Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic

Matthew J. S. Windle, George A. Rose, Rodolphe Devillers, and Marie-Josée Fortin

Analyses of fisheries data have traditionally been performed under the implicit assumption that ecological relationships do not vary within management areas (i.e. assuming spatially stationary processes). We question this assumption using a local modelling technique, geographically weighted regression (GWR), not previously used in fisheries analyses. Outputs of GWR are compared with those of global logistic regression and generalized additive models (GAMs) in predicting the distribution of northern cod off Newfoundland, Canada, based on environmental (temperature and distance from shore) and biological factors (snow crab and northern shrimp) from 2001. Results from the GWR models explained significantly more variability than the global logistic and GAM regressions, as shown by goodness-of-fit tests and a reduction in the spatial autocorrelation of model residuals. GWR results revealed spatial regions in the relationships between cod and explanatory variables and that the significance and direction of these relationships varied locally. A k-means cluster analysis based on GWR t-values was used to delineate distinct zones of species–environment relationships. The advantages and limitations of GWR are discussed in terms of potential application to fisheries ecology.

M. J. S. Windle and G. A. Rose: Fisheries Conservation Group, Marine Institute, Memorial University of Newfoundland, St John's, NL, Canada A1C 5R3. R. Devillers: Department of Geography, Memorial University of Newfoundland, St John's, NL, Canada A1C 5R3. M-J. Fortin: Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada M5S 3G5. Correspondence to M. J. S. Windle: tel: +1 709 778 0504; fax: +1 709 778 0669; e-mail: matt.windle@mi.mun.ca.

Introduction

Fisheries researchers are faced with the challenging task of studying complex patterns and processes in marine resources that occur over large spatial scales (Ricklefs, 1990). These processes are often examined using population and environmental variables averaged over management units, resulting in a single "global" model applied to an entire study region (Ciannelli et al., 2008). Although global models are useful for testing hypotheses, they function under the assumption of spatial stationarity in the processes under study, whereby the parameters of a process (e.g. variance and mean) are independent of location and direction (Fortin and Dale, 2005). Given the highly mobile nature of fisheries resources and the dynamic spatial interactions between biological and environmental variables in marine ecosystems (Ottersen et al., 2004; Rose, 2005; Ciannelli et al., 2008), stationarity of these processes may be the exception rather than the rule.

As the presence of spatial stationarity is required to make accurate model inferences at unsampled locations (Fortin and Dale, 2005), it is important to investigate whether the influences of environmental and trophic variables are relatively uniform across a study area. One possible solution is the use of local spatial statistics that highlight differences across space (Fotheringham et al., 2002). A relatively new local spatial technique that addresses both spatial heterogeneity (e.g. spatial non-stationarity) and spatial dependence (i.e. spatial dependence and spatial autocorrelation) in ecological data is geographically weighted regression (GWR; Brunsdon et al., 1998). In contrast to regression models that estimate one set of global parameters, GWR estimates a set of local parameter coefficients for each observation point by specifically incorporating the geographic locations of all sampled locations in the model (Dormann et al., 2007; Fortin and Melles, 2009). Observations spatially closer to the location being predicted are given more weight than those farther away. GWR therefore provides a method of exploring how regression model parameters vary across space (i.e. spatial non-stationarity in the processes under study) and represents a spatial modification to normal techniques, such as ordinary least-squares (OLS) regression. Indeed, GWR outperforms other global methods such as OLS, generalized linear models (GLMs), generalized additive models (GAMs), and linear mixed models in the context of modelling animal and vegetation distributions in terrestrial ecology (Zhang and Gove, 2005; Shi et al., 2006; Kupfer and Farris, 2007; Osborne et al., 2007; Kimsey et al., 2008) and shows promise as an exploratory tool for investigating spatial non-stationarity in other ecological settings. To date, no studies have been published that apply the GWR method to marine fisheries

data. This represents an opportunity to explore the issue of spatial non-stationarity of processes to shed light on the current state of this ecosystem.

Here, we focus on the northern Atlantic cod (*Gadus morhua*) stock as a case study for the exploratory GWR analysis. Northern cod has been managed as a single stock within the Northwest Atlantic Fisheries Organization (NAFO) Divisions 2J, 3K, and 3L (2J3KL) off the coast of Labrador and eastern Newfoundland, although it is thought to consist of several sub-populations (Lear and Parsons, 1993). After supporting one of the world's largest fisheries for nearly 400 years, the northern cod stock experienced a dramatic decline from the 1960s to the early 1990s to <1% of its maximum historical biomass (Rose, 2007). In the aftermath of this collapse, many studies have focused not only on the overfishing, but also on the underlying environmental, trophic, and demographic processes that affect the stock (Rose, 1993; Hutchings, 1996; Atkinson *et al.*, 1997; Drinkwater, 2002; Rice, 2002; Worm and Myers, 2003; Lilly, 2008). The relationships between species, fishing, and the environment have proven difficult to quantify and have resulted in uncertainty in the basic ecological knowledge of recent changes, in particular with respect to the current state of the ecosystem and prognoses for stock rebuilding.

Our overall goal was to explore the GWR method as applied to fisheries data. We compared the performance of GWR relative to the more typically used global logistic regression model and GAM in models of the distribution of northern cod based on both environmental and biological variables. Environmental variables included depth, temperature, and salinity that are typically strongly correlated, with all three structuring the distribution of cod through physiological and ecological constraints (Rose, 2005). Biological variables included snow crab (*Chionoecetes opilio*) and northern shrimp (*Pandalus borealis*), known prey of Atlantic cod in the 2J3KL region (Sherwood *et al.*, 2007). The analysis uses data collected during the autumn multispecies trawl survey of 2001. We adopted a logistic presence/absence approach to the analysis of cod distribution for several reasons: (i) cod, crab, and shrimp were not present in many areas, resulting in zero-inflated distributions for each variable; (ii) where present, cod catches were consistently very small, with little variance; and (iii) the relationships between cod weight and the dependent variables were not consistently linear. Unlike OLS regression, logistic regression does not assume a normal distribution for the variables used. Moreover, GAMs allow modelling of non-linear relationships between the dependent and independent variables through the use of smoothing functions (Hastie and Tibshirani, 1990) and are increasingly used in fisheries research for this reason (Ciannelli *et al.*, 2008). The specific goals of our study were to (i) compare the performance of logistic GWR with global logistic and binomial GAM regressions; (ii) examine spatial non-stationarity in the localized regression coefficients; (iii) map the parameter coefficients of GWR for interpretation; and (iv) explore any zones of unique species–environment relationships revealed by the GWR analysis.

Material and methods
Study area
The study area included some 520 000 km^2 of continental shelf off the east coast of Newfoundland and Labrador, encompassing the historical distribution of the northern cod stock in NAFO

Divisions 2J3KL, except the coastal region (Figure 1). Oceanographic conditions are primarily influenced by cold Arctic water of the Labrador Current, which flows south over the Labrador and northeast Newfoundland shelves and portions of the Grand Bank (Colbourne *et al.*, 1997). The bathymetry of the continental shelf is characterized by relatively shallow banks intersected by deeper channels and troughs. Mean annual bottom temperatures range between less than $-1.0°C$ over much of the Newfoundland Shelf and Grand Bank to $>3.0°C$ at the edges of the banks and are strongly influenced by local bathymetry (Colbourne *et al.*, 1997). In the region, distributions of cod are generally highly aggregated during winter and spring when the fish spawn, and more dispersed during summer and autumn, the feeding periods.

