A spatial kernel density method to estimate diet composition of fish

Journal:	<i>Canadian Journal of Fisheries and Aquatic Sciences</i>
Manuscript ID	cjfas-2017-0306.R2
Manuscript Type:	Article
Date Submitted by the Author:	12-Apr-2018
Complete List of Authors:	Binion-Rock, Samantha; North Carolina State University, Department of Applied Ecology Reich, Brian; North Carolina State University, Department of Statistics Buckel, Jeffrey; North Carolina State University, Department of Applied Ecology
Is the invited manuscript for consideration in a Special Issue? :	N/A
Keyword:	ESTUARIES < Environment/Habitat, TROPHIC RELATIONSHIPS < General, SPATIAL ANALYSIS < General, DIET < General, MARINE FISHERIES < General
<p>Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.</p> <p>Creating S Predict Grids.R Cross Validation using Grid Search.R Cross Validation using Optimization.R Data Simulation.R Estimating Diet using NonSpatial Estimator.R Estimating Diet using Spatial Estimator (SDKE).R Monte Carlo Simulation Study and Sample Size Comparison.R Bhattacharyya Distance.R</p>	

SCHOLARONE™
Manuscripts

A spatial kernel density method to estimate diet composition of fish

Samantha M. Binion-Rock^{1*}, Brian J. Reich², and Jeffrey A. Buckel¹

¹Department of Applied Ecology, North Carolina State University, Center for Marine Sciences and Technology, 303 College Circle, Morehead City, North Carolina, 28557.

²Department of Statistics, North Carolina State University, 2311 Stinson Drive, Campus Box 8203, Raleigh, North Carolina, 27695.

*corresponding author: smbinion@ncsu.edu

Draft

Abstract

We present a novel spatially-explicit kernel density approach to estimate the proportional contribution of a prey to a predator's diet by weight. First, we compare the spatial estimator to a traditional cluster-based approach using a Monte Carlo simulation study. Next we compare the diet composition of three predators from Pamlico Sound, North Carolina to evaluate how ignoring spatial correlation affected diet estimates. The spatial estimator had lower MSE values compared to the traditional cluster-based estimator for all Monte Carlo simulations. Incorporating spatial correlation when estimating the predator's diet resulted in a consistent increase in precision across multiple levels of spatial correlation. Bias was often similar between the two estimators but when it differed it mostly favored the spatial estimator. The two estimators produced different estimates of proportional contribution of prey to the diets of the three field-collected predator species, especially when spatial correlation was strong and prey were consumed in patchy areas. Our simulation and empirical data provide strong evidence food habits data should be modeled using spatial approaches and not treated as spatially-independent.

16 **Key Words**

17 Spatial correlation, kernel density, food habits, fish, Pamlico Sound, North Carolina

Draft

Introduction

Multispecies approaches are increasingly being used to gain a better understanding of ecosystem structure and population dynamics (Latour et al. 2003; Hollowed et al. 2000; Link 2010; Fogarty et al. 2012). Many of these approaches (e.g. MSVPA, Ecopath) require information on the diet composition of fish. Until recently (Ainsworth et al. 2010; Masi et al. 2014), there has been less attention on the quality of diet composition data relevant to multispecies model characteristics.

A potential source of error in food habits data is the failure to account for spatial correlation. Fisheries data are inherently spatially correlated with autocorrelation being found among nearby sites and decreasing as sites become further apart (Legendre 1993; Ciannelli et al. 2008). However, spatial correlation in food habits studies is often ignored. In studies where spatial correlation was explicitly taken into account, there was an increase in accuracy and precision for parameter estimates; moreover, functional relationships between abundance and environmental covariates were more clearly defined (Ciannelli et al. 2008; Cressie and Wikle 2011).

We provide a few detailed examples, from studies estimating the distribution and abundance of target species, to illustrate the importance of incorporating spatial correlation. Juvenile coho salmon (*Oncorhynchus kisutch*) abundance was compared using two Bayesian hierarchical models (Webster et al. 2008). The first model treated the sampling sites as independent, while the second model, a conditional autoregressive (CAR) model accounted for spatial correlation among sampling sites. The effects of cover class on juvenile coho salmon abundances were overestimated in the spatially-independent model while the CAR model provided more precise estimates of total stream abundance. Similarly, the precision of relative

abundance estimates for 28 groundfish species sampled on the U.S. Pacific Coast increased when spatial correlation was incorporated using a spatially-explicit delta generalized linear mixed model (Thorson et al., 2015a). On average, the spatially-explicit model had a 60% reduction in the coefficient of variation when compared to a traditional delta generalized linear mixed model that post-stratified the sampling areas. Moreover, the accuracy and precision of density-dependence estimates for three rockfish species (*Sebastes* spp.) were improved when a Gompertz model was updated to account for spatial correlation using Gaussian random fields (Thorson et al., 2015b).

Food habits data often meet the assumptions of point-referenced (or geostatistics) spatial data. One approach to account for spatial correlation in point-referenced data is a kernel density estimator. A kernel density is a nonparametric approach that can be used for any number of dimensions and is able to accurately estimate densities of any shape (Silverman 1986). In wildlife ecology, kernel density estimators have commonly been used for home range estimation (e.g. Seaman and Powell 1996), but this approach can be applied to a wide range of topics such as wildfire risk modeling (de la Riva et al. 2004; Kuter et al. 2011), identifying drought clusters (Kim et al. 2003; Dabo-Niang et al. 2014), and for establishing criteria for pollution bioindicators (Durrieu et al. 2005). Traditionally, kernel density estimators assume the data are identical and independently distributed (Silverman 1986), yet these estimators are flexible enough that they can be updated to account for lack of independence among samples (Dabo-Niang et al. 2014; Fleming et al. 2015).

To date, little work has focused on understanding the effects of spatial correlation on the accuracy and precision of diet estimators that describe fish feeding habits. The first diet estimators treated all individual fish stomachs as independent samples, regardless of sampling

method (Chipps and Garvey 2007). Cluster-based estimators (Bogstad et al. 1995; Buckel et al. 1999; Tirasin and Jørgenson 1999) were developed to address the lack of independence from multiple predators sampled from a single site using a net (e.g. trawl). Cluster-based estimators account for lack of independence within a site, but not for lack of independence between nearby sites. There is an increase in the precision of diet estimates when for the same number of fish, the number of fish sampled per location is decreased while the number of stations is increased (Bogstad et al. 1995). Increasing the number of stations per survey results in more stations being within close proximity of each other and increases the likelihood the stations will be spatially correlated and not independent. The inclusion of spatial correlation should reduce bias and increase precision when estimating diets of fish.

