Exploring spatial nonstationary environmental effects on species distribution: a case study of Yellow Perch in Lake Erie

Changdong Liu [1] , **Junchao Liu** [1] , **Yan Jiao** [2] , **Yanli Tang** [Corresp., 1] , **Kevin B. Reid** [3]

[1] Department of Fisheries, Ocean University of China, Qingdao, Shandong, China

[2] Department of Fish and Wildlife Conservation, Virginia Polytechnic Institute & State University, Blacksburg, Virginia, USA

[3] Department of Integrative Biology, University of Guelph, Guelph, Canada

Corresponding Author: Yanli Tang
Email address: changdong_ouc@163.com

Background: Global regression models under an implicit assumption of spatial stationarity were commonly applied to estimate the environmental effects on aquatic species distribution. However, the relationships between species distribution and environmental variables may change among spatial locations, especially at large spatial scales with complicated habitat. Local regression models are appropriate supplementary tools to explore species-environment relationships at finer scales.

Method: We applied geographically weighted regression (GWR) models on Yellow Perch in Lake Erie to estimate spatially-varying environmental effects on the presence probabilities of this species. Outputs from GWR were compared with those from generalized additive models (GAMs) in exploring the Yellow Perch distribution. Local regression coefficients from the GWR were mapped to visualize spatially-varying species-environment relationships. *K*-means cluster analyses based on the *t*-values of GWR local regression coefficients were used to characterize the distinct zones of ecological relationships.

Results: GWR resulted in a significant improvement over the GAM in goodness-of-fit and accuracy of model prediction. Results from the GWR revealed the magnitude and direction of environmental effects on Yellow Perch distribution changed among spatial location. Consistent species-environment relationships were found in the east basin for juveniles and in the west and east basins for adults. The different kinds of species-environment relationships found in the central management unit implied the variation of relationships at a scale finer than the management unit.

Conclusions: This study draws attention to the importance of accounting for spatial nonstationarity in exploring species-environment relationships. The superiority of GWR over the GAM highlights the limitations of using one global regression model to explore species-environment relationships at a large spatial scale and provides insights for managing Yellow Perch at finer scales.

Exploring spatial nonstationary environmental effects on species distribution: a case study of Yellow Perch in Lake Erie

Changdong Liu[1], Junchao Liu[1], Yan Jiao[2], Yanli Tang[1], Kevin B. Reid[3]

[1]Department of Fisheries, Ocean University of China, Qingdao, Shandong, China
[2]Department of Fish and Wildlife Conservation, Virginia Polytechnic Institute & State University, Blacksburg, VA, USA
[3]Department of Integrative Biology, University of Guelph, Guelph, Canada

Corresponding Author:
Yanli Tang[1]
No.5 Yushan Road, Qingdao, Shandong, 266003, China
Email address: changdong_ouc@163.com

Abstract

Background: Global regression models under an implicit assumption of spatial stationarity were commonly applied to estimate the environmental effects on aquatic species distribution. However, the relationships between species distribution and environmental variables may change among spatial locations, especially at large spatial scales with complicated habitat. Local regression models are appropriate supplementary tools to explore species-environment relationships at finer scales.

Method: We applied geographically weighted regression (GWR) models on Yellow Perch in Lake Erie to estimate spatially-varying environmental effects on the presence probabilities of this species. Outputs from GWR were compared with those from generalized additive models (GAMs) in exploring the Yellow Perch distribution. Local regression coefficients from the GWR were mapped to visualize spatially-varying species-environment relationships. *K*-means cluster analyses based on the *t*-values of GWR local regression coefficients were used to characterize the distinct zones of ecological relationships.

Results: GWR resulted in a significant improvement over the GAM in goodness-of-fit and accuracy of model prediction. Results from the GWR revealed the magnitude and direction of environmental effects on Yellow Perch distribution changed among spatial location. Consistent species-environment relationships were found in the east basin for juveniles and in the west and east basins for adults. The different kinds of species-environment relationships found in the central management unit implied the variation of relationships at a scale finer than the management unit.

Conclusions: This study draws attention to the importance of accounting for spatial nonstationarity in exploring species-environment relationships. The superiority of GWR over the GAM highlights the limitations of using one global regression model to explore species-environment relationships at a large spatial scale and provides insights for managing Yellow Perch at finer scales.

Introduction

Estimating the key relationships between species distribution and environmental variables is essential for natural resource conservation and ecosystem-based fishery management (*Grüss et al., 2017*). A large number of published papers reported that environmental variations caused the change of species abundance and distribution (e.g. *Tseng et al., 2013; Barbeaux & Holloweb, 2017; Liu et al., 2017; Muška et al., 2018*). Many biotic and abiotic factors, as well as the interactions among them, drive the species distribution at variable spatial-temporal scales. It is challenging to detangle the environmental effects on species distribution because of the spatial dynamic response of species to environmental variations. Exploring the environmental effects on species distribution at only one large spatial scale may mask the intrinsic relationships between them at finer scales. Accounting for spatial nonstationarity can improve our understanding of the interactive process between species distribution and environmental variables at various spatial scales (*Windle et al., 2010; Windle et al., 2012; Sadorus et al., 2014; Liu et al., 2017; Li et al., 2018; Bi et al. in press*).

Global regression models are the predominant methods to estimate environmental effects on species distribution presently. Generalized additive models (GAMs) are the popularly used methods and show priority over generalized linear models (GLMs) in estimating the nonlinear relationships between species distribution and environmental variables (e.g. *Canepa et al., 2017; Grieve et al., 2017; Hemami et al., 2018*). These models estimate one average regression parameter, independent of locations and directions, for each explanatory variable on the whole study area. Because of the complexity of aquatic ecosystem and the dynamic interaction between biology and environment, the assumption of spatial stationary relationships between biological and environmental factors may be violated, especially at large spatial scales. Local regression models can be effective complements for global models in inferring species-environment relationships at finer scales (*Fotheringham et al., 2002*).

SEM (spatial expansion model) is one of the early methods to estimate spatially-varying ecological relationships (*Fotheringham et al., 1997*). In SEM, each regression parameter itself is a function of spatial location and the form of the function (e.g. linear, polynomial) is determined by prior knowledge. The results from SEM are sensitive to the spatial expansion function and are hard to explain as the complex function (e.g. high-order polynomial) used. Geographically weight regression (GWR) model is a natural evolution of SEM. In GWR, local regression model is fitted at each sample location using its neighborhood observations. The weights of neighborhood observations on regression point are defined according to spatial dependence and the weighted least square (WLS) method can be used to fit local regression model. The main advantage of GWR is that it yields a set of parameter estimates at each sample location and the regression coefficients for each environmental variable can be mapped over the study area to visualize spatially-varying ecological relationships (*Fotheringham et al., 2002*).