Data sources
Fisheries data for the Newfoundland region were provided by the Canadian Department of Fisheries and Oceans (DFO) and were collected in 2001 during multispecies surveys conducted in NAFO Divisions 2J3KL during the months October–December (Figure 1). Surveys followed a stratified random design and employed a Campelen 1800 Shrimp trawl. At each trawl location, data were collected on the biomass and abundance of various groundfish and shellfish species, and biological samples were collected.

Atlantic cod were classified as present (1) or absent (0) based on a threshold catch weight of 5 kg (5–10 fish) in each trawl. The 5 kg threshold was used to avoid a small catch weight having too much influence on the logistic regression and was based on a frequency distribution of catch weights per trawl. Average trawl catches of cod in the 2001 autumn multispecies surveys, including the many zero catches, were low (2.22 ± 6.19 kg s.d.). Trawl catch weights (kg) of snow crab and northern shrimp were standardized according to area swept (0.8 nautical miles in 15 min with a wing spread of 16.84 m). Mean bottom depth (m), temperature (°C), and salinity (psu) were recorded during each trawl, and the approximate distance to shore (km) from each trawl location was determined using a geographical information system (GIS). Before the regression analyses, the weights of crab and shrimp catches for each set were log-normalized to adjust for their zero-inflated skewed distributions.

Statistical analyses
A global logistic regression equation takes the form

$$\ln\left[\frac{y^*}{(1 - y^*)}\right] = \beta_0 + \sum_k \beta_k x_k, \qquad (1)$$

where y^* is the predicted value of response variable y, β_0 the intercept coefficient, and β_k the coefficient for the explanatory variable x_k ($k = 1, 2, 3, \ldots, n$). In a GAM, the coefficient β_k is replaced by a smoothing function, such that

$$\ln\left[\frac{y^*}{(1 - y^*)}\right] = \beta_0 + \sum_k f_k x_k, \qquad (2)$$

where f_k is a non-parametric function describing the effect of x_k on y^*. We fitted a binomial GAM using a logit link function, with the shape of the f_k function determined by penalized regression splines with automatic smoothness selection (Wood, 2006). All
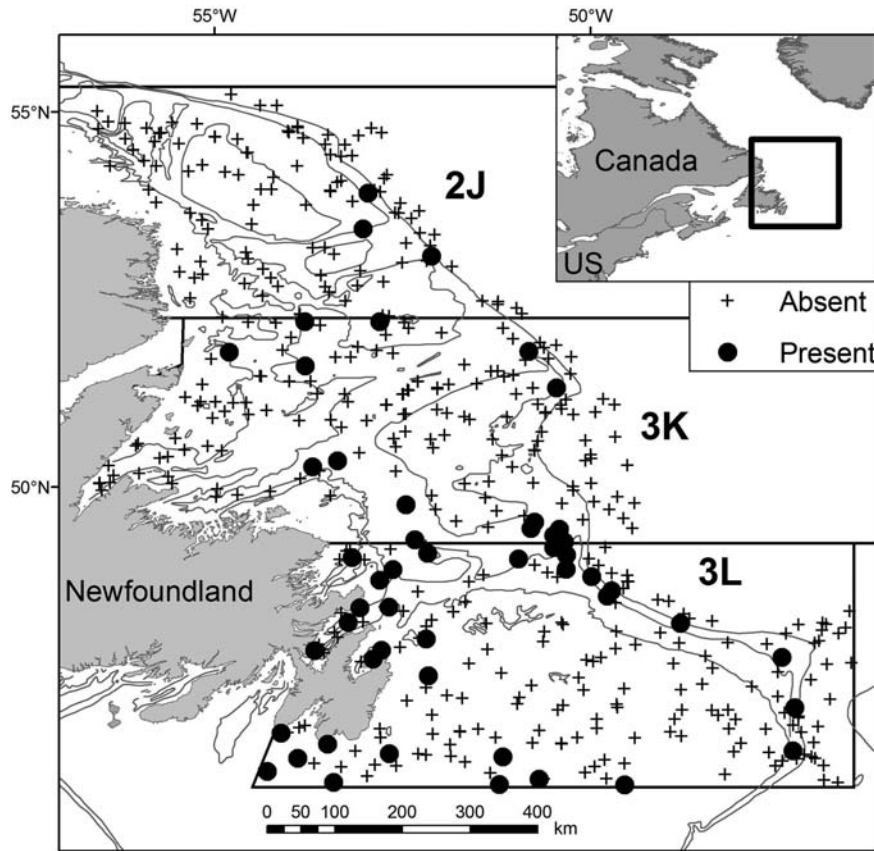
Figure 1. The continental shelf off Newfoundland and Labrador (with 200, 300, and 500 m contours identified), showing the locations of NAFO Divisions 2J3KL. The inset shows the location of the study area. The shelf-scale presence (filled circles) and absence (plus signs) distribution of Atlantic cod is shown, based on scientific survey trawl catches of cod >5 kg in autumn 2001.

explanatory variables were fitted as smoothers to allow for possible non-linear effects. GAM analyses were carried out in R (v. 2.8.1; R Development Core Team, 2008), using the "mgcv" package (v. 1.5-5; Wood, 2006).

The GWR model is essentially a modified global regression which incorporates a set of geographic coordinates for each location i, taking the form

$$\ln\left[\frac{y_i^*}{(1 - y_i^*)}\right] = \beta_{0i} + \sum_k \beta_{ki} x_{k_i}, \quad (3)$$

where β_{0i} is the intercept parameter specific to location i, and β_{ki} is the parameter coefficient of independent variable x_k at location i. Local variable coefficients in the GWR model are determined by a weighting matrix that uses a distance-decay function, resulting in local regression points being more influenced by observations closer in space. The spatial kernel controlling the distance-decay function can take either a fixed (distance) or adaptive (number of samples) approach to establishing the radius of the local GWR model, in effect creating a moving window regression for each observation point in the study area (Fotheringham *et al.*, 2002). The size of the kernel bandwidth has a large impact on the outcome of the GWR analysis and should be selected carefully. Increasingly smaller bandwidths result in parameter estimates that are highly localized and have a large degree of variance, whereas increasingly larger bandwidths tend towards the normal global

regression estimates. In our study, a fixed Gaussian kernel was used, where the weight of observation j to the ith regression point was determined by

$$w_{ij} = \exp\left[-\left(\frac{d_{ij}}{b}\right)^2\right], \quad (4)$$

where d_{ij} is the Euclidean distance between the two observation sites i and j, and b is the bandwidth in map units. The kernel bandwidth was determined by minimizing the corrected Akaike's Information Criterion (AIC$_c$) for the fitted regression model. The AIC is a relative goodness-of-fit statistic for comparing competing models, where the model with the smallest AIC provides the closest approximation to reality. For a more detailed overview of the GWR technique, the reader is referred to Fotheringham *et al.* (2002). The GWR analysis was performed with GWR 3.0 software (available at http://ncg.nuim.ie/ncg/GWR/index.htm).