We present a novel spatial kernel density estimator (“spatial estimator”) that accounts for inter-sample spatial correlation when estimating the proportional contribution of prey to a predator’s diet. We compare the spatial estimator to a non-spatial estimator using simulated and empirical data to better understand the importance of spatial dependence on prey contribution estimates. Diet data were simulated using a Monte Carlo approach and the precision and accuracy of the spatial and non-spatial estimator are assessed and compared. Additionally, empirical estimates of prey contribution to three predators collected from Pamlico Sound, North Carolina are determined using the non-spatial estimator and the spatial estimator, which estimated the degree of spatial correlation between sampling sites for each predator.

Methods

Description of diet estimators

We used spatial kernel density estimation (SKDE) to account for non-independence for fish collected from both a single site and from nearby sites when estimating the diet composition

of a predator. SKDE estimates a kernel (*i.e.* a 2-dimensional probability density with dimensions $x = \text{longitude}$ and $y = \text{latitude}$) for each sampling location by interpolating the diet composition for a series of grid (*i.e.* knots) locations (s) overlaid on the area where the diet is being estimated. To compute s we overlaid a $100 * 100$ grid across the spatial domain (*i.e.* Pamlico Sound). This resulted in grids spaced 0.2 decimal degrees (DD) apart in the North-South direction and 0.1 DD apart in the East-West direction. The SKDE of the composition at a location is a locally-weighted average of the n observed compositions. Let y_{ij} be the proportion prey j contributes to the diet of the predators in sample i , and s_i be the spatial location of sample i . The estimated proportion for prey j at location s is

$$\hat{y}_j(s) = \frac{1}{W} \sum_{i=1}^n w(s - s_i) y_{ij}$$

where w is the kernel function that decreases with the distance (d) between the prediction location s and the observation location s_i and $W = \sum_{i=1}^n w(s - s_i)$ is the normalizing constant to make the weights sum to one.

We used the Gaussian kernel function $w(d) = e^{-(d_x/h_x)^2 - (d_y/h_y)^2}$ where $d = (d_x, d_y)$.

The bandwidths $h_x, h_y > 0$ are the widths of the kernel and determine the amount of spatial smoothing. If the bandwidths are close to zero then strong spatial patterns for prey are present and only the few nearest neighbors affect the estimate, whereas many points affect the estimate if the bandwidths are large and there is less spatial clustering of prey (Silverman 1986; Brunsdon and Comber 2015). The most common scenario (isotropy) is to have the same bandwidth in both directions (*i.e.* $h_x = h_y$). We allow for a different bandwidths (anisotropy) in the East-West (h_x) and North-South (h_y) directions (Brundson and Comber 2015) to account for environmental covariates such as salinity, which have an East-West gradient in our system (increasing salinity

from west to east). We incorporated a tuning parameter “stretch” to allow h_x and h_y to vary. We parameterize the bandwidth as $h_x = h$ and $h_y = h * \text{stretch}$. Stretch determines the degree of anisotropy with $\text{stretch} = 1$ giving the isotropic model with bandwidth $h_x = h_y = h$. We considered values of stretch ranging from 0.5 – 2 (increasing in increments of 0.25).

We used leave-one-out cross validation to select the bandwidths h_x and h_y . For each observation, we predicted its composition using the other $n-1$ observations and compared the predicted $(\hat{y}_j)$ and observed (y_{ij}) values using mean square error,

$$MSE = \frac{1}{n} \sum_j \sum_{i=1}^n (\hat{y}_j(s_i) - y_{ij})^2.$$

We repeated this for all possible combinations of h ranging from 0.1 – 1.0 DD (by 0.1 increments) and stretch ranging from 0.5 – 2 (by 0.25 increments) and selected the combination that minimized MSE. We would like to note, that when cross validation is repeated using different test and training data sets, different combinations of h and stretch are sometimes selected; however, the final diet estimates are always similar. To produce the final diet estimate, we used the selected bandwidths to compute the estimate on a grid of locations covering the spatial domain. The final mean proportional contribution of prey j to the predator diet (M_j) is the average of the predictions $\hat{y}_j(s)$ and is calculated as

$$M_j = \frac{1}{l} \sum_{k=1}^l \hat{y}_j(s_k)$$

where k are the individual prediction points in s and l is the total number of points in s .

Hereafter, we refer to this SKDE as the spatial estimator.

For our non-spatial estimator, we used the cluster-based estimator developed by NOAA’s Northeast Fisheries Science Center and presented in Buckel et al. (1999). The non-spatial estimate of the mean proportional contribution of a prey type by weight (W_j) is calculated as

$$W_j = \frac{\sum_{i=1}^n M_i y_{ij}}{\sum M_i}$$

where

$$y_{ij} = \frac{w_{ij}}{w_i}$$

and w_i is the total weight of all prey in the predator stomachs sampled from station i and w_{ij} is the weight of prey type j from all predator stomachs from station i . M_i is the total number of predators (e.g. bluefish, spot) captured at station i and n is the number of sites where the predator was collected and is used in weighting the samples. Samples where larger number of predators are collected are given a higher weight than samples with smaller number of predators. The non-spatial estimator accounts for lack of independence for fish at a single site by treating all fish collected in a single net as one sample and not treating the individual fish as samples. Standard errors for the non-spatial estimator were calculated based on the equation for variance provided in Buckel et al. (1999). The non-spatial estimator adjusts for lack of independence for fish collected at a single site, but not for non-independence between fish collected from nearby sites.