Lake Erie is the smallest in volume but biologically most productive of the Laurentian Great Lakes (*Hartman, 1972*). The lake is separated into west, central and east basins as the significant environmental differences among them. Yellow Perch (*Perca flavescens*) is one of the most important commercial and recreational species in Lake Erie and plays an important role in regional economic development (*YPTG, 2015*). The variations of habitat quality cause the spatial heterogeneous distribution of Yellow Perch (*Bacheler et al., 2011; Liu et al., 2018*). Water temperature, water depth, water transparency and dissolved oxygen (DO) had been proved to be the key habitat variables to affect Yellow Perch distribution and several published studies had applied global regression models to estimate the environmental effects on Yellow Perch distribution (*Power & Heuvel, 1999; Arend et al., 2011; Bacheler et al., 2011; Yu et al., 2011; Manning et al., 2013; Liu et al., 2018*). Considering the significant environmental variations in space and the large spatial extent of Lake Erie (surface area 25,874 km^2), we expect the emergence of spatially-varying species-environment relationships in the lake. Estimating Yellow Perch distribution at finer scales can give us more insights into the dynamic interaction of Yellow Perch to environmental variations. However, we did not find the published researches on this project.

The procedure and objective of this study are that: (1) we first applied GAMs to estimate the nonlinear relationships between the presence probability of Yellow Perch and environmental factors; (2) we then used GWRs to explore spatially-varying species-environment relationships based on the same data used in the GAMs and tested the significance of spatial variation for each local regression parameter; (3) thirdly, we evaluated the goodness-of-fit and prediction accuracy of GAMs and GWRs and examined whether GWRs are better than GAMs; (4) fourthly, we interpolated and mapped the GWR local regression parameters to visualize the spatial patterns of species-environment relationships; (5) Finally, we characterized the special zones of species-environment relationships and discussed the implications for Yellow Perch management in Lake Erie.

Materials & Methods

Study area and data sources

The partnership index survey (PIS) was conducted using the standard gillnets by the Ontario Ministry of Natural Resources and Forestry and the Ontario Commercial Fisheries' Association in the Canadian waters of Lake Erie in annual late summer andearly fall. The survey gillnets comprised of 14 different mesh sizes were set on the bottom and suspended (canned) in the water column with a mean soak time of 20h. A depth-based stratified random sampling design was used and the number of sample sites was determined according to the surface area of each depth stratum (*see Berger et al., 2012; Pandit et al., 2013; Liu et al., 2018*). The PIS data from 1989 to 2015 comprised of a total of 2502 samples were analyzed in this study when removing the missing/erroneous values in the catch or environmental data (Fig.1).

For each sample, the catch weights and numbers of Yellow Perch were measured by age (age 0, age 1, age 2, etc.). The age of fish was estimated by otoliths or scales. Water temperature (℃), water transparency (m) and dissolved oxygen concentration (mg/L) were measured at the depth of gillnet (the depth from the water surface to the top of gillnet). Water transparency was estimated based on the visual distance of Secchi disk. To calculate distance, sample location coordinates denoted as longitude and latitude were converted to plane coordinates by the North American Datum 1983 Universal Transverse Mercator 17N projection. As benthivorous fish, only 6% in weight of the total catch was found in the canned gillnets, and therefore we only analyzed the bottom sampling data. The environmental effects on the Yellow Perch distribution depended on the life stages (*Liu et al., 2018*). Accordingly, we separated the fish caught into juveniles (age$<$2) and adults (age$\geqslant$2) because the age of recruitment to the Yellow Perch fishery was defined as age-2 fish (*YPTG, 2015*).

Model development

The fish caught data for each sample were simplified as 0/1 to indicate absence/presence of Yellow Perch. We built models to estimate the relationships between the presence probability of Yellow Perch and environmental factors. Water temperature, water depth, water transparency and dissolved oxygen concentration were used as explanatory variables in the model analysis because they were surveyed contemporaneously with the fish data and were proved to be the key habitat variables to Yellow Perch distribution (*Liu et al., 2018*). A preliminary variance inflation factor (VIF) analysis was conducted to test for multicollinearity of explanatory variables. The environmental factors with VIFs greater than 3 were excluded in the next model analysis (*Sagarese et al., 2014*). As all the VIFs less than 2, the four environmental variables were included in the following model analysis.

We first applied GAMs to estimate the environmental effects on the presence probabilities of juveniles and adults. GAMs extend the generalized linear models (GLMs) by replacing the linear predictors with spline functions to estimate the nonlinear relationships between response and explanatory variables (*Wood, 2006*). In the study, GAMs are denoted as:

$$\ln\left(\frac{y^*}{1-y^*}\right) = \beta_0 + \sum_{k=1}^{4} s_k(x_k) \qquad (1)$$

where y^* is the predicted presence probability, β_0 is the intercept coefficient, s is the penalized cubic regression spline function to describe the nonlinear environmental effects on the response variable, x_k is the *k*th explanatory variable. We used automatically selected degree of freedom to determine the smoothness of *s* (*Wood, 2006*). The GAM analysis was performed using the "gam" function of the "mgcv" package in the R platform and the gamma parameter was set to 1.4 to avoid overfitting (*Wood, 2014*).

The GWR model is the extension of GLM by accounting for spatial location in the parameter estimates and thus allows for exploring spatially-varying species-environment relationships. The GWR model in this study can be denoted as:

$$\ln\left(\frac{{y_i}^*}{1-{y_i}^*}\right) = \beta_0(u_i,v_i) + \sum_{k=1}^{4} \beta_k(u_i,v_i)x_k \qquad (2)$$

where ${y_i}^*$ is the predicted presence probability at location i, (u_i,v_i) is the coordinates of location i, β_0 is the intercept parameter specific to location i, β_k is the regression parameter for the *k*th environmental variables specific to location *i*. The fixed number of observations (adaptive bandwidth) nearest to the regression point are used to calibrate the local regression models in this study. The weights of observations to local parameter estimates are commonly set as decreasing with the distance to regression point and several forms of function can be used to calculate weights. We used the Gaussian weighting function (Eq.3) as its continuity easier for differential calculation.

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{h}\right) \qquad (3)$$

where d_{ij} is the Euclidean distance between the two sample sites *i* and *j*; h is the bandwidth and has a great impact on the model results. The optimal value of h was selected by minimizing the

Akaike's Information Criterion (AIC). The GWR analysis was performed based on the "GWmodel" package in the R platform.

The spatial variability of local regression parameter for each environmental variable from the GWR was estimated as the stationary index (SI) (*Brunsdon et al., 1998*). SI was calculated by dividing the interquartile range of a GWR regression coefficient by twice the *s.e.* of the same parameter estimate from the global logistic regression model (*Windle et al., 2010*). SI>1 indicates spatial non-stationarity.

The local regression parameter estimates from the GWR for juveniles (GWR-J) and adults (GWR-A) were interpolated to continuous surfaces and then mapped to visualize spatially-varying environmental effects on the presence probabilities of Yellow Perch. Lake Erie was divided into three basins as the environmental difference among them and four management units for Yellow Perch fishery (*YPTG, 2015*). In order to characterize the special zones of species-environment relationships, the *t*-values of local regression coefficients from the GWR were separated into different groups using a *k*-means cluster analysis method. The number of clusters (*k*) was set a prior to 3 and 4 for comparison with basins and management units respectively. Furthermore, the best number of clusters was estimated based on a gap statistic (*Tibshirani et al., 2001*). The spatial distribution of clusters was mapped. All the maps were produced by the ArcGIS (ESRI, v. 10.2) software.