The initial explanatory analysis of the environmental and biological variables was performed using a global logistic regression, binomial GAM, and logistic GWR approach using mean depth, mean bottom temperature and salinity, distance to shore, and log-normalized crab and shrimp weight. Variable and model selections were determined using the AIC$_c$ (Fotheringham *et al.*, 2002). Many regression models were calculated using different combinations of predictor variables, and AIC$_c$ statistics were compared to assess the goodness-of-fit for each model. As the environmental variables of

depth, temperature, and salinity are highly correlated, only one of each was selected for model testing to reduce multicollinearity. The GWR model with the smallest AIC_c was taken as the best representation of reality and selected for further analysis.

Using the optimal model variables determined by AIC_c minimization, global logistic regression, GAM, and GWR models were calculated for the entire NAFO 2J3KL study area. The AIC_c scores were compared to assess whether the local GWR model was a significant improvement over the global logistic model and the GAM for predicting cod occurrence. AIC_c differences >3 are assumed to represent significant differences between two models. We also calculated the area under the receiver operating characteristic curve (AUC) to compare the ability of each model to correctly predict cod presence and absence. The AUC statistic is equal to the probability that a randomly chosen positive cod occurrence will be ranked higher than a randomly chosen negative one, and ranges in value between 0 and 1, with values of 0.5 corresponding to random chance (Zou *et al.*, 2007).

To determine the spatial scale of non-stationarity in the local parameter coefficients, additional GWR analyses were carried out by increasing the fixed kernel bandwidth from a 100 to a 400 km search radius at 50 km increments. These bandwidths were chosen to reflect a range of values around the AIC_c optimized bandwidth. The variation in the regression coefficient for each variable at each location was subsequently used as an approximate index of stationarity (Brunsdon *et al.*, 1998). This was accomplished by dividing the interquartile range of the GWR coefficient by twice the s.e. of the same variable from the global logistic regression. As ∼68% of the values are within 1 s.d. and 50% are within the interquartile range, ratio values >1 suggest that the relationship between the independent and dependent variables might be spatially non-stationary. These index values were plotted against the kernel bandwidth distances to examine at what spatial scale the relationships became non-stationary.

Given that the errors for a well-fitted model are randomly distributed across a study region, the spatial fit of the global logistic, GAM, and GWR models was examined using Moran's *I* analysis of the model residuals. Moran's *I* coefficients between 0 and 1 indicate positive spatial autocorrelation, negative coefficients between 0 and −1 indicate dissimilar neighbouring values, and those near 0 indicate weak or no spatial autocorrelation (Fortin and Dale, 2005). Correlograms of Moran's *I* coefficients in 20-km distance classes were generated to examine scales of spatial autocorrelation in the residuals.

An advantage of GWR is its ability to map the local parameter coefficients produced for each datapoint, as well as model diagnostics, to visualize and interpret spatial non-stationarity in the dataset. This can be challenging given the large amount of spatial data produced by GWR, as well as the need to present simultaneously both the parameter estimates and their associated significance (e.g. *t*-values) for accurate interpretation (Mennis, 2006). Here, we mapped the spatial distributions of local r^2 and GWR parameter coefficients, applying a significant threshold of 95% to mask out points where the relationship between cod and the predictor variable was not significant.

In addition to the spatial analysis of single-parameter coefficients, it is also useful to examine whether there are any spatial groupings of parameters with relatively homogenous values. For example, these groupings might identify zones of distinct species–environment relationships (Wimberly *et al.*, 2008). A *k*-means cluster analysis was used to separate the parameters

into different zones, using the *t*-values of the local GWR parameter coefficients and with the number of clusters (*k*) set *a priori* to *k* = 2, 3, and 4. Average parameter coefficients for each cluster were calculated, and the spatial distribution of the clusters was mapped using a GIS. All maps were generated using ArcGIS (ESRI, v. 9.3) software.

Results

The mapping of catches from the 2001 autumn survey in 2J3KL revealed that cod were mostly inshore near the Avalon Peninsula and offshore near the border between 3K and 3L (Figure 1). There was a general absence of cod in shallower water as well as in the northern part of 2J. These distributions correspond to the known locations of remaining populations of cod within the stock's range (DFO, 2002; Rose, 2003).

Based on AIC_c minimization scores, the model that best approximated reality for the GWR method included the variables of temperature, distance to shore, and crab and shrimp weight (AIC_c = 271.4). The AIC_c values for the equivalent global logistic regression (323.4) and binomial GAM (313.8) were significantly higher. AIC_c scores for GWR were lower than those for the global logistic regression by a range of 16–52, indicating that GWR resulted in a significantly better fit for all tested combinations of variables. For subsequent analysis, we elected to continue with the variables temperature, distance to shore, and crab and shrimp weight to generate logistic, GAM, and GWR models for the entire 2J3KL region. The results of the global logistic model suggested that, throughout the study area, crab had a significant negative association and distance from shore a weak negative association with cod, whereas shrimp had a stronger positive association and temperature a weaker positive relationship (Table 1). Descriptive statistics for the local parameter coefficients produced by GWR revealed much variance in the parameter values (Table 2), suggesting the presence of spatial non-stationarity in the relationships between cod distribution and the explanatory variables. The temperature, distance-from-shore, and crab variables had both negative and positive parameter values, although temperature was mostly positive and the distance and crab variables were mostly negative. In contrast, all parameter values for the shrimp variable were 100% positive (Table 2). The GWR model was an improvement over the global logistic and GAM regressions for predicting cod presence/absence, as indicated by the significantly lower AIC_c score and higher AUC value (Table 3). The proportion of variance explained by the local GWR models for each location *i* was higher (r^2 = 0.11–0.26) than the global measures for both the global logistic (r^2 = 0.013) or GAM (0.072) models (Table 3). Spatial mapping of the local pseudo-r^2 values suggested that the GWR models had greater explanatory power nearshore, but performed worse in the 2J area (Figure 2).

Table 1. Parameter estimates for the global logistic regression model.

Variable	B	s.e.	t-value	Exp(B)
Intercept	−1.769	0.346	−5.108	0.171
Temperature	0.058	0.120	0.486	1.060
Distance	−0.000001	−0.000001	−1.827	1.000
Crab	−0.407	0.187	−2.183	0.665
Shrimp	0.145	0.078	1.869	1.156

Table 2. Summary statistics of the logistic GWR parameter estimates[a], including the overall percentage of negative (% −) and positive (% +) values.

Variable	Minimum	Lower quartile	Median	Upper quartile	Maximum	% −	% +
Intercept	−8.63	−3.99	−2.00	−1.04	0.22	97.3	2.7
Temperature	−0.05	0.28	0.32	0.40	0.58	1.5	98.5
Distance	−0.000012	−0.000009	−0.000007	−0.000006	0.000016	84.8	15.2
Crab	−1.49	−0.68	−0.46	−0.02	0.89	76.9	23.1
Shrimp	0.16	0.22	0.30	0.41	0.48	0.0	100.0

[a]AIC$_c$ optimized bandwidth of 212 km.