Simulated data

A Monte Carlo simulation study was used to compare the spatial and non-spatial estimators. Monte Carlo simulations were conducted for ten bandwidths ranging from 0.1 – 0.9 decimal degrees (DD), by 0.2 increments, which reflects the level of spatial correlation present for predators in Pamlico Sound, North Carolina (Fig. 1). For these simulations, we only modeled

isotropic conditions and had $h_x = h_y$ (*i.e.* stretch = 1). For each bandwidth, food habits data for the entire Pamlico Sound surface were simulated 1,000 times. Each data set contained 10 prey types and represented the percent contribution, by weight, of each prey type to the diet of the simulated predator at each site. The simulated diets are generated by first sampling mean counts from a log-Gaussian process, then sampling Poisson counts for each species and normalizing. Let $Z_j(s)$ be a Gaussian process for species j at location s generated independently over species with mean $E[Z_j(s)] = j/5 - 2$ and $Cov[Z_j(s), Z_j(s')] = \exp(-||s - s'||)$, where $||s - s'||$ is the Euclidean distance between locations s and s' . The number of individuals from species j observed at location s is sampled as $X_j(s) \sim \text{Poisson}\{\exp[Z_j(s)]\}$. Weights were randomly assigned to the prey individuals $X_j(s)$ as $Y_i(s) \sim \text{Gamma}(2,1)$ and the composition is the proportion $y_j(s) = Y_j(s) / \sum_i Y_i(s)$. Empirical diet data contains many 0's, with not all prey being detected in each sample and entire samples containing predators with empty stomachs. We first simulated the data as prey counts because this allowed us to incorporate these 0's in our simulated data.

Fifty sites were then randomly sampled from the simulated diets and used to estimate the predator diet using both the spatial and non-spatial estimators. MSE, Bhattacharyya distance, and bias were used to determine which estimator is most robust. First, a single MSE for all prey types was calculated from all 1,000 simulations for each estimator. MSE was calculated as

$$MSE_k = \frac{1}{n} \sum_j \sum_{i=1}^n (\hat{y}_{ijk} - y_{ijk})^2$$

where n is the number of simulated samples, y is the true diet composition, and $\hat{y}$ is the estimated diet composition using estimator k (*i.e.* spatial or non-spatial). Then, we calculated the Bhattacharyya distance (BD) for each estimator as

$$BD_k = -\ln(BC_k)$$

166

167 where

$$BC_k = \sum_{i=1}^n \sum_{j=1}^J \sqrt{(\hat{y}_{ijk})(y_{ijk})}$$

168 and J is the number of prey types and the remaining terms are defined in the above equation.

169 Finally, treating all 1,000 diet estimates as a distribution, MSE and bias were calculated for each

170 prey type. MSE was calculated as above, except by prey type and bias was calculated as

$$Bias_{jk} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_{ijk} - y_{ijk})$$

171

172 where j is the prey type and the remaining terms are defined in the above equations. A Wilcoxon

173 signed ranked test was used to compare the MSE and bias values between the spatial and non-

174 spatial estimators. We repeated these Monte Carlo simulations ($n = 500$) and analyses with

175 varying diet sample sizes (25, 50, 100, 200, and 500) to evaluate if diet sample size influenced

176 estimator choice. In the Appendix, we have included R code for creating s , simulating the data,177 selecting hx and hy using leave-one-out cross validation either through a grid search or using an

178 optimization routine, and calculating BD. The associated data files needed for our code are

179 available on Dryad.

180 *Empirical data*

181 The spatial and non-spatial estimators were compared using empirical data from the

182 stomachs of three predators: bluefish (*Pomatomus saltatrix*), inshore lizardfish (*Synodus foetens*;183 hereinafter referred to as lizardfish), and spot (*Leiostomus xanthurus*). Predators were collected

in 2012 and 2013 from two fisheries independent surveys conducted by the North Carolina Division of Marine Fisheries (NCDMF) in Pamlico Sound, North Carolina (Fig. 2). The first is a gill net survey conducted from mid-February to mid-December. For this survey, Pamlico Sound is divided into four geographical regions: Pamlico River, Neuse River, western Pamlico Sound, and eastern Pamlico Sound. Each region is further subdivided into four strata with each stratum sampled once in February and December and twice all other months (Fig. 2A). For each sampling event, one set of nets is fished in “shallow” (typically < 0.61 m) water and a second set is fished in “deep” (1.8 – 2.4 m) water with each set treated as a single sample. Floating gill nets that fish the entire water column are used to sample both the shallow and deep sites. A set is composed of eight (7.6, 8.9, 10.2, 11.4, 12.7, 14.0, 15.2, and 16.5 cm stretched mesh) 27.4 m nets for a total of 219 meters per sample (Paramore et al. 2009; Rock and Stewart 2011).

The second survey is a trawl survey conducted in June and September (Shepherd and Rock 2011a; Shepherd and Rock 2011b). Pamlico Sound is divided into seven strata (Neuse River, Pamlico River, Pungo River, Pamlico Sound east of Bluff Shoal, shallow and deep, and Pamlico Sound west of Bluff Shoal, shallow and deep) based on depth and geographic location (Fig. 2B). Shallow grids have water depths ranging from 1.8 – 3.7 m and deep grids have water depths > 3.7 m. A total of 54 stations are randomly sampled in each survey with a minimum of three stations per strata. Samples are collected by towing double-rigged demersal mongoose trawls (9.1 m headrope, 1.0 * 0.6 m doors, 2.2 cm bar mesh body, 1.9 cm bar mesh cod end, and 100 mesh tailbag extension) for 20 minutes. Each tow constitutes a single sample and 108 samples are collected annually (Rock 2009).

For both surveys, predators were stored on ice and brought back to the lab where length (TL or FL; mm), weight (0.1 kg), and sex were recorded. Predator stomachs were removed and

frozen until they were ready to be processed. Frozen stomachs were thawed and the stomach contents were identified to the lowest possible taxa and individual prey items were weighed (0.01 g) separately.

Bluefish were sampled from the gill net survey, lizardfish were predominately sampled from the trawl net survey, and spot were sampled from the trawl net survey. The spatial distribution of each predator is taken into account before the diet is estimated by only using prediction grids (s) overlapping the predator's distribution. The proportional contribution of a prey to a predator's diet for the spatial and non-spatial estimators was calculated using the equations provided above for the Monte Carlo simulation study comparisons. For the spatial estimator, we considered h_y values of 0.1 – 1.0 DD, in increments of 0.1 and stretch values of 0.5 – 2 with 0.25 increments when using cross validation to select the appropriate bandwidth values for each predator. We included empty stomachs when estimating predator diets using both the spatial and non-spatial estimators. We have included R code for both estimators in the Appendix and our Spot data set is available on Dryad.

Results

Using simulated data to compare spatial and non-spatial diet estimators

The spatial estimator performed better than the non-spatial diet estimator in predicting predator diet composition using simulated data. We first describe attributes of the simulated data and then present comparisons of MSE, BD, and bias for the two estimators.