Model evaluation and comparison

AIC and deviance explained (%) were calculated to assess goodness-of-fit for each model. The model with the lower AIC and higher deviance explained would be judged to have better fitting performance. Modelling the binary data can be treated as classified algorithm and a larger area under the receiver operating characteristic (ROC) curve (AUC) value indicated the higher discrimination accuracy (*Bradley, 1997*). To evaluate whether a model captured the spatial patterns in the response variable, we calculated Moran's I to test for the spatial autocorrelation in model residuals. Value of Moran's I close to -1 and 1 indicates strong clustering and dispersing respectively. A permutation test for Moran's I statistic was used to test for significance of spatial autocorrelation (*Bivand and Wong, 2018*).

To assess the predicted accuracy of the model, the survey data were split into training and testing data randomly as a ratio of 75%:25%. The training data were used to fit the model and the testing data were used to validate the model. AUC was used to assess the discrepancy between the predicted and observed values. The cross-validation was repeated 100 times for calculating the mean AUC value and its 95% confidence interval.

Results

Juveniles are present at 58% of sample sites, while adults are present at 90% of sample sites. The spatial distribution map indicates that juveniles are mainly distributed in the central and west basins and a significant high absence is found in the east basin. Adults are present in most sample areas and high absences are found in the deep waters of the east basin and near-shore areas in the central basin (Fig.2).

GAM results show that water temperature, water depth, water clarity and dissolved oxygen have significant effects on the presence probability of juveniles, yet only the first three variables significantly affect the adults distribution (p<0.01). The presence probability of juveniles significantly increases with water temperature and dissolved oxygen, decreases with water clarity, and first increases and then decreases with water depth. The presence probability of adults shows similar change trend with that of juveniles to the variation of water temperature, water depth and dissolved oxygen.

Based on the AIC criteria, GWRs with adaptive bandwidths of 64 and 241 points have the best performances for juveniles and adults, respectively. GWRs result in significant decreases of AIC values and increases of deviance explained indicating better goodness-of-fit compared with the equivalent GAMs. GWRs also present the high prediction accuracy indicating by the higher AUC values than the equivalent GAMs. Moran's I test results show that spatial autocorrelations of model residuals from the GAMs and GWRs are not significant, implying the two types of models can capture the spatial patterns of the response variable (Table 1).

Descriptive statistics of local regression coefficient estimates from the GWRs reveal the much variations of coefficient values. SI values are all greater than 1 indicating the significantly spatial nonstationary relationships between the presence probability of Yellow Perch and environmental variables (Table 2). The estimated coefficient values of water temperature, water depth, water clarity and DO for juveniles from the GWR vary between -0.40-0.28, -0.16-0.47, -0.86-0.25 and -0.51-0.55, respectively (Fig.4). Though the positive associations between water temperature and the presence of juveniles found in most areas, the strong negative associations are present in the east basin and the middle areas of the central basin (Fig.4a). The presence of juveniles is positively correlated with water depth in the west basin and negatively correlated with water depth in the deep waters of east basin (Fig.4b). Water clarity and DO present strong negative and positive effects on the presence of juveniles respectively in the west basin (Fig.4c, d). The estimated coefficient values of water temperature, water depth, water clarity and DO for adults from the GWR vary between -0.038-0.69, -0.15-0.33, -1.0-0.027 and -0.12-0.41, respectively (Fig.5). The presence of adults increases with water temperature in almost all areas (Fig.5a). Water depth provides positive effects in the west and central basins and negative effects in the east basin on the presence of adults (Fig.5b). The negative associations between water clarity and presence of adults are present in all the areas except a small section in the east basin (Fig.5c). The strong negative associations of DO with the presence of adults are found in the east basin and the strong positive associations are found in the west basin and the west of central basin (Fig.5d).

The k-means cluster analysis of t-values of local regression coefficients from the GWRs characterized the special zones of environmental effects on the Yellow Perch distribution (Fig.6, 7). The k-means cluster analysis when k=2 indicates the species-environment relationships for juveniles in the central of Lake Erie with relative deep waters are specialized as cluster 1 and the rest of Lake Erie is specialized as cluster 2 (Fig.6a). As k changed from 2 to 3, the areas of cluster 1 do not change and the areas of cluster 2 are further divided into two groups. The consistent species-environment relationships are found in the west basin and in most areas of central and east basins (Fig.6b). As *k* changed from 3 to 4, the areas of cluster 1 and 2 change little and the areas of cluster 3 are further divided into two groups. The ecological relationships for juveniles in each management unit are not classified as one group (Fig.6c). The *k*-means cluster analysis of *k*=2 divides Lake Erie into distinct longitudinal zones of environmental effects on the adult distribution (Fig.7a). The areas of cluster 1 are further cut into two adjacent parts as *k* changed from 2 to 3. The west and east basins show consistent species-environment relationships (Fig.7b). As *k* changed from 3 to 4, the east areas (cluster 2) are further separated into two adjacent parts and the two groups located in the west of Lake Erie do not change. The boundary of cluster 2 is consistent with that of management unit 4 (Fig.7c). Based on the gap statistics, the best numbers of clusters for juveniles and adults are both two.

Discussions

Water temperature is an essential factor for the growth of juvenile Yellow Perch. Juveniles prefer to live in the warmer waters when the water temperature below the optimal range (20.0 to 23.3℃) (*McCauley & Read, 1973; Power & Heuvel, 1999*). This was verified by the GWR results of the presence of juveniles increasing with water temperature in the cold waters of eastern Lake Erie. By contrast, GWR results also proved that the presence of juveniles decreased with water temperature as it over the optimal range in the west basin. GAM pooled all the survey data and got a mean trend in the association of the presence of juveniles with water temperature, which masked the interaction between water temperature and juvenile distribution at finer scales.

Water depth and water clarity in Lake Erie increase from west to east. According to the GAM results, the presence of Yellow Perch first increased and then decreased with water depth as it over 20m. This result projected to space by GWR was that the presence of Yellow Perch increased with water depth in the west and central basins with shallow water and decreased with water depth in the east basin with deep water. Juveniles prefer to inhabit in the shallower, more turbid waters for avoiding pelagic, visual predators (*Manning et al., 2013*). This finding was verified by the GAM results of the significant decrease of Yellow Perch presence with increasing water clarity. However, clearer waters are good for the growth of juveniles by improving the visual field and increasing the foraging success rates (*Manning et al., 2013*). This is probably the reason that the presence of juveniles increasing with the water clarity in parts of Lake Erie based on the GWR results.

Dissolved oxygen concentrations below threshold or fluctuating diurnally are not conducive to the growth of juveniles (*Bejda et al., 1992*). Hypolimnetic hypoxia (<2 mg O_2/L^{-1}) can cause Yellow Perch to avoid hypoxic habitats to more oxygenated areas and alter the fish distribution (*Roberts et al., 2012*). Over 99% of the sample sites have dissolved oxygen concentrations above the hypoxic threshold and this is probably the reason to cause the insignificant effect of dissolved oxygen on the presence of adults. *Liu et al. (2018)* also found dissolved oxygen did not affect adult Yellow Perch distribution significantly. GAM results indicated the general trend of the presence of juveniles significantly increasing with dissolved oxygen concentration. This finding may not be appropriate for applying at the local scale. Juveniles prefer to live in the more oxygenated areas for optimizing the growth in the shallower, warmer waters. However, as dissolved oxygen concentration over a certain value, it is not an important factor to affect juvenile's distribution. *Liu et al. (2018)* found the significant interactive effect of dissolved oxygen with water depth on the distribution of juvenile Yellow Perch in Lake Erie. GWR results proved that the effect of dissolved oxygen on juvenile's distribution depending on water depth and are consistent with the findings of the published research.

