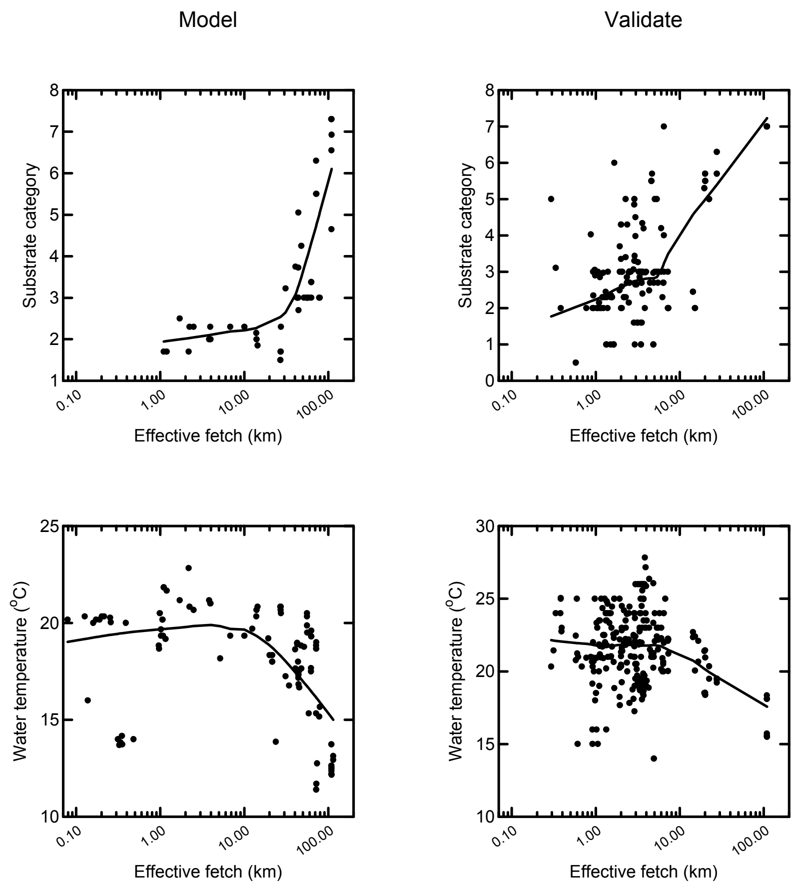


Figure 3. Relationship between maximum effective fetch and substrate category (upper) and average water temperature (lower) for the model and validation data sets.

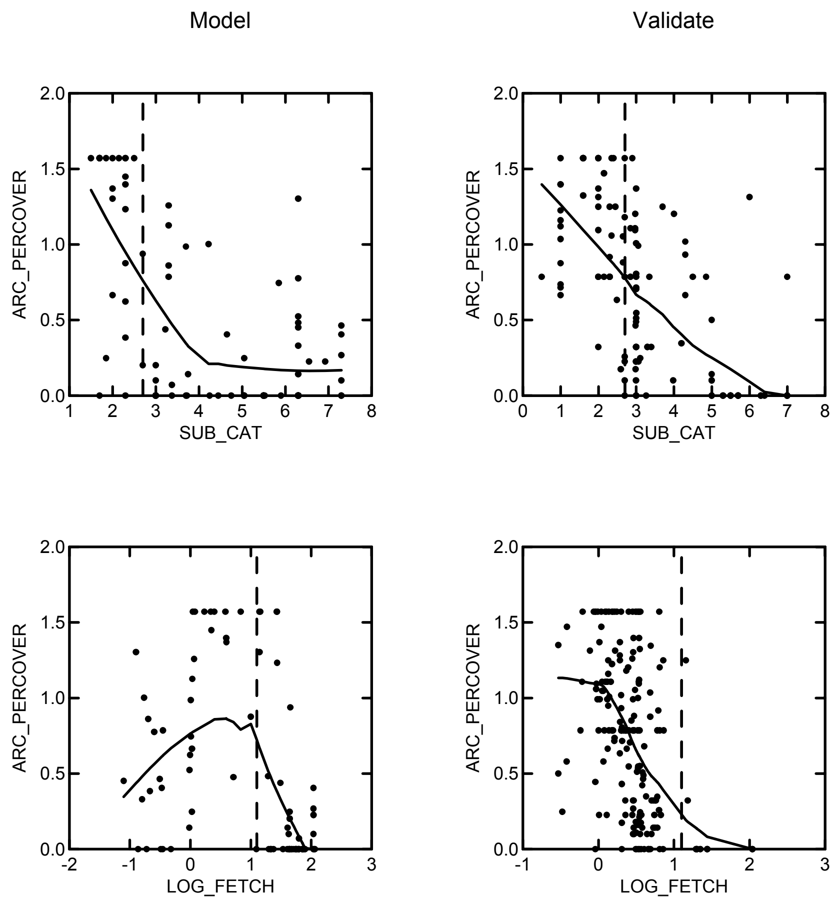


Figure 4. Scatter plots of macrophyte cover (ARC_PERCOVER is arcsine percent cover) versus substrate size (SUB_CAT) and log fetch for the model and validation data sets.