When averaged over all 1,000 simulations, the percent contribution of each prey type increased as the prey type increased from 1 – 10. The average percent contribution for each prey type was similar across bandwidths; however, there was variability among the individually simulated data sets. The effects of increasing bandwidths had the desired effect of reducing the

amount of spatial clustering of prey. At a bandwidth of 0.1 DD, some simulated prey items (e.g. Prey 9, Prey 10) are found throughout large portions of Pamlico Sound, however, several others (e.g. Prey 1, Prey 4, and Prey 8) are more patchily distributed and form more localized clusters (Fig 3A). As the bandwidth increases to 0.5 DD the level of spatial clustering of prey among sites decreases (Fig 3B). For this particular simulation, Prey 10 is the dominant prey type and is more widespread than the common prey types were in the 0.1 DD data set, yet there are still several spatial clusters of prey items (e.g. Prey 1, Prey 3, and Prey 4). The level of spatial clustering further decreases when the bandwidth is 0.9 and there is less overall variability in the diet across Pamlico Sound. Prey 7 and Prey 8 are the dominant prey types and found at the majority of sites and although there are still rare prey types (e.g. Prey 1, Prey 3, and Prey 5), there is very little localized clustering of prey (Fig. 3C).

The overall MSE for the spatial estimator was significantly smaller than the non-spatial estimator at all bandwidths when compared using a Wilcoxon sign rank test (Fig. 4). As the bandwidth increases, MSE decreases. This holds true for both estimators. The spatial estimator was also favored using BD and had significantly smaller BD values at all bandwidths. Complete BD results are available in the Appendix (Table A1).

The spatial estimator performed better at estimating prey-specific contributions to predator diets relative to the non-spatial estimator. Except for one comparison, MSE values from the spatial estimator were smaller than the non-spatial estimator. The majority of comparisons (68%) were statistically significant with at least half of the comparisons significantly favoring the spatial estimator for each bandwidth (Fig. 5).

For the majority of bias comparisons (74%), there was no significant difference between the spatial and non-spatial estimator. Both estimators tend to be biased in the positive direction

and overestimate the contribution of a prey type to the predator diet. The spatial estimator was predominately favored and only 1 of 13 significant comparisons favored the non-spatial estimator (figure presented in Appendix; Fig. A1). Thus, the significant differences in MSE between the two estimators represent mostly differences in precision and less differences in bias.

These results held constant for the Monte Carlo simulations using varying (25, 50, 100, 200, and 500) numbers of diet samples (see Appendix Table A2 and Figures. A3 – A13 for diet sample size analysis results). The spatial estimator had significantly smaller overall MSE and BD values than the non-spatial estimator for all bandwidths and sample sizes. Prey-specific MSE values for the spatial estimator were always smaller than the non-spatial estimator and the difference was significant for the majority (86%) of comparisons. Similar to the initial prey-specific bias comparisons, most (75%) prey-specific bias comparisons were not significantly different regardless of diet sample size. Only 6 of the 32 significant bias results favored the non-spatial estimator. Diet sample size did not influence the performance of the estimators relative to each other, but as expected, MSE and bias decreased with increasing sample size for both estimators.

Using empirical data to compare spatial and non-spatial diet estimators

Bluefish ($n = 1,315$) sampled from the gill net survey ranged in fork length (FL) from 134 - 820 mm with a mean FL ($\mu \pm \text{SE}$) of 318.7 ± 2.5 mm. Bluefish were distributed throughout the entire sampling area, so the prediction sites for diet estimation did not need to be truncated. Overall, bluefish collected at a single site do not show much variability in their diets and typically all bluefish from a single collection only consumed a few prey types (Fig. 6A). Spatial correlation was strong for bluefish diets, with an estimated bandwidth of 0.2 DD (~ 22.4 km) for both the North-South and East-West directions (stretch = 1). This strong correlation

results in there being small-scale clustering of sites where a single prey type was consumed, resulting in high variability in diets across the sampling area. Also, the high-level of spatial clustering leads to the two estimators producing very different estimates for the majority of prey types (Fig. 7A).

For bluefish, the spatial and non-spatial estimators differed the most for Atlantic menhaden (*Brevoortia tyrannus*) and spot. The contribution of Atlantic menhaden to bluefish diet ($\mu \pm \text{SE}$) was 19% lower using the spatial estimator (0.317 ± 0.008) than the non-spatial estimator (0.377 ± 0.004). Atlantic menhaden were predominately found in samples from the western portion of the sampling area (Fig. 6A). The spatial estimator adjusts for Atlantic menhaden being more important in western sites and decreases their overall contribution to bluefish diet. For spot, the contribution to bluefish diet was nearly doubled using the spatial estimator (0.187 ± 0.003) rather than the non-spatial estimator (0.101 ± 0.002). Spot were commonly found in diets of bluefish throughout the entire sampling area, and were especially important in the Pamlico and Neuse Rivers.

Lizardfish FL ranged in size from 105 – 358 mm with a mean FL of 199.1 ± 1.6 mm. We combined lizardfish from the gill net survey ($n = 2$) with those sampled from the trawl survey ($n = 574$). Lizardfish are distributed in the main portion of Pamlico Sound and only enter the lower portions of the Pamlico and Neuse Rivers. Therefore, their diets were only estimated in this truncated portion of the sampling area. The level of spatial correlation for diets in the North-South direction was moderate with a bandwidth of 0.5 DD (~ 55.9 km). Spatial correlation was stronger in the East-West direction with a bandwidth of 0.25 DD (stretch = 2). Unlike bluefish, there were no small spatial scale clusters where prey were consumed, but some larger-scale clusters are present. Spot were consumed more in the western half of the lizardfish distribution

while squid (Loliginidae) were consumed more heavily in the eastern half. Pigfish (*Orthopristis chrysoptera*) and anchovies (*Anchoa mitchilli* and *A. hepsetus*) were consumed by lizardfish throughout their entire distribution (Fig 6B). The two estimators differ for the majority of prey types, but these differences are not as pronounced as those observed for bluefish (Fig. 7B). For weakfish (*Cynoscion regalis*), the estimate of proportional contribution was much larger with the spatial estimator (0.077 ± 0.003) compared to the non-spatial estimator (0.029 ± 0.001). Weakfish were only found in lizardfish diets at 24 of 74 sampled sites, but these stations were clustered in the northeast portion of the sampling area (Fig. 6B).