Our cluster analysis characterized special zones of species-environment relationships. *Liu et al. (2017)* achieved similar results in analyzing the relationships between walleye distribution and environmental factors in Lake Erie. In order to detect whether a consistent species-environment relationship exists in each basin, we divided the local regression coefficients of GWR into three groups based on the *k*-means cluster analysis. Consistent ecological relationships were found in the west basin for juveniles and in the west and east basins for adults. The distinctive environmental attributes with warmer, shallower, more turbid and colder, deeper, clearer waters in the west and east basins respectively may be the reasons to shape the special zones of ecological relationships. Lake Erie was partitioned into four management units (MUs) and total allowable catch (TAC) of Yellow Perch was allocated based on MUs each year (*YTPG, 2015*). The MU boundaries were identified with full consideration of socioeconomic concerns (e.g., at least one major port exists within each MU) and political boundaries (e.g., counties in Ontario) (*Kocovsky & Knight, 2012*). Hence, MUs are convenient for landing and reporting of harvest and may lack of ecological significance to some degree. When comparing the *k*-means cluster analysis (*k*=4) results for adults with MUs, consistent species-environment relationships were found in MU1 and MU4 and two different kinds of species-environment relationships were found in MU2 and MU3. This implied the variation of species-environment relationships at a scale finer than the management unit. *Kocovsky & Knight (2012)* provided the morphological evidence of discrete stocks of Yellow Perch at management unit scale. Comprehensive analysis with additional explanatory variable included in the GWR in combination with genetic research can be used to refine the current MU structure in consideration of ecological relevance for sustainable management of Yellow Perch.

The predominant advantage of GWR is the ability to capture the spatially-varying ecological relationships. Furthermore, GWR can be used as an identifier to determine at which scale the species-environment relationships become stationary (*Windle et al., 2010*). Although the superiority of GWR over the global regression models, it should be used with cautions. Due to local regression coefficients estimated based on the neighborhood observations, GWR cannot be used to predict species distribution outside the study area. Spatial coordinates are the only information required by GWR to estimate local regression coefficients at unobserved locations. Thus, GWR cannot be used to predict future distribution of species. The possible collinearity in local regression coefficients may limit the interpretation of species-environment relationships (*Wheeler & Tiefelsdorf, 2005*). Attention should be given when including multi-level categorical variable (e.g. year in this study) in the GWR because of the strong risk to cause collinearity in the local regression coefficients. The prediction accuracy of GWR is sensitive to data quantity. Thus, developing the GWR separately for each year in this study may not be sufficient to get ecologically meaningful results. The large data quantity required to estimate local regression coefficients limits the application of GWR.

Conclusions

Though the convenience in the statistical test of ecological relationships, developing a global regression model by pooling all the survey data in the large region may mask the local variability in the processes being studied. We applied the GWR to question the assumption of spatial stationarity in estimating the relationships between Yellow Perch distribution and environmental variables in Lake Erie. The superiority of GWR over the GAM highlights the limitations of using one global regression model to explore species-environment relationships at a large spatial scale. The results from GWR provide insights for managing Yellow Perch at finer scales. The zonation of species-environment relationships supports informative views for refining the current MUs in consideration of ecological significance. Though some limitations, GWR has been recommended as a complementary tool for global regression models in exploring spatially-varying ecological relationships. To the end, an expanded research was prepared to explore the spatio-temporal nonstationary species-environment relationships for Yellow Perch in Lake Erie using a geographically and temporally weighted regression (GTWR) model.

Acknowledgements

This project was supported in part by a grant ‘‘Integration of spatial stock structure and multiple stocks into stock assessment for Yellow Perch in Lake Erie’’ to Y. Jiao. We thank the Ontario Commercial Fisheries Association and Great Lakes Fishery Commission for providing data to us. We also acknowledge the support from Department of Fisheries, Ocean University of China.

References

Arend KK, Beletsky D, Depinto JV, Ludsin SA, Roberts JJ, Rucinski DK, Scavia D, Schwab DJ, HÖÖK TO. 2011. Seasonal and interannual effects of hypoxia on fish habitat quality in central Lake Erie. *Freshwater Biology* **56**: 366-383. DOI 10.1111/j.1365-2427.2010.02504.x

Canepa A, Fuentes V, Bosch-Belmar M, Acevedo M, Toledo-Guedes K, Ortiz A, et al. 2017. Environmental factors influencing the spatiotemporal distribution of Carybdea marsupialis (Lineo, 1978, Cubozoa) in South-Western Mediterranean coasts. PLoS ONE 12: e0181611. DOI 10.1371/journal.pone.0181611.

Bacheler NM, Paoli TJ Schacht GM. 2011. Controls on abundance and distribution of Yellow Perch: predator, water quality, and density-dependent effects. *Transactions of the American Fisheries Society* **140**: 989-1000 DOI 10.1080/00028487.2011.603979

Barbeaux SJ, Holloweb AB. 2017. Ontogeny matters: Climate variability and effects on fish distribution in the eastern Bering Sea. *Fisheries Oceanography* **27:**1-15. DOI 10.1111/fog.1222.

Bejda AJ, Phelan BA, Studholme AL. 1992. The effect of dissolved oxygen on the growth of young-of-the-year winter flounder, *Pseudopleuronectes americanus*. *Environmental Biology of Fishes* **34**: 321. DOI: 10.1007/BF00004780.

Berger AM, Jones ML, Zhao Y. 2012. Improving fishery independent indices of abundance for a migratory walleye population. *Journal of Great Lakes Research* **38**: 755-765. DOI 10.1016/j.jglr.2012.09.012.

Bivand RS, Wong DWS. 2018. Comparing implementations of global and local indicators of spatial association. *Test* **27**: 716-748. DOI 10.1007/s11749-018-0599-x

Bradley A. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* **30**: 1145-1159. DOI 10.1016/S0031-3203(96)00142-2.

Fotheringham AS, Charlton M, Brunsdon C. 1997. Two techniques for exploring non-stationarity in geographical data. *Geographical System* **4:** 59-82.

Fotheringham AS, Charlton ME, Brunsdon C. 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A* 30: 1905-1927. DOI 10.1068/a301905.

Fotheringham AS, Brunsdon C, Charlton M. 2002. Geographically Weighted Regression: the

Analysis of Spatially Varying Relationships. John Wiley & Sons, Chichester, England.

Grieve BD, Hare JA , Saba VS. 2017. Projecting the effects of climate change on *Calanus finmarchicus* distribution within the U.S. Northeast Continental Shelf**.** *Scientific Reports* **7:**1-12**.** DOI 10.1038/s41598-017-06524-1.

Grüss A, Thorson JT, Babcock EA, Tarnecki JH. 2017. Producing distribution maps for informing ecosystem-based fisheries management using a comprehensive survey database and spatio-temporal models. *ICES Journal of Marine Science* **75**:1-20 DOI 10.1093/icesjms/fsx120.

Hartman WL. 1972. Lake Erie: effects of exploitation, environmental changes, and new species on the fishery resources. *Journal of the Fisheries Research Board of Canada* **29**:931–936.