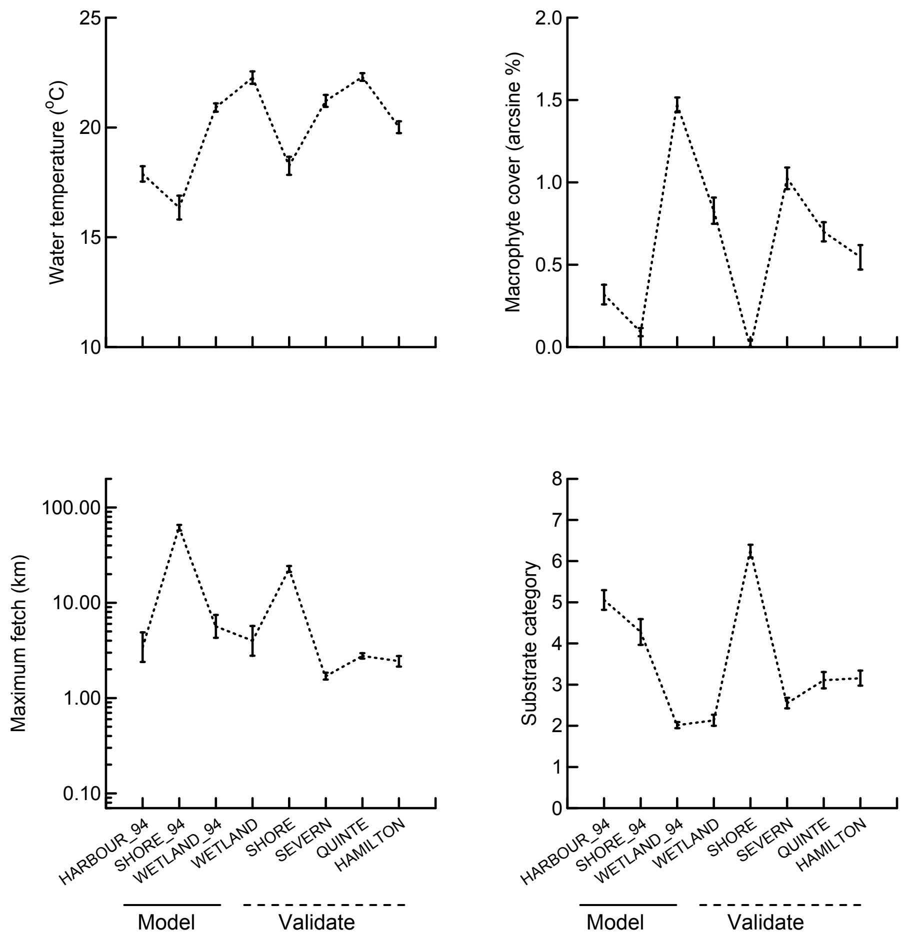


Figure 5. Comparison of average water temperature, macrophyte density, maximum fetch distance, and substrate category at the different model and validation survey areas.