Spot ($n = 719$) were sampled from the trawl net survey and ranged in size from 100 – 223 mm; with a mean FL of 143.4 ± 0.7 mm. Spot are distributed throughout the entire sampling area, so the distribution of their prediction sites was not truncated. There was a weak correlation in the North-South direction, bandwidth = 0.7 DD (~ 78.0 km) and a moderate correlation East-West direction with a bandwidth of 0.35 DD (stretch = 2). Since there were not strong spatial patterns in either direction, the estimates from the two estimators were fairly similar for the majority of prey types, but there were a few differences. Spot diet is dominated by mollusks (Bivalvia and Gastropoda) and polychaetes (Polychaeta) and these two prey types comprise over 50% of the diet, regardless of which estimator is being used (Fig. 7C). Both of these prey types are consumed throughout the entire sampling area and are typically the most common prey types for each sample (Fig. 6C). The spatial estimator had slighter lower estimates for both mollusks (0.355 ± 0.004) and polychaetes (0.287 ± 0.004) than the non-spatial estimator (mollusks = 0.366 ± 0.005 , polychaetes = 0.327 ± 0.005). Two prey types, gammarid amphipods (Gammaridae) and porcelain crabs (Porcellanidae), were important as prey items in more localized clusters which resulted in an $\sim 50\%$ relative increase in diet contribution using the spatial estimator (Figs.

6C, 7C); Gammarid amphipods (non-spatial = 0.046 ± 0.002 , spatial = 0.073 ± 0.001) were consumed in the central part of Pamlico Sound, while porcelain crabs (non-spatial = 0.056 ± 0.002 , spatial = 0.082 ± 0.004) formed a cluster in the eastern portion.

Discussion

Food habits data are required for multispecies and ecosystem models which are increasingly used for fisheries management (Hollowed et al. 2000; Link 2010; Fogarty et al. 2012). The quality of food habits data used in these models can be of varying quality and sources of error associated with these data should be minimized (Ainsworth et al. 2010, Masi et al. 2014). We present a novel approach for estimating the proportional contribution of a prey to a predator's diet that accounts for both intra- and inter- haul correlation. The spatial estimator performed better than the non-spatial estimator with lower MSE and BD values at all levels of prey clustering (small to large bandwidths). Bias was fairly similar between the two estimators (or slightly favored the spatial estimator); thus, the differences in MSE values for the two estimators largely result from lower variance with the spatial estimator. Diet estimates for predators sampled in Pamlico Sound also differed between the two estimators, especially when prey consumption was locally clustered and not homogenous throughout the region. To our knowledge, this is the first time a spatially-explicit diet estimator has been developed and evaluated.

When correlated data are treated as independent, estimates of variance are often biased low which leads to unrealistically small standard errors (Tirasen and Jorgensen 1999). This pattern was observed in our analysis of survey data for bluefish, which had the strongest degree of correlation. By treating spatially-correlated data as independent, the non-spatial model assumes the data are more informative than they truly are resulting in smaller standard errors for

the majority of bluefish prey types. However, for lizardfish and spot, we saw increases in precision when using the spatial estimator. This matches other studies (Webster et al. 2008; Thorson et al. 2015a) and results from our simulated data which also saw increases in the precision when using spatial models. When the data have low to moderate levels of correlation, incorporating spatial correlation increases the precision for model estimates. Our findings suggest that even for weakly spatially correlated data, spatial correlation needs to be included when estimating fish diets.

For empirical data, the spatial and non-spatial estimator provided different results when a particular prey type was widespread or clustered. For example, spot had a widespread spatial distribution as bluefish prey; this resulted in the spatial estimator having a higher contribution of spot than the non-spatial estimator. When the diets are being interpolated, the diets at each site are essentially weighted averages with nearby sites being more heavily weighted than further sites. When spatial correlation is strong, nearby sites are providing a majority of the information used to interpolate the diet at a site. Since spot are commonly consumed throughout the entire sampling area by bluefish, they are providing information to more of the interpolated sites than any other prey type. This was observed in lizardfish predators for weakfish prey and in spot predators for gammarid amphipod and porcelain crab prey. For lizardfish, weakfish were consumed in a third of the samples covering ~ 20% of lizardfish distribution. Weakfish were very important prey in this region and when estimating the diet composition, the spatial estimator incorporates the importance of weakfish in this region whereas the non-spatial estimator did not. This is also why the proportional contributions for gammarid amphipods and porcelain crabs to spot diet were increased with spatial estimates of diet despite only being consumed at 18 and 12 of the possible 122 sites, respectively. The non-spatial estimator only accounts for the number of

samples the prey was found in, not where the prey was consumed, so the contributions of locally important prey are often underestimated.

The two estimators use different approaches for weighting the observed diet data when estimating the final diet composition of predators. The spatial estimator weights the sample based on location, with nearby sites being more heavily weighted than distant sites (the degree to which site is weighted is determined using the bandwidth and stretch parameters). The non-spatial estimator weights the samples based on the number of predators collected at a single site. Sites with large sample sizes of predators are weighted more heavily than sites with lower sample sizes. Both approaches can introduce bias into the final diet estimates. For the spatial estimator, not weighting by the number of predators can increase bias if the proportional diet composition varies as a function of predator sample size. The spatial estimator is flexible and can incorporate both location and number of predators collected at a site when estimating the final diet composition.

From our simulated data, the majority of prey-specific bias comparisons for both estimators were positive (Figure A1). We feel this is an artifact of how we simulated the data. We wanted the simulated data to closely resemble “real” diet data, so we incorporated a large portions of zeroes and simulated the diet data to be right-skewed, which resulted in bias being predominately positive. To further investigate this, we resimulated the diet data using a Uniform distribution with parameters $a = 1$ and $b = 10$. As expected, when the diet data were homogenous and not skewed, the bias results were more of an equal mix of positive and negative values (Figure A2) and was very low.