Hemami MR, Ahmadi M, Sadegh-Saba M, Moosavi MH. 2018. Population estimate and distribution pattern of Indian Ocean humpback dolphin (*Sousa plumbea*) in an industrialised bay, northwestern Persian Gulf. *Ecological Indicators* **89**:631-638 DOI 10.1016/j.ecolind.2018.02.031.

Kocovsky PM, Knight CT. 2012. Morphological evidence of discrete stocks of yellow perch in Lake Erie. *Journal of Great Lakes Research* **38**:534-539. DOI 10.1016/j.jglr.2012.04.006

Liu C, Wan R, Jiao Y, Kevin B. R. 2017. Exploring non-stationary and scale-dependent relationships between walleye (*Sander vitreus*) distribution and habitat variables in Lake Erie. *Marine & Freshwater Research* **68**:270-281. DOI 10.1071/MF15374. DOI 10.1007/s10750-017-3440-8.

Li B, Cao J, Guan L, Mazur M, Chen Y, Wahle, RA. 2018. Estimating spatial non-stationary environmental effects on the distribution of species: a case study from American lobster in the Gulf of Maine. *ICES Journal of Marine Science* **75**:1473–1482. DOI 10.1093/icesjms/fsy024.

Li Y, Jiao Y, Browder JA. 2016. Modeling spatially-varying ecological relationships using geographically weighted generalized linear model: A simulation study based on longline seabird bycatch. *Fisheries Research* **181**:14-24. DOI 10.1016/j.fishres.2016.03.024.

Liu CD, Wan R, Jiao Y, Reid KB. 2017. Exploring non-stationary and scale-dependent relationships between walleye (*Sander vitreus*) distribution and habitat variables in Lake Erie. *Marine and Freshwater Research* **68**:270-281. DOI 10.1071/MF15374

Liu CD, Jiao Y, Reid KB, Tang YL, Ma SY. 2018. Habitat supply for Yellow Perch (*Actinopterygii, Percidae*) varies with space, time, and life stage in Lake Erie. *Hydrobiologia* **808**:371-386. DOI 10.1007/s10750-017-3440-8.

Manning NF, Mayer CM, Bossenbroek JM, Tyson JT. 2013. Effects of water clarity on the length and abundance of age-0 Yellow Perch in the Western Basin of Lake Erie. *Journal of Great Lakes Research* **39**: 295-302. DOI 10.1016/j.jglr.2013.03.010.

McCauley RW, Read LAA. 1973. Temperature selection by juvenile and adult yellow perch (*Perca flavescens*) acclimated to 24℃. *Journal of the Fisheries Research Board of Canada* **30**:1253-1255.

Muška M, Tušer M, Frouzová J, Mrkvička T, Ricard D, Seda J, Morelli F, Kubečka J. 2018. Real-time distribution of pelagic fish: combining hydroacoustics, gis and spatial modelling at a fine spatial scale. *Scientific Reports* **8**:5381.

Pandit SN, Zhao YM, Ciborowski JJH, Gorman AM, Knight CT. 2013. Suitable habitat model for walleye (*Sander vitreus*) in Lake Erie: implications for inter-jurisdictional harvest quota allocations. *Journal of Great Lakes Research* **39**: 591-601. DOI 10.1016/j.jglr.2013.09.011.

Power M, Heuvel MR. 1999. Age-0 Yellow Perch growth and its relationship to temperature. *Transactions of the American Fisheries Society* **128**: 687-700. DOI 10.1577/1548-8659(1999)128<0687:aypgai>2.0.co;2

Roberts JJ, Grecay PA, Ludsin SA , Pothoven SA, Vanderploeg HA, Höök TO. 2012. Evidence of hypoxic foraging forays by yellow perch (*Perca flavescens*) and potential consequences for prey consumption. *Freshwater Biology* **57**:922-937. DOI 10.1111/j.1365-2427.2012.02753.x

Sadorus LL, Mantua NJ, Essington T, Hickey B, Hare S. 2014. Distribution patterns of Pacific halibut (*Hippoglossus stenolepis*) in relation to environmental variables along the continental shelf waters of the US West Coast and southern British Columbia. *Fisheries Oceanography* **23**:225-241. DOI 10.1111/fog.12057.

Tibshirani R, Walther G, Hastie T. 2001. Estimating the number of data clusters via the gap statistic. *Journal of the Royal Statistical Society. Series B. Methodological* **63**: 411-423. DOI 10.1111/1467-9868.00293.

Tseng CT, Su NJ, Sun CL, Punt AE, Yeh SZ, Liu DC, Su WC. 2013. Spatial and temporal variability of the Pacific saury (*Cololabis saira*) distribution in the northwestern Pacific Ocean. *ICES Journal of Marine Science* **70**: 991–999. DOI:10.1093/icesjms/fss205.

Wheeler D, Tiefelsdorf M. 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems* **7**:161–187.

DOI 10.1007/S10109-005-0155-6.

Windle MJ, Rose GA, Devillers R, Fortin MJ. 2010. Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic. *ICES Journal of Marine Science* **67**:145154. DOI 10.1093/ICESJMS/FSP224.

Windle MJ, Rose GA, Devillers R, Fortin MJ. 2012. Spatiotemporal variations in invertebrate-cod-environment relationships on the Newfoundland-Labrador Shelf, 1995-2009. *Marine Ecology Progress Series* **469**:263-278. DOI 10.3354/MEPS10026

Wood S. 2006. Generalized Additive Models: an Introduction with R. CRC Press: Boca Raton, FL.

Wood S. 2014. mgcv: mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation. R package ver. 1.8–4. Available at http://CRAN.R-project.org/package=mgcv.

YPTG (Yellow Perch Task Group). 2015. Report of the Lake Erie Yellow Perch Task Group. Wheatley, Ontario.

Yu H, Jiao Y, Winter A. 2011. Catch-Rate standardization for Yellow Perch in Lake Erie: a comparison of the spatial generalized linear model and the generalized additive model. *Transactions of the American Fisheries Society* **140**:905-918. DOI 10.1080/00028487.2011.599258.

Zuur AF, Ieno EN, Elphick CS. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* **1**: 3-14. DOI 10.1111/j.2041-210X.2009.00001.x.

Table 1(on next page)

Table 1 Summary of optimal bandwidths and model performances for generalized additive models (GAMs) and geographically weighted regression (GWR) models.

The unit of bandwidth is the number of points. J denotes juveniles, A denotes adults. AIC is Akaike's information criterion. AUC is area under the receiver operating characteristic (ROC) curve. CV_AUC±SD is the mean AUC±standard deviance calculated based the 100 repeated cross-validations. Moran test is the *p*-values of testing for the significance of residual spatial autocorrelations.

Model	Bandwidth	AIC	Deviance (%)	AUC	CV_AUC±SD	Moran test
GAM-J	-	2955.6	14.3	0.73	0.72±0.02	0.18
GAM-A	-	1018.8	36.2	0.88	0.74±0.02	0.51
GWR-J	64	2809.9	23.2	0.80	0.81±0.01	0.96
GWR-A	241	982.3	41.5	0.91	0.90±0.02	0.86

Table 2 (on next page)

Table 2. Summary statistics of the logistic GWR local parameter estimates and spatial stationarity index (SI).