Table 3. Comparison of fit for global logistic, GAM, and logistic GWR[a] models applied to the distribution of Atlantic cod off the coast of Newfoundland in 2001, with the range of pseudo-r^2 values for the GWR model provided.

Model	n	k_e	−2 log likelihood	AIC$_c$	AUC ± s.e.	r^2 (adj.)
Global logistic	481	5.0	313.3	323.4	0.65 ± 0.038	0.013
GAM	481	5.0	290.2	313.7	0.73 ± 0.038	0.072
GWR	481	12.4	245.7	271.4	0.87 ± 0.023	0.11−0.26

n, number of observations; k_e, effective number of parameters; AIC$_c$, corrected Akaike's Information Criterion; AUC, area under the receiver operating characteristic curve.
[a]212 km bandwidth.

The AIC$_c$ optimized bandwidth for the GWR model was 212 km. By varying the bandwidth of the GWR model, we noted that predictor variables became spatially stationary at different scales (Figure 3). The results suggested that the temperature component of the model became stationary beyond 150–200 km, whereas shrimp was non-stationary up to ∼250 km. The crab component of the model remained non-stationary up to a bandwidth of 300 km (Figure 3).

The global Moran's I analysis of model residuals confirmed the improved performance of GWR over the global methods. Weak non-significant spatial autocorrelation was detected in the residuals from the logistic regression (Moran's $I = 0.11$; $p = 0.24$) and GAM (Moran's $I = 0.067$; $p = 0.48$), and the GWR model residuals had a completely random spatial pattern (Moran's $I = 0.002$; $p = 0.97$). Global Moran's I coefficients were plotted against lag distance to examine any spatial patterns in the residuals (Figure 4). The global logistic and GAM residuals showed significant positive spatial autocorrelation up to ∼80 km, but there was no significant positive spatial autocorrelation for the GWR residuals, except for a lag distance of 50 km (Figure 4). Global logistic, GAM, and GWR residuals resulted in similar spatial patterns, although the global residuals were larger in both negative and positive directions.

Model performance and non-stationarity were also explored visually by mapping the local coefficient estimates of the temperature, distance-from-shore, snow crab, and shrimp variables produced by GWR. Parameter estimates varied spatially according to both the locations where they had a significant association with cod presence and the direction (negative or positive) of this relationship (Figure 5). The relationship between cod and temperature was mostly non-significant throughout the study area, except a moderate positive relationship at the eastern edge of the Grand Bank (3L) and an inshore area in 3K (Figure 5a). Distance from shore had a weak negative association with cod presence/absence in most of 3K and all of 3L, with a pronounced

inshore/offshore gradient indicating that more cod were likely to be inshore in these areas (Figure 5b). A stronger negative crab–cod relationship was found in the south and in most nearshore areas, with the relationship becoming weaker with increasing distance from shore (Figure 5c). Crab did not have a significant association with cod in 2J or northern 3K, as well as at the eastern tip of the Grand Bank. The relationship with shrimp was significant and positive throughout the entire study area, with the highest parameter estimates in the north and along the edges of the continental shelf (Figure 5d).

The k-means cluster analysis revealed distinct latitudinal zones of species–environment effects on cod distribution (Figure 6). A northern zone was identified (Cluster 2) and was characterized by stronger positive relationships with crab and shrimp, and a weak positive relationship with distance from shore (Table 4; Figure 6). That zone remained spatially consistent through all cluster analyses. A distinct latitudinal band (Cluster 2) also emerged extending from White Bay in 3K along the southern edge of Cluster 1 for both three- and four-cluster analyses (Figure 6). A third zone (Cluster 3) was characterized by a relatively strong negative relationship with crab. A final zone emerged at the eastern tip of the Grand Bank (Cluster 4) that was differentiated from Cluster 3 by a stronger positive relationship with temperature and a weaker negative relationship with crab (Table 4). Average shrimp parameter estimates were most useful for distinguishing the northern zone (Cluster 2) and had similar values in all other zones.

Discussion

We found that relationships between Atlantic cod and both abiotic and biotic variables that influence distribution could be investigated over multiple spatial scales by incorporating geographic information into GWR analyses. The improved performance of GWR over the global logistic and GAM regression models was demonstrated by significantly lower values of AIC$_c$ and deviance as well as an improved non-spatial AUC score. An additional strength of the GWR method was the ability to explore spatial variability in the relationships between cod and explanatory variables by mapping the variation in local parameter coefficients. GWR revealed spatial non-stationarity in the relationships between cod and temperature, distance from shore, crab, and shrimp weight, and that the significance and direction of these relationships varied throughout the entire 2J3KL study area. Overall, the results highlight the limitations of using a single set of global parameters to model the distribution of a species in a large marine ecosystem.

Non-stationarity of ecological processes is not a new concept in fisheries science (Walters, 1987), although it has almost exclusively been associated with temporal (rather than spatial) variation in species–environment relationships. By shifting the GWR
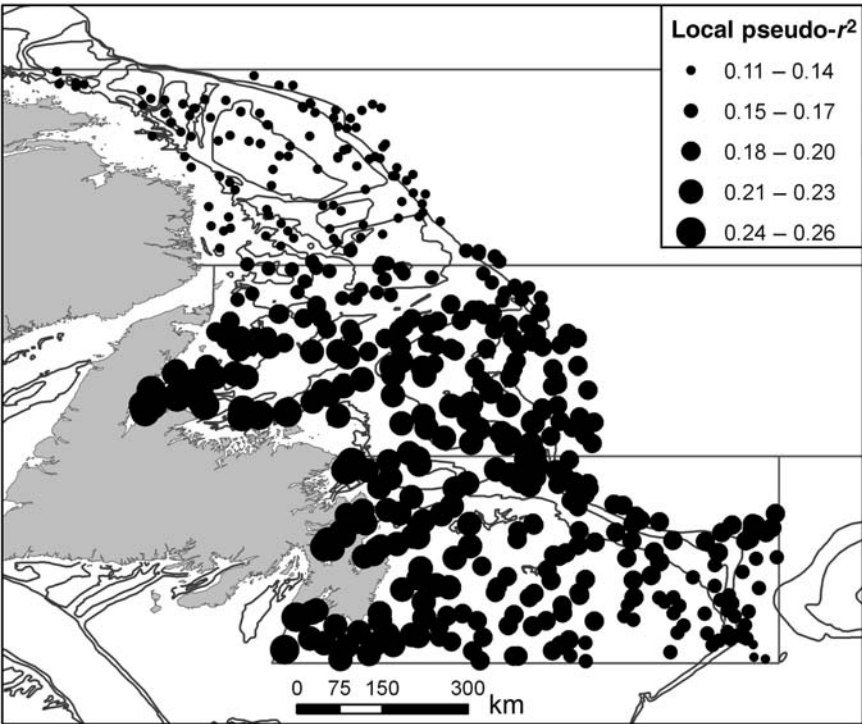
Figure 2. Mapped local pseudo-r^2 values from the GWR models of cod presence/absence in the NAFO 2J3KL region for autumn 2001.
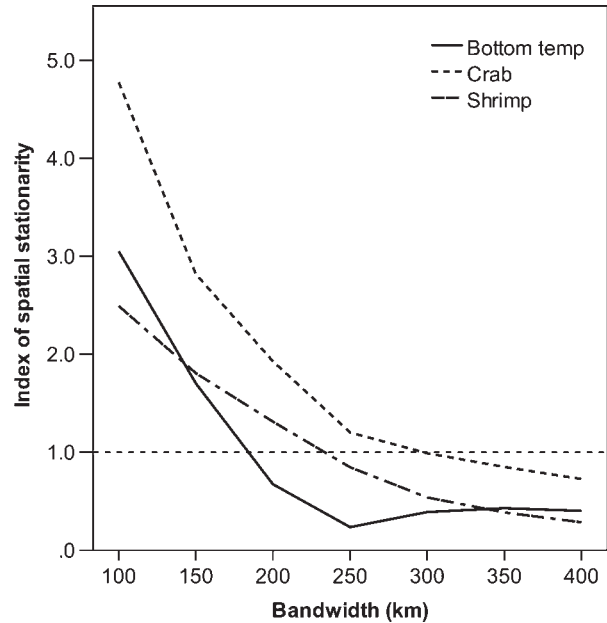