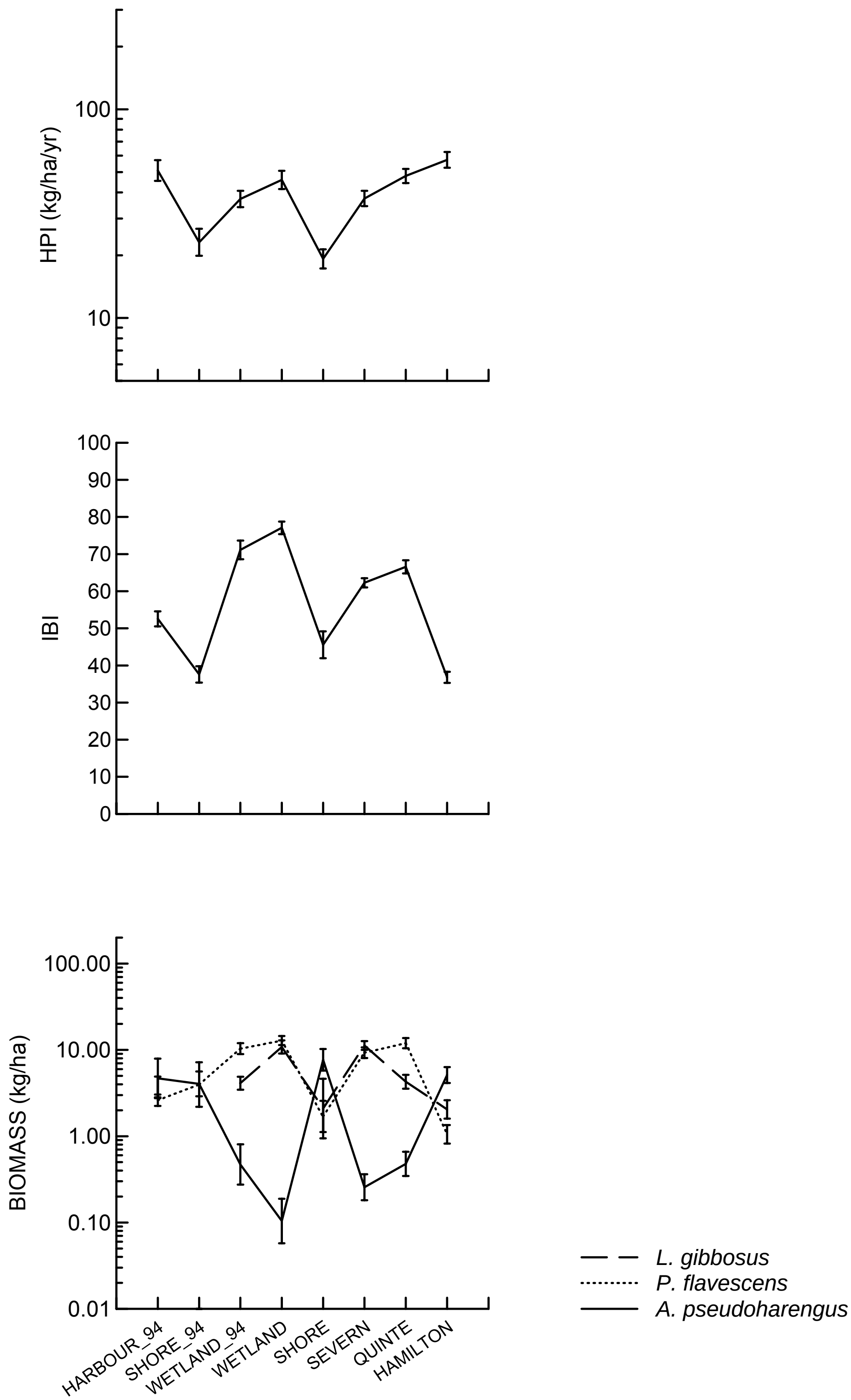


Figure 6. Average HPI, IBI and biomass of *Lepomis gibbosus*, *Perca flavescens* and *Alosa pseudoharengus* at the different survey areas.

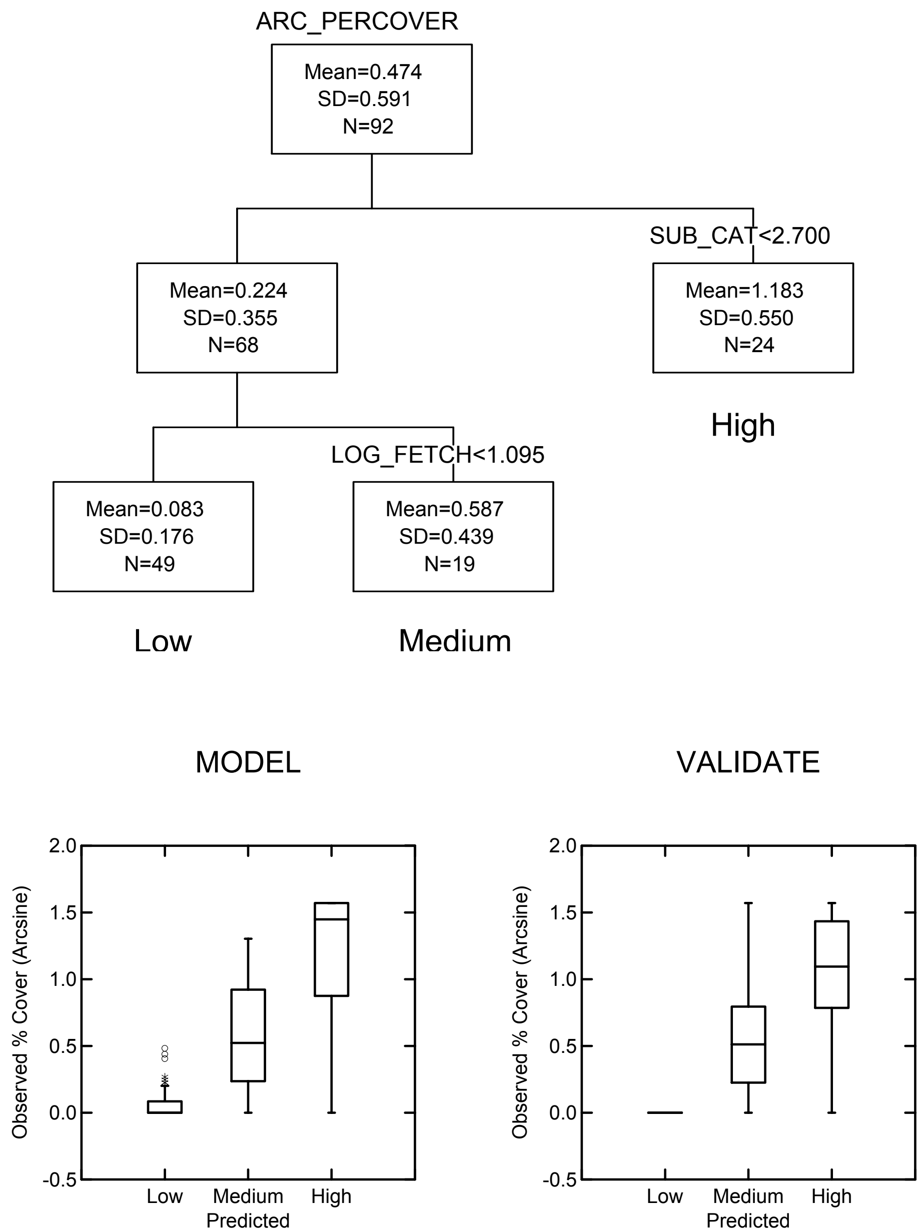


Figure 7. Upper: Regression tree (PRE = 0.62, N = 92) to classify macrophyte density (ARC_PERCOVER: arcsine percent cover) from substrate size (SUB_CAT) and fetch (LOG_FETCH). Cut values: substrate value of 2.7 is mud, silt or fine sand; fetch of 1.095 is 12.4 km. Lower: comparison of predicted versus observed macrophyte cover for the model and validation data sets.