J denotes juveniles, A denotes adults, DO denotes dissolved oxygen. SI was calculated by dividing the interquartile range of a GWR regression coefficient by twice the s.e. of the same parameter estimate from the global logistic regression model. SI>1 indicates spatial non-stationarity.

Model	Variable	Minimum	Lower quartile	Median	Upper quartile	Maximum	SI
GWR-J	Intercept	-9.37	-2.47	-0.27	1.40	6.09	4.68
	Temperature	-0.40	-0.06	0.04	0.10	0.28	4.86
	Depth	-0.16	0.00	0.06	0.08	0.12	8.23
	Transparency	-0.86	-0.23	-0.09	0.05	0.25	5.67
	DO	-0.51	-0.11	-0.02	0.12	0.55	5.59
GWR-A	Intercept	-10.34	-4.57	-0.70	3.64	5.64	5.76
	Temperature	-0.04	0.09	0.14	0.40	0.69	3.70
	Depth	-0.15	-0.08	0.06	0.18	0.33	11.01
	Transparency	-1.01	-0.63	-0.36	-0.26	0.03	5.73
	DO	-0.12	-0.07	0.01	0.22	0.41	4.13

3

Figure 1

Figure 1 Sample sites of the partnership index survey (PIS) in the Canadian side of Lake Erie from 1989 to 2015.

The middle thick line through the lake represents the Canada-USA border. The bold black lines in Lake Erie are the separate lines among basins.

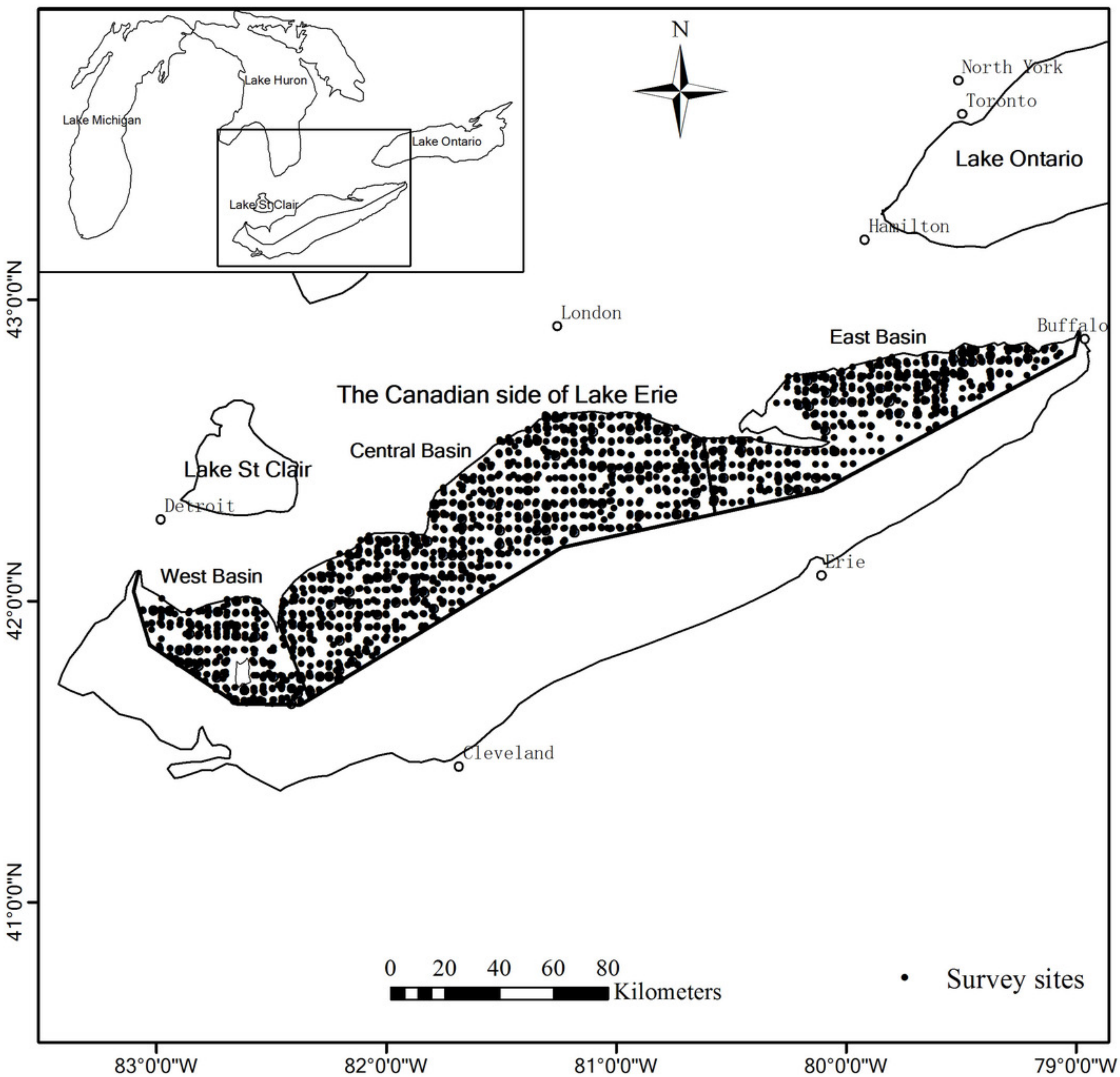

Figure 2

Figure 2 Spatial distributions of absence (o) and presence (x) for (a) juvenile and (b) adult Yellow Perch in the Canadian side of Lake Erie based on the partnership index survey (PIS) data.

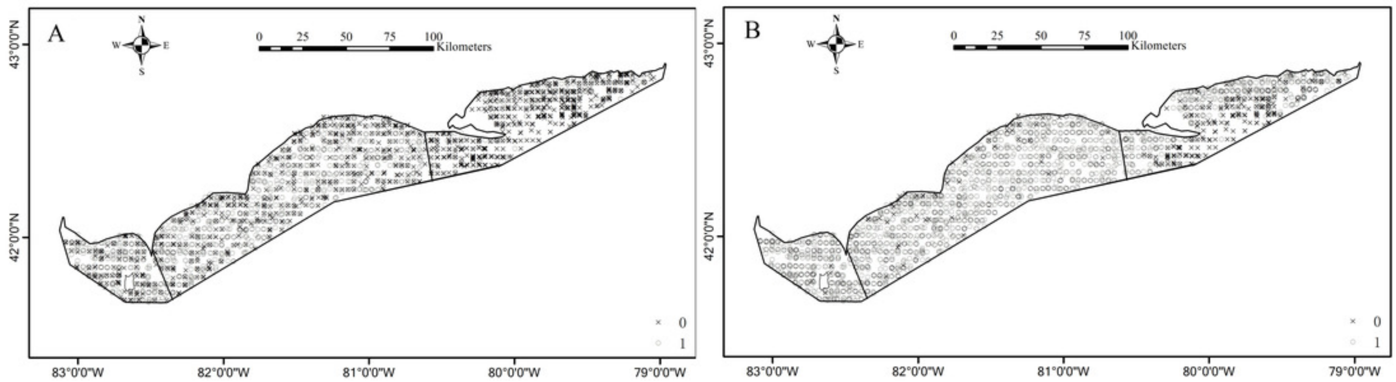

Figure 3

Figure 3 Environmental effects on the presence probabilities of (a) juveniles and (b) adults based on the generalized additive models (GAMs).