Figure 3. Index of spatial stationarity for variables used in the GWR models for the occurrence of Atlantic cod in NAFO Divisions 2J3KL. The index is calculated by dividing the interquartile range of a GWR regression coefficient by twice the s.e. of the same parameter estimate from the global model. Spatial non-stationarity is suggested by index values >1.

bandwidth, we were able to detect variability in the spatial stationarity of the relationship between cod, temperature, and crustacean prey. This was perhaps expected given that spatial predator–prey associations can vary with scale (Rose and

Leggett, 1990; Ciannelli *et al.*, 2008) and geographic location, depending on alternative prey distributions (Sherwood *et al.*, 2007). The relationship between cod and temperature was the most spatially stable, whereas the relationship with crab proved to be spatially inconsistent up to large bandwidths (300 km). These results suggest that the observed spatial patterns of cod in the 2J3KL region are determined by ecological factors acting concurrently at different spatial scales.

Although the use of a single year's data to demonstrate the GWR methodology on fisheries survey data does not merit much ecological interpretation, a few comments may be in order to help assess the present results and their potential. It is useful to first examine the spatial context of the variables used in this analysis. The seasonal distribution of the northern cod stock has been described in many studies in relation to the geography of the continental shelf (Atkinson *et al.*, 1997), oceanographic variables (Rose, 1993; Rose *et al.*, 1994), and prey species (Rose, 1993; Lilly *et al.*, 2000; Rose and O'Driscoll, 2002), all of which provide context for the main finding of non-stationarity in the relationship between cod distribution and environmental factors. For example, shrimp are primarily distributed offshore north and east of the Grand Bank (DFO, 2008), whereas snow crab are found over a broad range of depths and more inshore, particularly in 3K and 3L (Dawe and Colbourne, 2002). Shrimp and crab indices were the strongest predictors of cod distribution, as found by both global and GWR methods. In 2001, the year currently analysed, expanding populations of shrimp had largely replaced capelin as the primary prey item for cod in offshore regions of 2J3KL, especially in the north (Rose and O'Driscoll, 2002; Sherwood *et al.*, 2007), and this likely explains the stronger positive relationship between cod and shrimp in the 2J region. The largely negative GWR coefficients between cod and snow crab throughout the study area are not so easily explained, but could
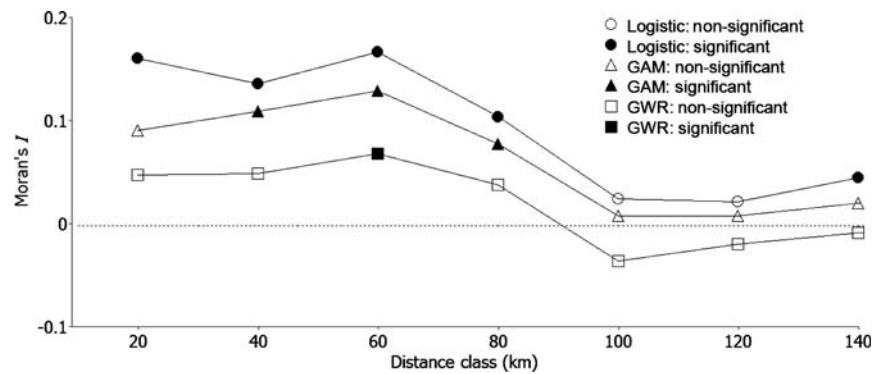
Figure 4. Correlograms of global Moran's *I* values for residuals of the global logistic (circles), binomial GAM (triangles), and logistic GWR (squares) models of cod occurrence in NAFO Divisions 2J3KL for the year 2001. Filled symbols indicate significant spatial autocorrelation ($p <$ 0.05).
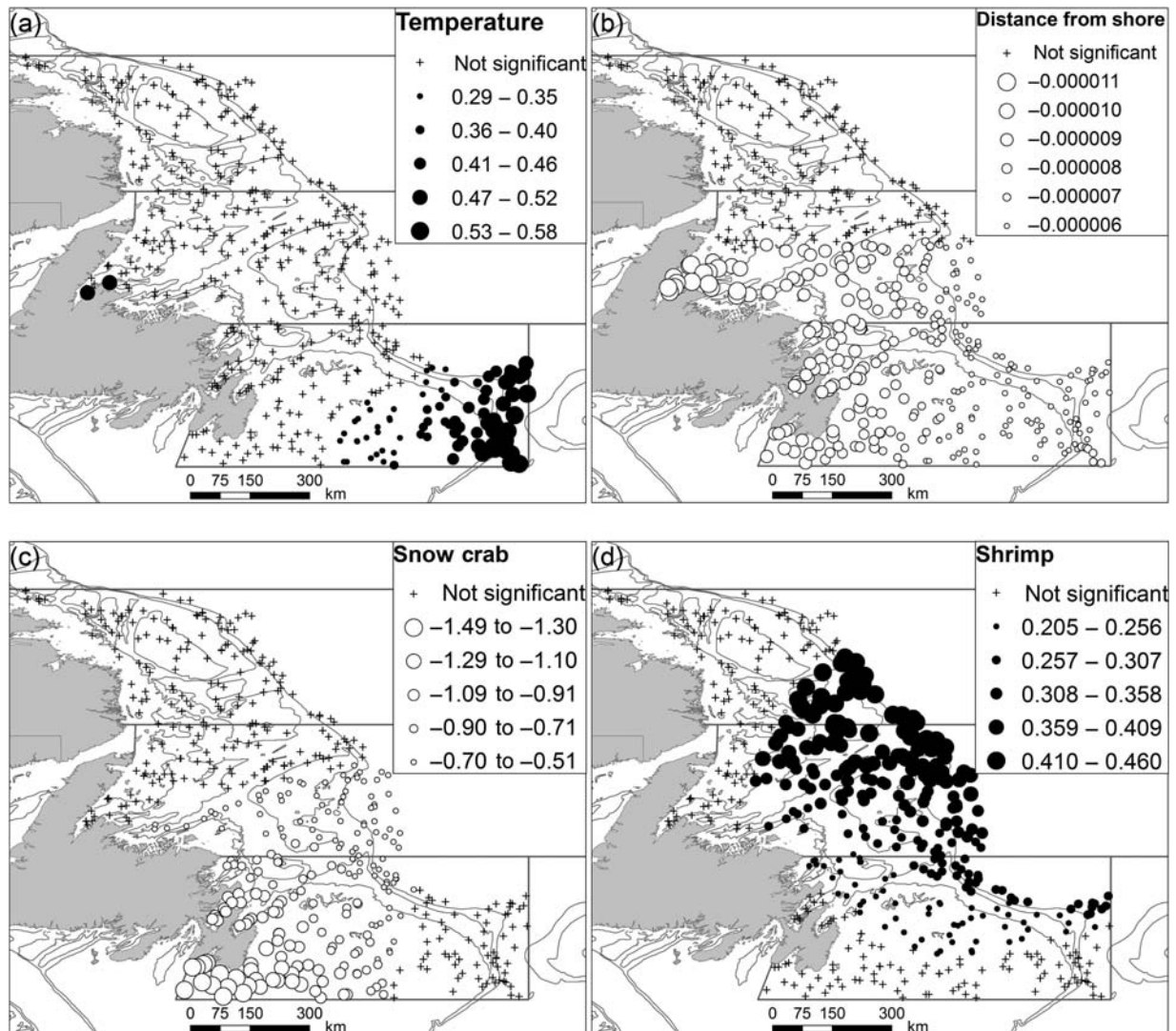