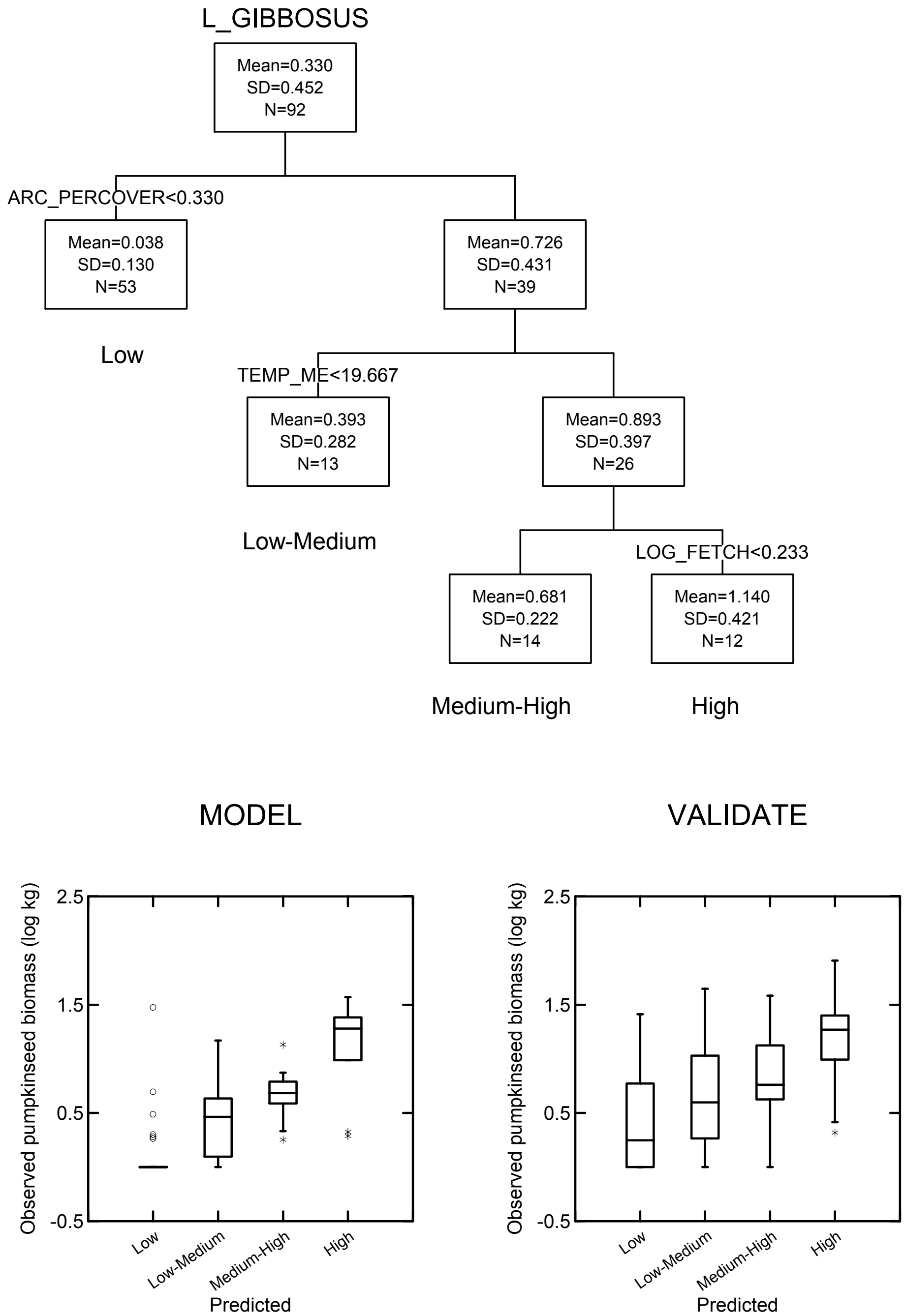
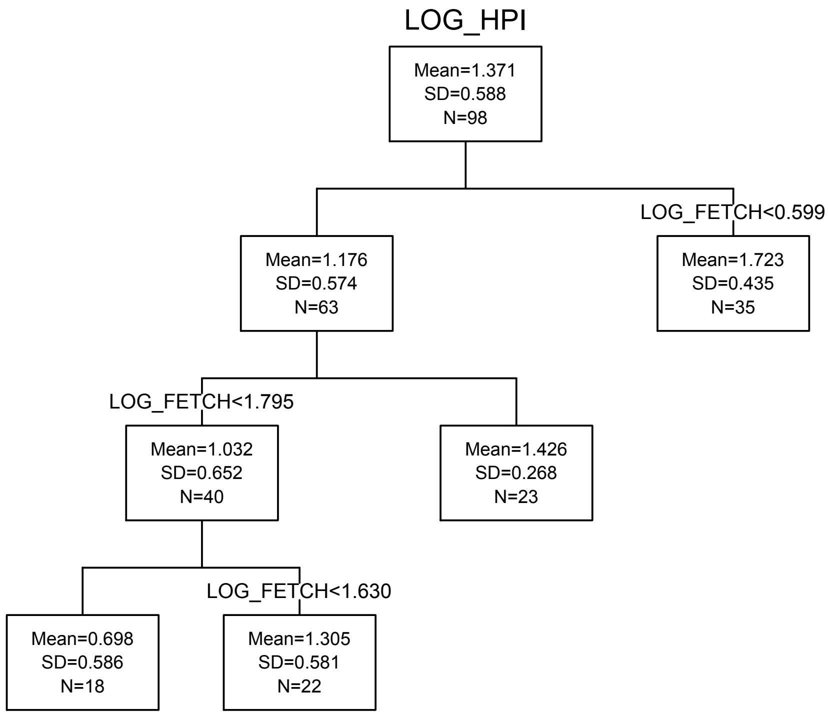


Figure 8. Upper: Regression tree (PRE = 0.76, N = 92) to classify pumpkinseed biomass (L_GIBBOSUS, log biomass) from macrophyte density (ARC_PERCOVER: arcsine percent cover) water temperature (TEMP_ME) and fetch (LOG_FETCH). Cut values: 10.5% cover; 19.7^{o}C; fetch of 1.7 km. Lower: comparison of predicted versus observed macrophyte cover for the model and validation data sets.