Tick marks on the x-axis are observed data points; s(x) represents the cubic spline function; and shaded areas indicate 95% confidence bounds.

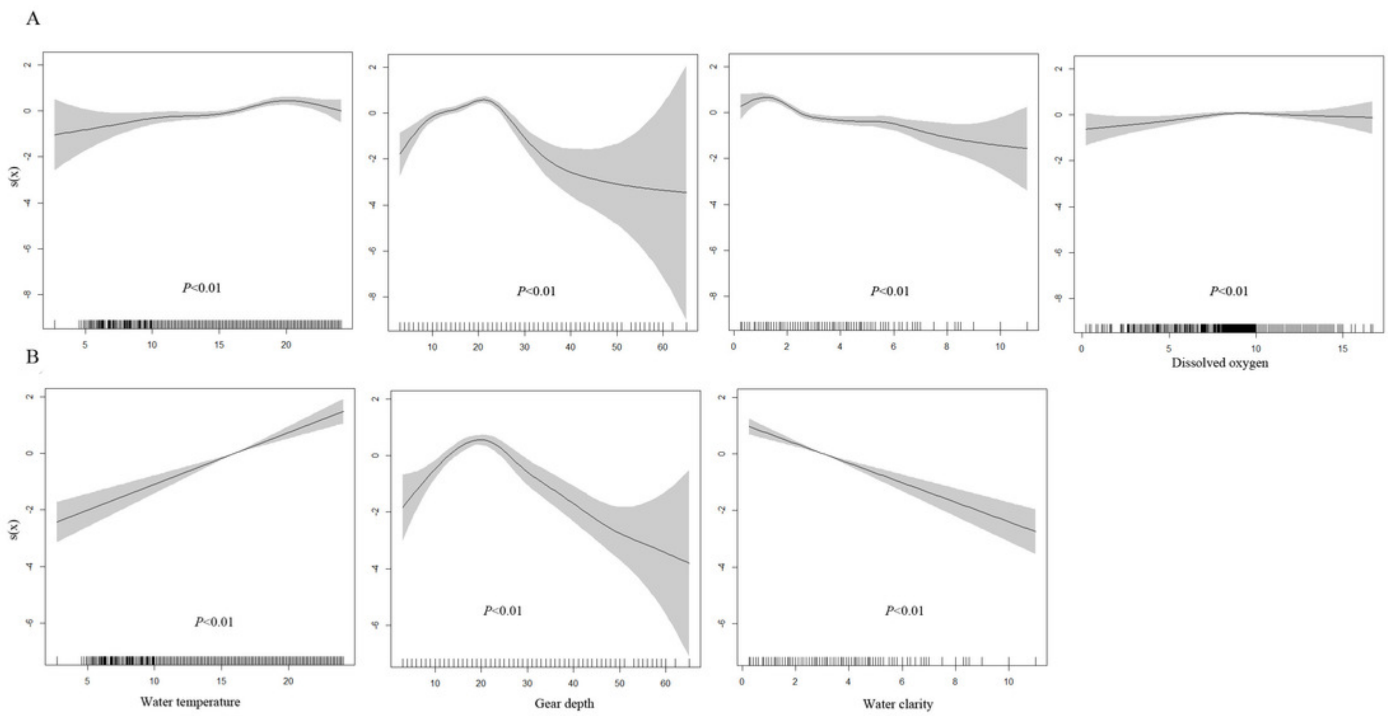

Figure 4

Figure 4 The interpolated continuous surfaces of the GWR local regression coefficient estimates for juveniles for (a) water temperature, (b) water depth, (c) water transparency, and (d) dissolved oxygen.

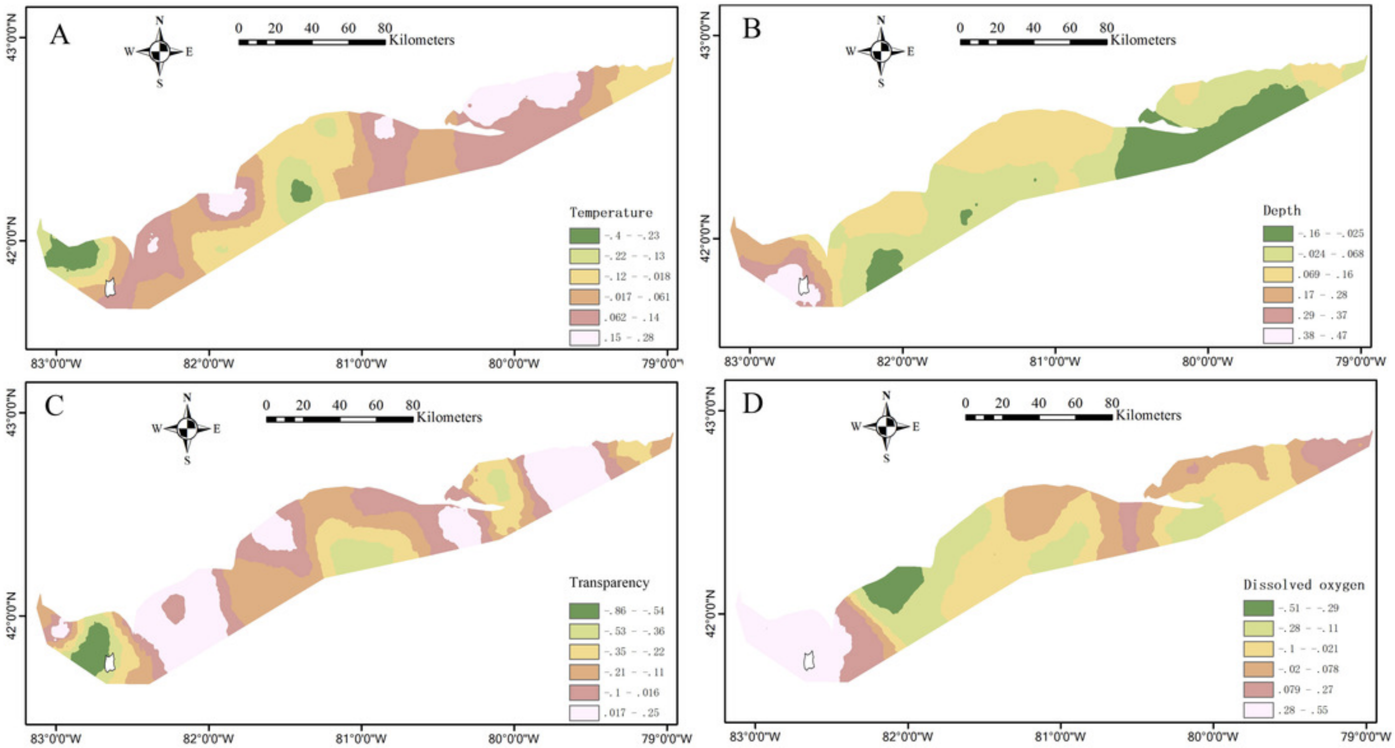

Figure 5

Figure 5 The interpolated continuous surfaces of the GWR local regression coefficient estimates for adults for (a) water temperature, (b) water depth, (c) water transparency, and (d) dissolved oxygen.