Figure 5. GWR-derived local coefficient estimates for (a) temperature, (b) distance from shore, and (c) snow crab and (d) shrimp as predictors of cod presence/absence in the 2J3KL region for autumn 2001. Positive values are shown as filled circles and negative values as unfilled circles. A significant threshold of 95% was used to mask out points where the relationship between cod and the predictor variable was not significant (plus signs).
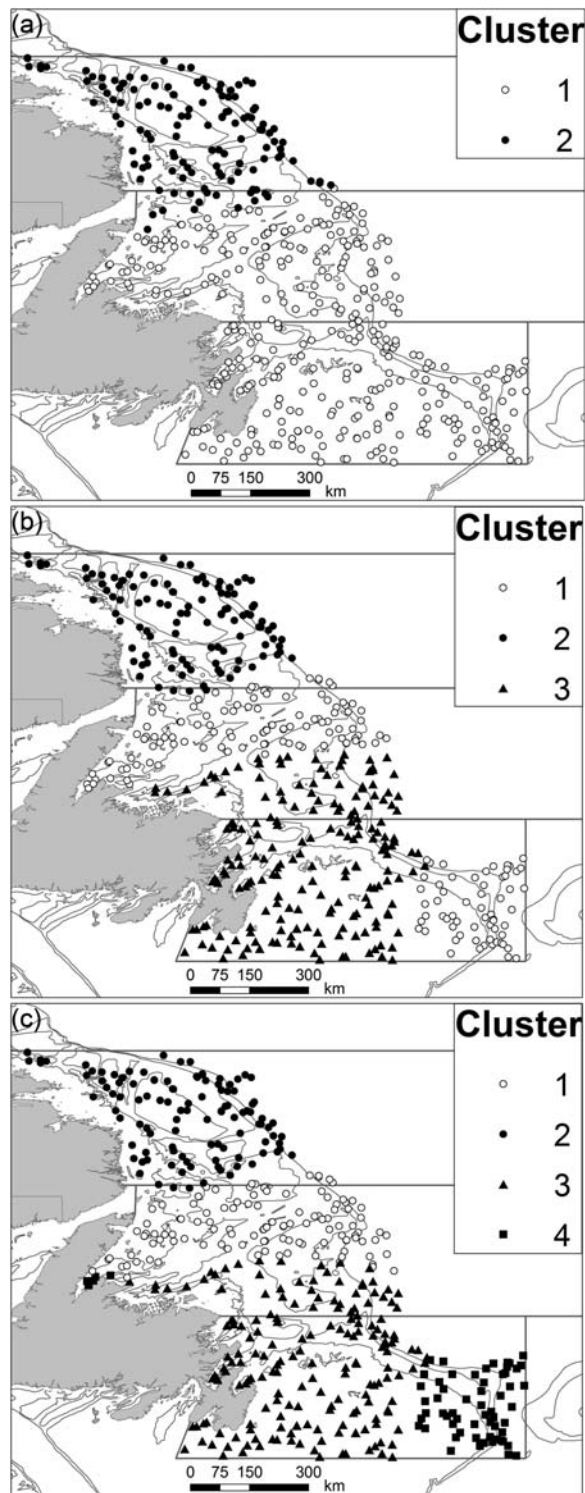
Figure 6. Mapped results of *k*-means cluster analyses of the pseudo *t*-values from the logistic GWR models, for three clusters, (a) $k = 2$, (b) $k = 3$, and (c) $k = 4$.

relate to autumn migration patterns of cod or to gear selection for large male crab differing from cod selection for smaller females or could be spurious based on a single year's data. The mostly non-significant relationship with temperature may have been a consequence of the dispersal of cod during autumn (Rose, 1993) and

low population size and spatial contraction both offshore (Atkinson *et al.*, 1997) and inshore (Rose, 2003), eliminating region-wide trends. That the GWR method can illustrate both plausible results and raise questions that potentially can be answered is reason to pursue such spatial analyses of species–environmental interactions.

Our cluster analysis of parameter coefficients characterized zones of unique spatial associations between covariates at the time of the survey. Wimberly *et al.* (2008) achieved similar results in a large-scale terrestrial study of tick-borne pathogens and suggested that a unique model could be applied to each zone to improve predictions at unsampled locations. For example, in the current study, 2J and 3L appear to represent unique zones at two- and three-cluster analyses, with a transition in 3K. These zones would undoubtedly be refined with the inclusion of additional explanatory variables, yet still suggest that both global and local models would benefit from spatial partitioning of the data.

This study achieved similar results to previous GWR models applied to forestry data (Zhang and Shi, 2004; Kupfer and Farris, 2007; Kimsey *et al.*, 2008) and terrestrial species (Foody, 2004; Shi *et al.*, 2006; Osborne *et al.*, 2007; Wimberly *et al.*, 2008), in which GWR was shown to be a significant improvement over global regression techniques. The lack of spatial patterns in the GWR model residuals suggests that this technique successfully corrected for spatial autocorrelation in the data, which is typically present given the patchy distribution of marine species and environmental gradients (Pennington, 1996; Ciannelli *et al.*, 2008). Both the global and local models performed poorly in distinguishing areas where cod were absent. This may again be related to the low population status and contracted distribution of the northern cod stock in the data used, or perhaps missing variables from the analysis. Mapping of the pseudo-r^2 values from the local models also suggested regional variability in the proportion of variance explained by GWR. Fotheringham *et al.* (2002) note that the local r^2 statistic needs to be interpreted with caution, as the performance of each local model near a regression point *i* may change if the processes under study exhibit some degree of spatial non-stationarity.