There is an increased focus on identifying forage fish (Pikitch et al. 2012; Houde et al. 2014) and a recent qualitative approach uses a minimum proportional contribution for a prey of

5% to multiple predator diets as one of their criteria (Houde et al. 2014). For each predator, there were prey whose forage fish label ($<$ or $>$ than 5% contribution) differed depending on the type of estimator used. For bluefish, both silver perch and Atlantic croaker (*Micropogonias undulatus*) proportional contributions exceeded 5% only with the spatial estimator while the contribution of anchovies was above 5% for the non-spatial but below this threshold for the spatial estimator. Relative to the non-spatial estimates, both silver perch and weakfish proportional contributions from the spatial estimator increased over 5% for lizardfish while gammarid amphipods did so for spot. Not accounting for spatial correlation led to underestimates of diet contribution for four important prey types and the resulting potential misclassification as non-forage fish.

When using the spatial estimator, it is important to determine the distribution of the predator before estimating the diet and to have prediction points mirror the predator's distribution within the sampling area. Failure to do so can lead to an increased bias in the estimates because the edge locations of the sampling area (outside of the predator's distribution) will have a greater impact on diet estimates. For the diets estimated from survey data, lizardfish were the only species with a truncated prediction area relative to the sampling area and their diets were estimated only for the area overlapping their distribution. At the edge of their distribution, anchovy and spot are the dominant prey types. If we estimate lizardfish diet for the entire sampling area, and not just the truncated area to align with their distribution, the estimates for anchovy and spot increase while the estimates for most other prey types decline (see Appendix, Fig. A14). For this example, lizardfish are not found in a small portion of the sampling area, so the differences between the two estimates are modest. This difference would become more pronounced for species with distributions only covering a small portion of the sampling area.

This issue may be more important in estuarine environments with strong salinity gradients or for species with habitat-specific distributions (e.g. reefs, seagrass beds).

The spatial estimator is very flexible and can be used to estimate diet composition for multiple scenarios. While we focused on the fact the spatial estimator addresses both intra- and inter-haul correlation, this estimator can also be used when predators are collected individually (e.g. hook and line). We presented the spatial estimator as a method to estimate the proportional contribution by weight of a prey to a predator's diet but this estimator can be easily modified to estimate frequency of occurrence instead.

One limitation of SKDE is that for data from long-tailed distributions erroneous noise often appears in the tails of the estimates. This is caused by having the bandwidth fixed across the entire sample. It is possible to overcome this noise by increasing the bandwidth, but this has a negative impact on the estimates from the main part of the distribution (Silverman 1986). Allowing h_x and h_y to vary and not be equal also helps reduce this noise. Another limitation is the inability to directly incorporate covariates into the model. Abiotic factors, such as salinity and temperature, typically influence fish distribution and can result in clustering of fish (both predators and prey). We were able to indirectly incorporate covariates like salinity, which has an East-West gradient in our system by allowing h_x and h_y to have different values, but we were not able to adjust for any covariates not following a spatial gradient. There are other approaches to modeling spatial data that allows for the direct modeling of covariates, such as fitting an inhomogeneous Poisson process where the covariates are related to the Poisson process intensity or through the use of hierarchical modeling with the kernel density as a process model and an observation model to account for observation error. Many hierarchical and Bayesian approaches model the entire data set instead of using "testing" and "training" data sets when identifying the

amount of spatial correlation present in the data. To our knowledge, spatial correlation has not been incorporated into an estimator describing the diets of fish and we wanted to use a simple spatial model as our first approach; however, these more complex approaches could potentially improve our diet estimates.

While this estimator does result in a decrease in MSE when compared to traditional diet estimators, it still has the usual shortcomings associated with estimating fish diets. Unidentifiable prey and differential digestion rates for different prey types both have the potential to lead to biased estimates, especially for biomass-based metrics (Hyslop 1980; Baker et al. 2014). There are study-based approaches to help mitigate these problems such as conducting feeding periodicity studies so food habits sampling can target active feeding times to help limit the amount of digestion (Hyslop 1980). This type of approach though would be harder for food habits studies targeting multiple predators whose feeding periods may not be synchronous and it does not completely eliminate unidentifiable prey. Another approach is to collect a sufficiently large sample (> 100 stomachs) because there tends to be good agreement between frequency-based estimators, which are less influenced by these problems, and biomass-based estimators for large sample sizes (Hynes 1950; Hyslop 1980).

It is imperative that food habits data capture as much of the true predator-prey dynamics as possible. For example, underestimates of diet contribution can cause unrealistically low natural mortality estimates and decreased prey importance within a food web model. While the estimates for proportional contribution were similar for many prey types, there were noticeable differences in the estimates when spatial correlation was high or when the prey were consumed in localized clusters. Many of the prey types with the most pronounced differences are species that are both harvested and managed. The non-spatial estimator underestimated the contribution

of weakfish to lizardfish diet and the contributions of both spot and Atlantic croaker to bluefish diet. For bluefish, the spatial estimator had an 85% relative change in the contribution of spot. If a multispecies model using the non-spatial estimator had annual predation estimates of 10 tonnes of spot consumed by bluefish this would increase to 18.5 tonnes using the spatial estimator. Both lizardfish and bluefish are abundant in Pamlico Sound in terms of both numbers and biomass, so even a modest change in prey contribution to diet translates to a large change in the number of prey removed through predation.

The spatial estimator outperformed the non-spatial estimator on simulated data and there were large differences between the two estimators in prey contribution to the three predators for particular prey types. This study strongly suggests the spatial estimator should be used in future diet studies when predators are collected at sites that have potential for non-independence, even if the level of correlation is weak. Adjusting for spatial dependence in conjunction with a well-designed food habits study should provide reliable data for analyses where food habits data are required as input.

Acknowledgements

This work was supported by a North Carolina Coastal Recreational Fishing License Grant (Grant Number 4044). We would like to thank R. Christian, D. Eggleston, and K. Shertzer, two anonymous reviewers, and the associate editor for their helpful comments and reviews which improved the quality of this manuscript.