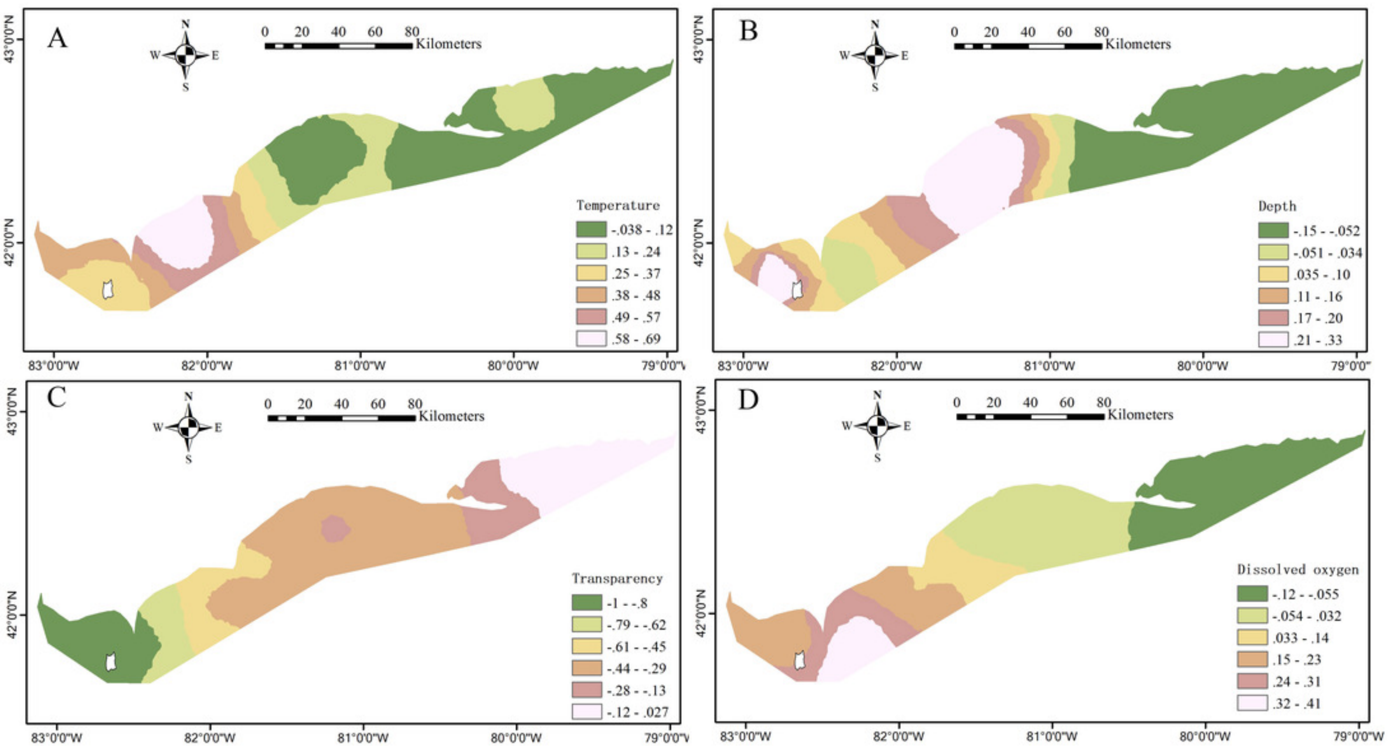

Figure 6

Figure 6 Mapped results of *k*-means cluster analyses of the pseudo *t*-vales from the logistic GWR local coefficient estimates for juveniles, for three clusters, (a) *k*=2, (b) *k*=3, (c) *k*=4.

The bold black lines in (b) and (c) are the separate lines among basins and management units, respectively.

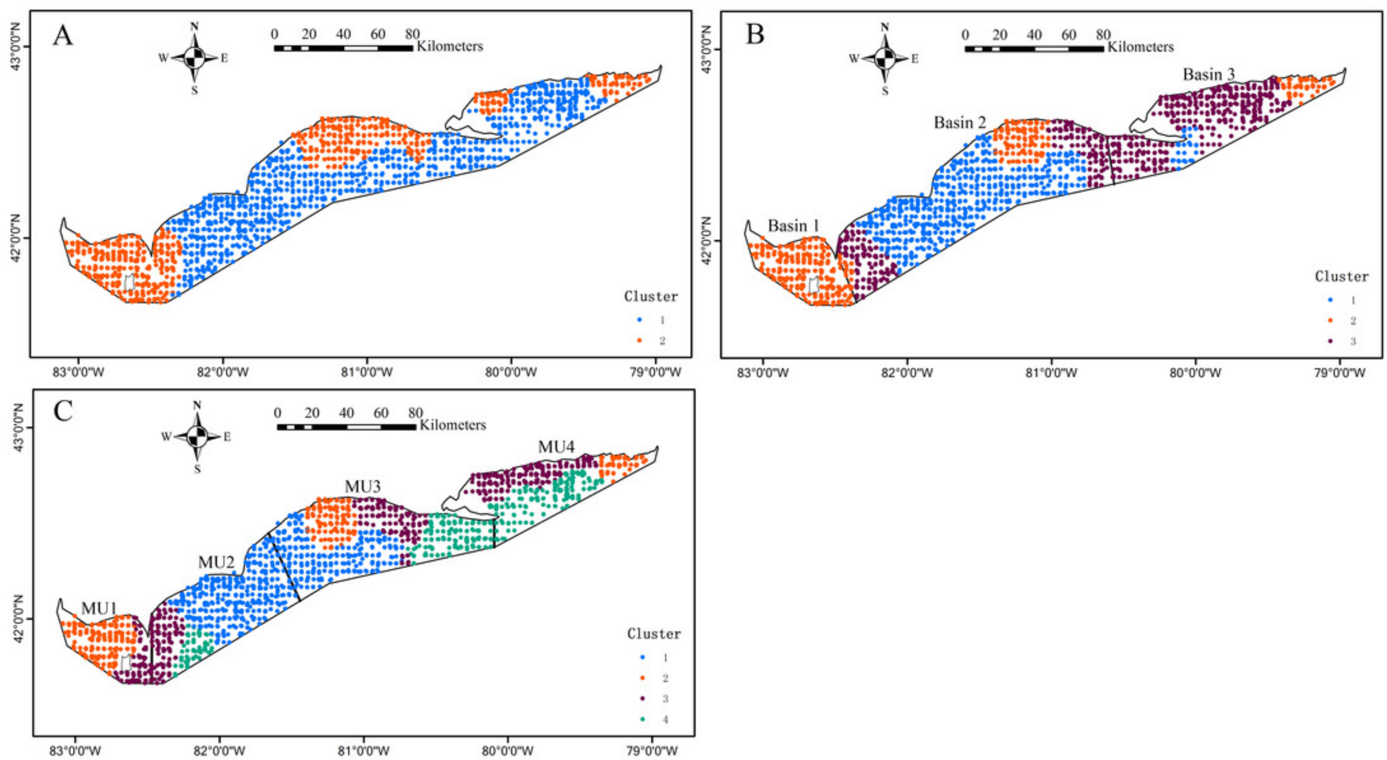

Figure 7

Figure 7 Mapped results of *k*-means cluster analyses of the pseudo *t*-vales from the logistic GWR local coefficients estimates for adults, for three clusters, (a) *k*=2, (b) *k*=3, (c) *k*=4.