Fisheries models are faced with the inherent difficulty of predicting the distribution of mobile fisheries resources in a dynamic marine environment, with processes that operate on different spatial and temporal scales (Ricklefs, 1990). These challenges are confounded further in cases when stocks have undergone rapid changes in abundance and spatial distribution. The results here appear to be sufficiently informative to warrant more comprehensive application and enhancement of these techniques in fisheries. Nevertheless, improvements are no doubt feasible. For example, although the predictive power of the present GWR model based on four variables (temperature, distance to shore, shrimp, and crab indices) was a significant improvement over the global logistic regression and GAM, it still failed to account for cod distributions in some areas, suggesting that other influential factors were not accounted for. For example, data on capelin (*Mallotus villosus*), a key prey of cod in the region (Rose and O'Driscoll, 2002), were not available. Moreover, oceanographic variables such as temperature may not always consistently predict cod distribution when other ecological factors take precedence, as in the case when cod leave preferred thermal paths to pursue prey (Rose, 1993). Therefore, perhaps there may be non-additive effects of environmental and trophic

Table 4. Mean GWR parameter estimates from the model of 2J3KL cod distribution in autumn 2001 (212 km bandwidth) for each group identified by the *k*-means cluster analyses (s.d. in parentheses).

k	*c*	*n*	Temperature	Distance from shore	Snow crab	Shrimp
2	1	351	0.338 (0.086)	−0.000008 (0.000001)	−0.601 (0.329)	0.270 (0.069)
	2	130	0.321 (0.122)	0.000002 (0.000006)	0.247 (0.308)	0.436 (0.024)
3	1	172	0.408 (0.071)	−0.000008 (0.000001)	−0.316 (0.189)	0.317 (0.074)
	2	113	0.308 (0.126)	0.000003 (0.000005)	0.317 (0.265)	0.441 (0.020)
	3	196	0.283 (0.040)	−0.000008 (0.000001)	−0.818 (0.238)	0.241 (0.051)
4	1	117	0.369 (0.068)	−0.000008 (0.000001)	−0.418 (0.126)	0.367 (0.035)
	2	112	0.308 (0.126)	0.000003 (0.000005)	0.321 (0.263)	0.441 (0.020)
	3	186	0.285 (0.042)	−0.000008 (0.000001)	−0.829 (0.239)	0.234 (0.045)
	4	66	0.452 (0.067)	−0.000007 (0.000001)	−0.176 (0.201)	0.234 (0.026)

variables in the GWR model, a common issue when modelling fisheries data (Ciannelli *et al.*, 2008). Additional improvements to predictions might come from recognizing the strong association between predator size and prey selectivity, with medium-sized cod (30–70 cm) tending to feed on pelagic prey and larger cod (>70 cm) selecting a greater proportion of invertebrates in their diet (Sherwood *et al.*, 2007). One of the strengths of local spatial methods such as GWR is that an analysis of the spatial variation in model performance and regression coefficients may help improve both model specification and enhance understanding of the spatial processes under study.

Although having many advantages, GWR should be used with some caveats. Owing to the local nature of parameter estimates, the GWR model cannot be used to predict species distributions outside the study area. The overall distribution patterns of the parameter estimates may also be indicative of model misspecification (Fotheringham *et al.*, 2002). Attention should be given to possible collinearity in the local regression coefficients, which may limit interpretation of their distributional patterns (Wheeler and Tiefelsdorf, 2005), as well as any curvilinear relationships, which may produce false results of non-stationarity (Austin, 2007). Given these limitations, GWR has been recommended as a supplementary exploratory tool with global regression methods to investigate how relationships between variables can vary across a landscape (Fotheringham *et al.*, 2002).

In conclusion, although useful in terms of testing general theories of ecological relationships, the common practice of averaging population and environmental data over large geographic regions to build global predictive models has the disadvantage of potentially masking local variability in the processes being studied (Fortin and Dale, 2005). The results of the current study indicate this to be true in marine as well as in the more commonly studied terrestrial ecosystems. This spatial non-stationarity can hamper efforts to achieve meaningful interpretation of ecological studies. Non-parametric statistical methods are increasingly being used in studies of fisheries ecology, in recognition of the fact that ecological data often do not meet the assumptions of normal distribution and linear relationship (Haddon, 2001; Ciannelli *et al.*, 2008; Espeland *et al.*, 2008), and GWR represents a relatively new tool to explore the parametric assumption of spatial stationarity. Our relatively simple analysis, the first GWR application to fisheries of which we are aware, was able to reveal significant local variation in the species–environment relationships under study and illustrates the potential for the GWR method in analyses of fisheries data, particularly for understanding and predicting the spatial dynamics of large marine ecosystems. To that end, an expanded study is planned to explore the full time-series of

fisheries data for the Newfoundland region using GWR and additional abiotic and biotic variables.

References

Atkinson, D. B., Rose, G. A., Murphy, E. F., and Bishop, C. A. 1997. Distribution changes and abundance of northern cod (*Gadus morhua*), 1981–1993. Canadian Journal of Fisheries and Aquatic Sciences, 54: 132–138.

Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecological Modelling, 200: 1–19.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. E. 1998. Geographically weighted regression—modelling spatial non-stationarity. Statistician, 47: 431–443.

Ciannelli, L., Fauchald, P., Chan, K. S., Agostini, V. N., and Dingsor, G. E. 2008. Spatial fisheries ecology: recent progress and future prospects. Journal of Marine Systems, 71: 223–236.

Colbourne, E. B., deYoung, B., Narayanan, S., and Helbig, J. 1997. Comparison of hydrography and circulation of the Newfoundland Shelf during 1990–1993 with the long-term mean. Canadian Journal of Fisheries and Aquatic Sciences, 54(Suppl. 1): 68–80.

Dawe, E. G., and Colbourne, E. B. 2002. Distribution and demography of snow crab (*Chionoecetes opilio*) males on the Newfoundland and Labrador Shelf. *In* Crabs in Cold Water Regions: Biology, Management, and Economics. Proceedings of the 19th Lowell Wakefield Symposium, pp. 577–594. Ed. by A. J. Paul, E. G. Dawe, R. Elner, G. S. Jamieson, G. H. Kruse, R. S. Otto, B. Sainte-Marie, *et al.* University of Alaska Sea Grant AK-SG-02-01, Fairbanks. 866 pp.

DFO. 2002. Northern (2J+3KL) cod stock status update. DFO Science Stock Status Report A2-01(2002).

DFO. 2008. Assessment of Divisions 2G-3K northern shrimp. Canadian Stock Assessment Secretariat Science Advisory Report 2008/008.

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography, 30: 609–628.

Drinkwater, K. F. 2002. A review of the role of climate variability in the decline of northern cod. American Fisheries Society Symposium, 32: 113–130.

Espeland, S. H., Olsen, E. M., Knutsen, H., Gjøsæter, J., Danielssen, D., and Stenseth, N. C. 2008. New perspectives on fish movement: kernel and GAM smoothers applied to a century of tagging data on coastal Atlantic cod. Marine Ecology Progress Series, 372: 231–241.