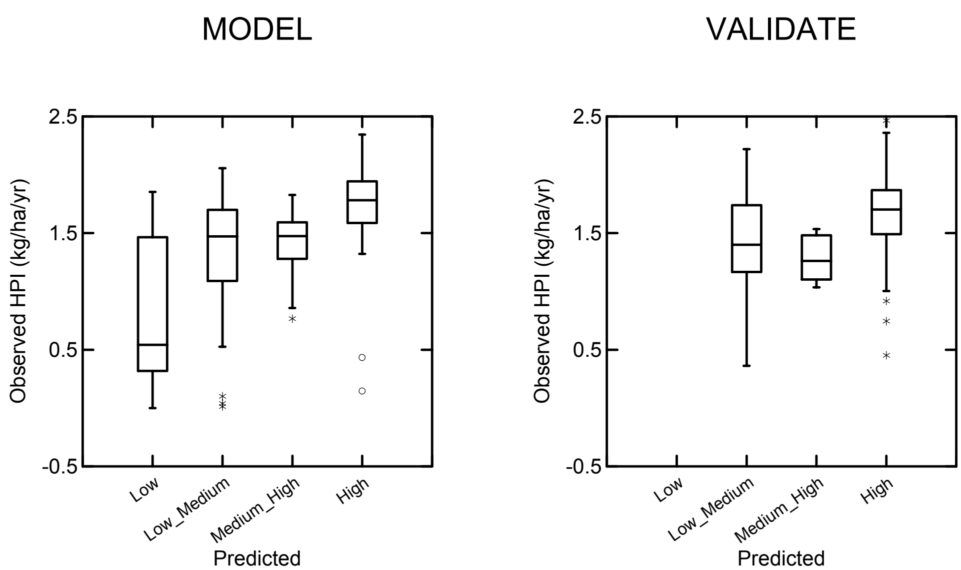


Figure 9. Upper: Regression tree (PRE = 0.38, N = 98) to classify the Habitat Productivity Index (LOG_HPI) from maximum fetch (LOG_FETCH). Lower: comparison of predicted versus observed HPI for the model and validation data sets. After pooling the medium groups, a fetch cut value of 4 km separated HPI into two categories of low and high.