Works Cited

- Ainsworth, C. H., Kaplan, I. C., Levin, P. S., and Mangel, M. 2010. A statistical approach for estimating fish diet compositions from multiple data sources: Gulf of California case study. *Ecological Applications* **20**(8): 2188-2202.
- Baker, R., Buckland, A., and Sheaves, M. 2014. Fish gut content analysis: robust measures of diet composition. *Fish and Fisheries* **15**: 170-177.
- Bogstad, B., Pennington, M., and Vølstad, J. H. 1995. Cost-efficient survey designs for estimating food consumption by fish. *Fisheries Research* **23**: 37-46.
- Brunsdon, C. and Comber, L. 2015. An introduction to R for spatial analysis and mapping. Sage Publications Inc., Thousand Oaks, California.
- Buckel, J. A., Conover, D. O., Steinberg, N. D., and McKown, K. A. 1999. Impact of age-0 bluefish (*Pomatomus saltatrix*) predation on age-0 fishes in the Hudson River estuary: evidence for density-dependent loss of juvenile striped bass (*Morone saxatilis*). *Canadian Journal of Fisheries and Aquatic Sciences* **56**(2): 275-287.
- Chipps, S.R. and Garvey, J. E. 2007. Assessment of diets and feeding patterns. *In* Analysis and interpretation of freshwater fisheries data. Edited by C.S. Guy and M.L. Brown, editors. American Fisheries Society Bethesda, Maryland. pp. 473-514.
- 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.
- Cressie, N. and Wikle, C. K. 2011. Statistics for spatio-temporal data. John Wiley and Sons, Inc. Hoboken, New Jersey.

- 503 Dabo-Niang, S., Hamdad, L., Ternynck, C., and Yao, A.-F. 2014. A kernel spatial density
504 estimation allowing for the analysis of spatial clustering. Application to Monsoon Asia Drought
505 Atlas data. Stochastic Environmental Research and Risk Assessment **28**(8): 2075-2099.
- 506 de la Riva, J., Pérez-Cabello, F., Lana-Renault, N., and Koutsias, N. 2004. Remote Sensing of
507 Environment **92**: 288-294.
- 508 Durrieu, G., Maury-Brachet, R., and Boudou, A. 2005. Goldmining and mercury contamination
509 of the piscivorous fish *Hoplias aimara* in French Guiana (Amazon basin). Ecotoxicology and
510 Environmental Safety **60**: 315-323.
- 511 Fleming, C. H., Fagan, W. F., Mueller, T., Olson, K. A., Leimgruber, P., and Calabrese, J. M.
512 2015. Rigorous home range estimation with movement data: a new autocorrelated kernel
513 density estimator. Ecology **96**(5): 1182-1188.
- 514 Fogarty, M. J., Overholtz, W. J., and Link, J. S. 2012. Aggregate surplus production models for
515 demersal fishery resources of the Gulf of Maine. Marine Ecology Progress Series **459**: 247-258.
- 516 Hollowed, A. B., Bax, N., Beamish, R., Collie, J., Fogarty, M., Livingston, P., Pope, J., and Rice,
517 J. C. 2000. Are multispecies models an improvement on single-species models for measuring
518 fishing impacts on marine ecosystems? ICES Journal of Marine Sciences **57**:707-719.
- 519 Houde, E., Gaichas, S., and Seagraves, R. 2014. Managing forage fishes in the mid-Atlantic
520 region: a white paper to inform the mid-Atlantic fishery management council. 42 pp.
- 521 Hynes, H. B. N., 1950. The food of fresh-water Sticklebacks (*Gasterosteus aculeatus* and
522 *Pygosteus pungitius*), with a review of methods used in studies of the food of fishes. Journal of
523 Animal Ecology **19**(1): 36-58.
- 524 Hyslop, E. J., 1980. Stomach contents analysis-a review of methods and their application.
525 Journal of Fish Biology **17**: 411-429.

- 526 Kim, T.-W., Valdés, J. B., and Yoo, C. 2003. Nonparametric approach for estimating return
527 periods of droughts in arid regions. *Journal of Hydrologic Engineering* **8**(5):237-246.
- 528 Kuter, S., Usul, N., and Kuter, N. 2011. Bandwidth determination for kernel density analysis of
529 wildfire events at forest sub-district scale. *Ecological Modelling* **222**: 3033-3040.
- 530 Latour, R. J., Brush, M. J., and Bonzek, C. F. Toward ecosystem-based fisheries management.
531 *Fisheries* **28**(9): 10-22.
- 532 Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74(3):1659-
533 1673.
- 534 Link, J. 2010. Ecosystem-based fisheries management confronting tradeoffs. Cambridge
535 University Press, New York.
- 536 Masi, M. D., Ainsworth, C. H., and Chagaris, D. 2014. A probabilistic representation of fish
537 diet compositions from multiple data sources: A Gulf of Mexico case study. *Ecological*
538 *Modelling* **284**: 60-74.
- 539 NC OneMap GeoSpatial Portal. <http://data.nconemap.gov/geoportal/catalog/main/home.page>.
- 540 Paramore, L., Stewart, C., and Rock, J. E. 2009. Biological Program Documentation: Program
541 915 Fisheries Independent Assessment Composed of Pamlico Sound Independent Gill Net
542 Survey, Pamlico and Neuse Watershed Survey, Southern District Gill Net Survey, and
543 Presence/Absence of Disease Sampling. Division of Marine Fisheries, Morehead City, North
544 Carolina.
- 545 Pikitch, E., Boersma, P. D., Boyd, I. L., Conover, D. O., Cury, P., Essington, T., Heppell, S. S.,
546 Houde, E. D., Mangel, M., Pauly, D., Plagányi, É., Sainsbury, K., and Steneck, R. S. 2012. Little
547 fish, big impact: managing a crucial link in ocean food webs. Lenfest Ocean Program.
548 Washington, DC.

- 549 Rock, J. E. 2009. Biological Program Documentation: Program 195 Pamlico Sound
550 Independent Fishery Survey. North Carolina Division of Marine Fisheries. Morehead City,
551 North Carolina.
- 552 Rock J. E. and Stewart, C. 2011. North Carolina Division of Marine Fisheries five-year project
553 for reoccurring funds from the marine resource fund job 4: fisheries independent assessment
554 program. Annual report. Morehead City, North Carolina.
- 555 Seaman, D. E. and Powell, R. A. 1996. An evaluation of the accuracy of kernel density
556 estimators for home range analysis. *Ecology* **77**(7): 2075-2085.
- 557 Shepherd M. and Rock, J. E. 2011(a). Pamlico Sound survey June 2010 cruise report. Final
558 Report. North Carolina Division of Marine Fisheries, Morehead City, North Carolina.
- 559 Shepherd M. and J. E. Rock. 2011(b). Pamlico Sound survey September 2010 cruise report.
560 Final Report. North Carolina Division of Marine Fisheries, Morehead City, North Carolina.
- 561 Silverman, B. W. 1986. Density Estimation for Statistics and Data Analysis. Chapman and
562 Hall Ltd. New York, New York.
- 563 Thorson, J. T., Shelton, A. O., Ward, E. J., and Skaug, H. J. 2015(a). Geostatistical delta-
564 generalized linear mixed models improve precision for estimated abundance indices for West
565 Coast groundfishes. *ICES Journal of Marine Science* **72**(5): 1297-1310.
- 566 Thorson, J. T., Skaug, H. J., Kristensen, K., Shelton, A. O., Ward, E. J., Harms, J. H., and
567 Benante, J. A. 2015(b). The importance of spatial models for estimating the strength of density
568 dependence. *Ecology* **96**(5): 1202-1212.
- 569 Tirasin, E. M. and Jørgensen, T. 1999. An evaluation of the precision of diet description.
570 *Marine Ecology Progress Series* **182**: 243-252.