Foody, G. M. 2004. Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. Global Ecology and Biogeography, 13: 315–320.

Fortin, M-J., and Dale, M. R. T. 2005. Spatial Analysis: a Guide for Ecologists. Cambridge University Press, Cambridge, UK. 382 pp.

Fortin, M-J., and Melles, S. J. 2009. Avian spatial responses to forest spatial heterogeneity at the landscape level: conceptual and statistical challenges. In Real World Ecology: Large-Scale and Long-Term Case Studies and Methods, pp. 137–160. Ed. by S. Miao, S. Carstenn, and M. Nungesser. Springer, New York. 312 pp.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. E. 2002. Geographically Weighted Regression: the Analysis of Spatially Varying Relationships. Wiley, Chichester, UK. 269 pp.

Haddon, M. 2001. Modelling and Quantitative Methods in Fisheries. Chapman and Hall, Boca Raton, FL. 480 pp.

Hastie, T. J., and Tibshirani, R. J. 1990. Generalized Additive Models. Chapman and Hall, London. 352 pp.

Hutchings, J. A. 1996. Spatial and temporal variation in the density of northern cod and a review of hypotheses for the stock's collapse. Canadian Journal of Fisheries and Aquatic Sciences, 53: 943–962.

Kimsey, M. J., Moore, J., and McDaniel, P. 2008. A geographically weighted regression analysis of Douglas-fir site index in north central Idaho. Forest Science, 54: 356–366.

Kupfer, J. A., and Farris, C. A. 2007. Incorporating spatial non-stationarity of regression coefficients into predictive vegetation models. Landscape Ecology, 22: 837–852.

Lear, W. H., and Parsons, L. S. 1993. History and management of the fishery for northern cod in NAFO Divisions 2J, 3K and 3L. In Perspectives on Canadian Marine Fisheries Management, pp. 55–89. Ed. by L. S. Parsons, and W. H. Lear. Canadian Bulletin of Fisheries and Aquatic Sciences, 226.

Lilly, G. R. 2008. The decline, recovery, and collapse of Atlantic cod (Gadus morhua) off Labrador and eastern Newfoundland. In Resiliency of Gadid Stocks to Fishing and Climate Change, pp. 67–88. Ed. by G. H. Kruse, K. Drinkwater, J. N. Ianelli, J. S. Link, D. L. Stram, V. Wespestad, and D. Woodby. Alaska Sea Grant College Program, Alaska. 364 pp.

Lilly, G. R., Parsons, G., and Kulka, D. W. 2000. Was the increase in shrimp biomass on the northeast Newfoundland shelf a consequence of a release in predation from cod? Journal of Northwest Atlantic Fishery Science, 27: 45–61.

Mennis, J. 2006. Mapping the results of geographically weighted regression. Cartographic Journal, 43: 171–179.

Osborne, P. E., Foody, G. M., and Suárez-Seoane, S. 2007. Non-stationarity and local approaches to modelling the distributions of wildlife. Diversity and Distributions, 13: 313–323.

Ottersen, G., Alheit, J., Drinkwater, K., Friedland, K., Hagen, E., and Stenseth, N. C. 2004. The responses of fish populations to ocean climate fluctuations. In Marine Ecosystems and Climate Variation, pp. 73–94. Ed. by N. Ch. Stenseth, G. Ottersen, J. W. Hurrell, and A. Belgrano. Oxford University Press, Oxford, UK. 252 pp.

Pennington, M. 1996. Estimating the mean and variance from highly skewed marine data. Fishery Bulletin US, 94: 498–505.

R Development Core Team. 2008. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Rice, J. C. 2002. Changes to the large marine ecosystem of the Newfoundland and Labrador shelf. In Large Marine Ecosystems of the North Atlantic, pp. 51–104. Ed. by K. Sherman, and H. R. Skoldal. Elsevier, Amsterdam. 464 pp.

Ricklefs, R. E. 1990. Scaling patterns and processes in marine ecosystems. In Large Marine Ecosystems: Patterns, Processes and Yields, pp. 169–178. Ed. by K. Sherman, L. M. Alexander, and D. B. Gold. American Association for the Advancement of Science, Washington, DC. 249 pp.

Rose, G. A. 1993. Cod spawning on a migration highway in the North-west Atlantic. Nature, 366: 458–461.

Rose, G. A. 2003. Monitoring coastal northern cod: towards an optimal survey of Smith Sound, Newfoundland. ICES Journal of Marine Science, 60: 453–462.

Rose, G. A. 2005. On distributional responses of North Atlantic fish to climate change. ICES Journal of Marine Science, 62: 1360–1374.

Rose, G. A. 2007. Cod: the Ecological History of the North Atlantic Fisheries. Breakwater Books Ltd, St John's, NF, Canada. 591 pp.

Rose, G. A., Atkinson, D. B., Baird, J., Bishop, C. A., and Kulka, D. W. 1994. Changes in distribution of Atlantic cod and thermal variations in Newfoundland waters, 1980–1992. ICES Marine Science Symposia, 198: 542–552.

Rose, G. A., and Leggett, W. C. 1990. The importance of scale to predator–prey spatial correlations: an example of Atlantic fishes. Ecology, 71: 33–43.

Rose, G. A., and O'Driscoll, R. L. 2002. Capelin are good for cod: can the northern stock rebuild without them? ICES Journal of Marine Science, 59: 1018–1026.

Sherwood, G. D., Rideout, R. M., Fudge, S. B., and Rose, G. A. 2007. Influence of diet on growth, condition and reproductive capacity in Newfoundland and Labrador cod (Gadus morhua): insights from stable carbon isotopes (δ^{13}C). Deep Sea Research II, 54: 2794–2809.

Shi, H., Laurent, E. J., LeBouton, J., Racevskis, L., Hall, K. R., Donovan, M., Doepker, R. V., et al. 2006. Local spatial modeling of white-tailed deer distribution. Ecological Modelling, 190: 171–189.

Walters, C. J. 1987. Nonstationarity of production relationships in exploited populations. Canadian Journal of Fisheries and Aquatic Sciences, 44: 156–165.

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

Wimberly, M. C., Yabsley, M. J., Baer, A. D., Dugan, V. G., and Davidson, W. R. 2008. Spatial heterogeneity of climate and land-cover constraints on distributions of tick-borne pathogens. Global Ecology and Biogeography, 17: 189–202.

Wood, S. N. 2006. Generalized Additive Models: an Introduction with R. Chapman and Hall, CRC Press, Boca Raton, FL. 416 pp.

Worm, B., and Myers, R. A. 2003. Meta-analysis of cod–shrimp interactions reveals top-down control in oceanic food webs. Ecology, 84: 162–173.

Zhang, L., and Gove, J. H. 2005. Spatial assessment of model errors from four regression techniques. Forest Science, 51: 334–346.

Zhang, L., and Shi, H. 2004. Local modeling of tree growth by geographically weighted regression. Forest Science, 50: 225–244.

Zou, K. H., O'Malley, A. J., and Mauri, L. 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation, 115: 654–657.