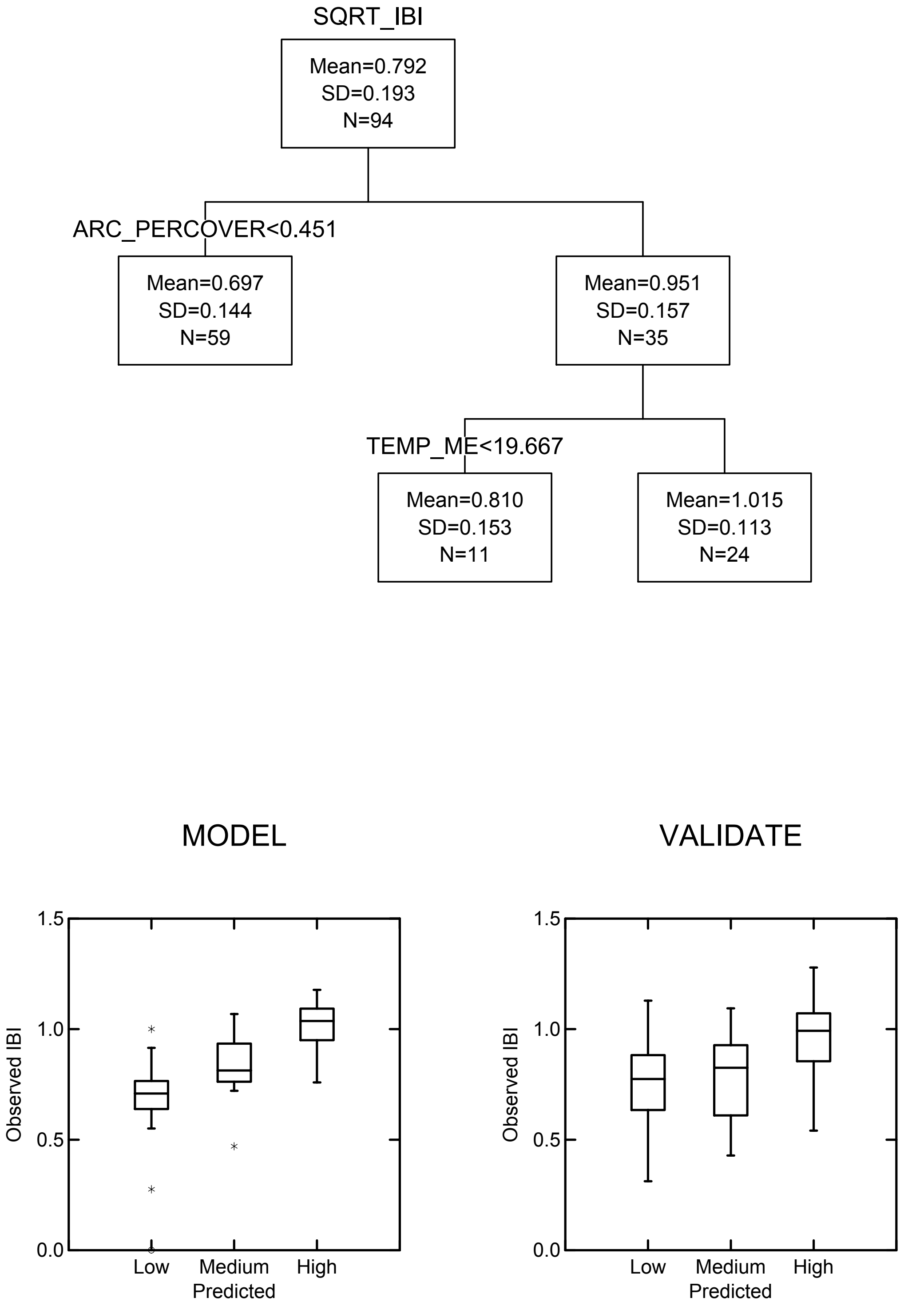


Figure 10. Upper: Regression tree (PRE = 0.50, N = 94) to classify the Index of Biotic Integrity (SQRT_IBI) using percent cover (ARC_PERCOVER) and water temperature (TEMP_ME) as predictors. Lower: comparison of predicted versus observed IBI for the model and validation data sets. A macrophyte cut value of >19% cover and water temperature >19.7°C produced the high IBI group.

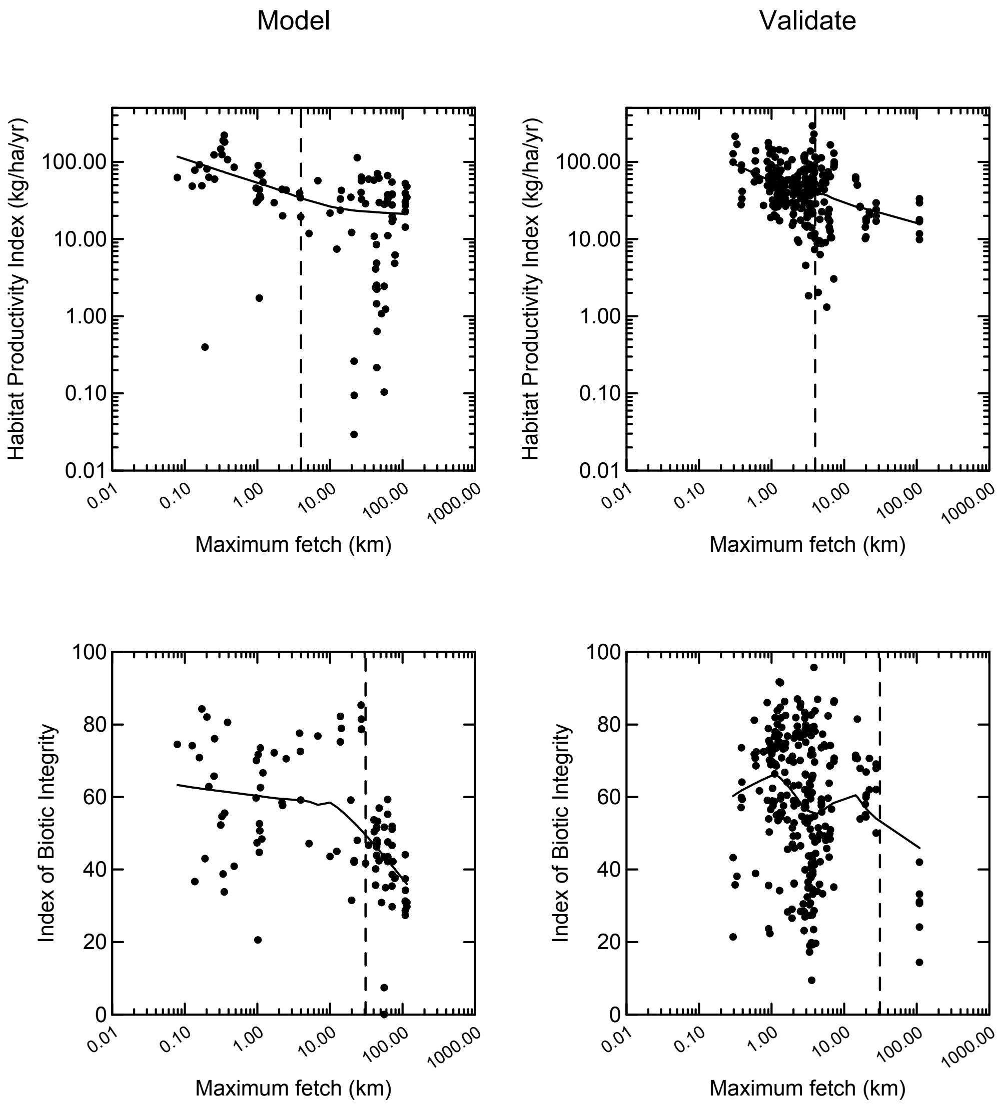


Figure 11. Relationship between HPI (upper) and IBI (lower) and coastal exposure as measured by maximum fetch in the model and validation data sets.

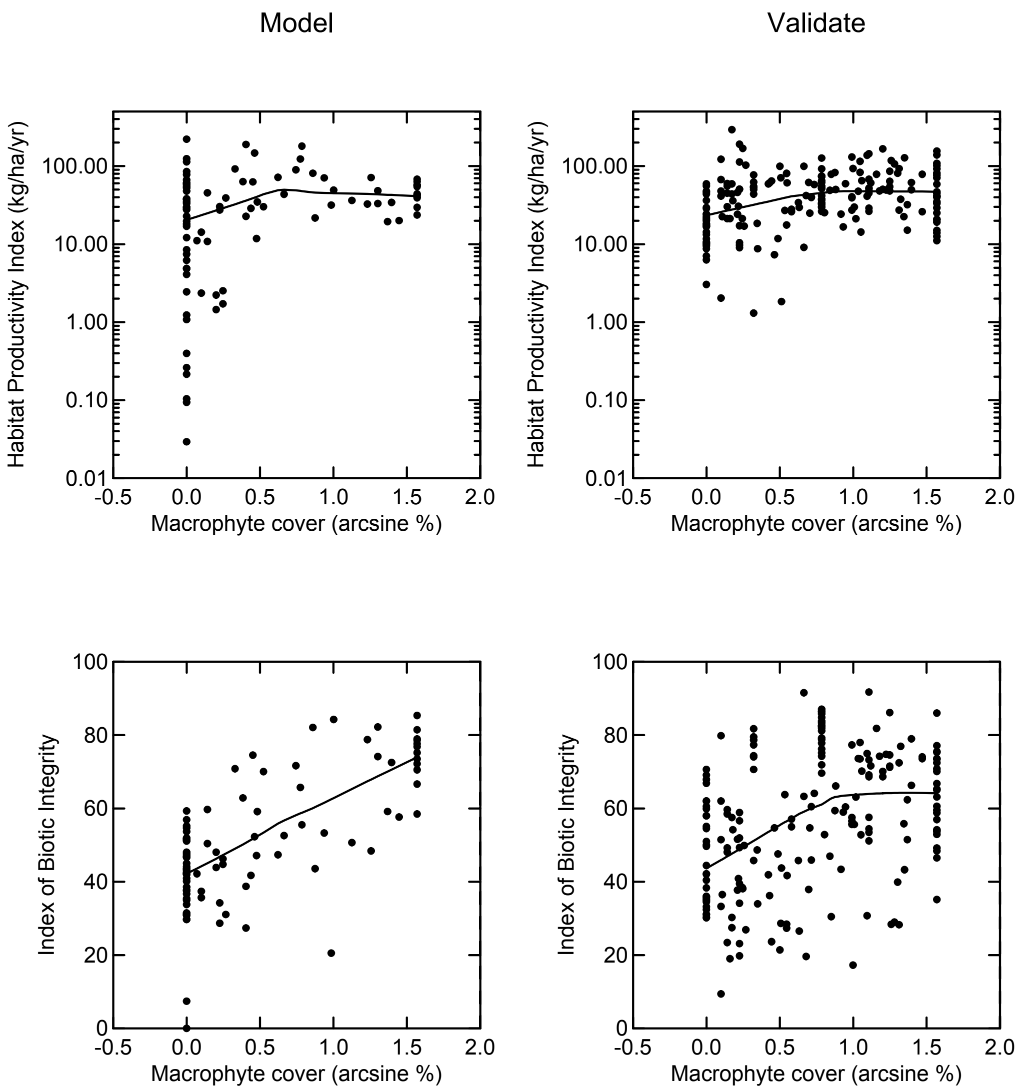


Figure 12. Relationship between HPI (upper) and IBI (lower) and macrophyte cover in the model and validation data sets.

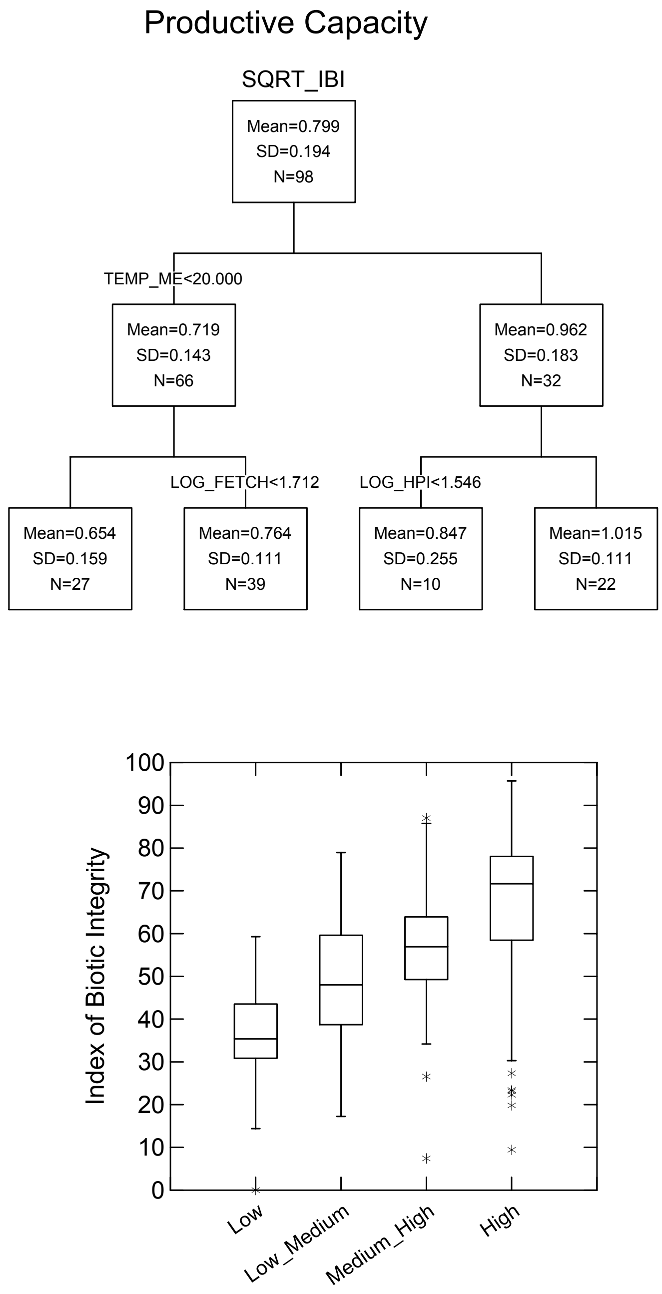


Figure 13. Two-axes approach for determining habitat productive capacity (see text). Upper: Tree diagram to classify IBI groups based on water temperature, fetch and HPI. Lower: Box plot of IBI for each of the 4 habitat classes, all data pooled (n = 342).

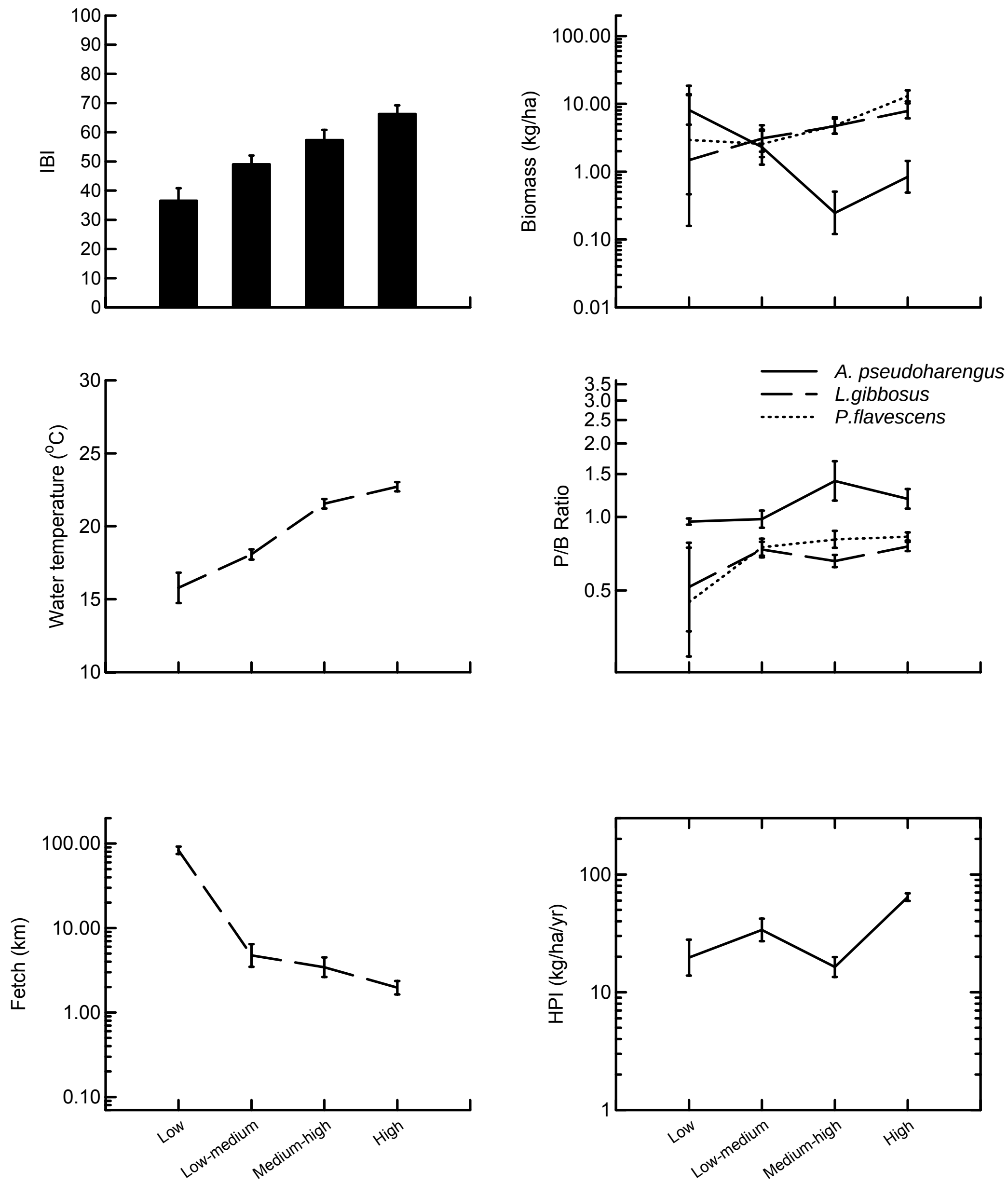


Figure 14. Summary of IBI values, water temperature, fetch, biomass and P/B ratios of *Alosa pseudoharengus*, *Lepomis gibbosus* and *Perca flavescens*, and HPI values for the four IBI groups determined by tree regression analysis.