571 Webster, R. A., Pollock, K. H., Ghosh, S. K., and Hankin, D. G. 2008. Bayesian spatial
572 modeling of data from unit-count surveys of fish in streams. Transactions of the American
573 Fisheries Society **137**(2): 438-453.
574

Draft

List of Figures

Fig. 1: Map showing bandwidths ranging from 0.1 to 1.0 decimal degrees (DD). The black circle is the reference site from which distances are measured. Sampling sites (gray squares) located at distances $\leq$ the bandwidth identified from cross-validation would not be considered independent. Map was created in ArcMap 10.3.1. The shape file for the state boundary of North Carolina was accessed through NC OneMap GeoSpatial Portal.

Fig 2: Locations of strata for the (A) gill net and (B) trawl surveys conducted by NCDMF in Pamlico Sound, North Carolina. The shape file for the state boundary of North Carolina was accessed through NC OneMap GeoSpatial Portal.

Fig. 3: Diet composition at 50 randomly sampled sites for 3 simulated data sets with bandwidths of (A) 0.1 decimal degrees (DD), (B) 0.5 DD, and (C) 0.9 DD. Stretch = 1 for all. The shape file for the state boundary of North Carolina was accessed through NC OneMap GeoSpatial Portal.

Fig. 4: Overall MSE (pooled across all prey types) results comparing the spatial and non-spatial estimators. Comparisons with *** indicates the p value for the Wilcoxon Rank Sum test was < 0.001 .

Fig. 5: Prey-specific MSE results comparing the spatial and non-spatial estimators at (A) decimal degrees (DD), (B) 0.3 DD, (C) 0.5 DD, (D) 0.7 DD, (E) 0.9 DD bandwidths. Comparisons with a * are significant at the $p < 0.5$ level, ** at the $p < 0.01$ level, and *** at the $p < 0.001$ level for a Wilcoxon Rank Sum test.

Fig. 6: Proportional diet by sampling site for (A) bluefish, (B) lizardfish, and (C) spot sampled in Pamlico Sound, North Carolina. The shape file for the state boundary of North Carolina was accessed through NC OneMap GeoSpatial Portal.

Fig. 7: Proportional diet composition by weight for (A) bluefish, (B) lizardfish, and (C) spot estimated using both the spatial and non-spatial diet estimators.

Figure 1

Figure 2

Figure 3

Figure 4

Draft

Figure 5

Figure 6

Figure 7

Appendix

Data Files (available as individual files on Dryad)

Pamlico_Lat_Longs.csv – List of latitude and longitude coordinates for Pamlico Sound, NC.
This file is used when generating *s*.

Lat – Latitude in decimal degrees

Long – Longitude in decimal degrees

Pamlico_Waters_Extended_wgs (file extensions .cpg, .dbf, .prj, .sbn, .sbx, shp, and .shx) –
Shape file for Pamlico Sound. This file is needed when generating *s*.

s_predict.csv – Locations for *s*, needed when estimating the diet using the spatial estimator and for simulating diet data across the Pamlico Sound surface.

long.pred – Longitude in decimal degrees

latitude.pred – Latitude in decimal degrees

spot_trawl.csv – Data used for estimating Spot diet using the spatial and non-spatial estimators.

SourceID – Sequence number linking diet data to corresponding field data

StomachCount – The number of predator stomachs sampled for food habits

NetCount – The number of predators collected in the net. *M* for the non-spatial estimator.

YEAR – Year

MONTH – Month

DAY – Day

Lat – Latitude in decimal degrees

Long – Longitude in decimal degrees

Columns 9 - 21 (*Goby*, *Mollusks*, *Squid*, *Polychaete*, *Gammarid_Amphipod*, *Mysids*, *Barnacle*, *Penaeid_Shrimp*, *Pea_Crab*, *Purse_Crab*, *Porcelain_Crab*, *Brittle_Star*, *Macroalgae*) – Proportional contribution of the prey to the diets of Spot sampled at that location. w_{ik} for the non-spatial estimator, y_{obs} for the spatial estimator.

- 31 R Code (available as separate text files)
- 32 Creating S Predict Grids
- 33 Cross Validation using Grid Search
- 34 Cross Validation using Optimization
- 35 Data Simulation
- 36 Estimating Diet using NonSpatial Estimator
- 37 Estimating Diet using Spatial Estimator
- 38 Monte Carlo Simulation Study and Sample Size Comparison
- 39 Bhattacharyya Distance
- 40

Draft

Table A1: Comparison of Bhattacharyya Distance (BD) for the spatial and non-spatial estimator at bandwidths 0.1, 0.3, 0.5, 0.7, and 0.9 DD. Values for BD represent the sum for all 1,000 Monte Carlo simulations and p is from a Wilcoxon Rank Sum Test.

Bandwidth (DD)	BD		p
	Spatial	Non-Spatial	
0.1	10.728	11.644	< 0.001
0.3	9.444	10.121	< 0.001
0.5	8.784	9.281	< 0.001
0.7	8.013	8.352	< 0.001
0.9	7.966	8.329	< 0.001

Table A2: Comparison of BD for spatial and non-spatial estimators at diet sample sizes of 25, 50, 100, 200, and 500 and bandwidths 0.1, 0.3, 0.5, 0.7, and 0.9 decimal degrees (DD). Values for Bhattacharyya distance (BD) represent the sum from 500 Monte Carlo simulation studies and p is from a Wilcoxon Rank Sum test.

Bandwidth (DD)	Sample Size	BD		p
		Spatial	Cluster	
0.1	25	11.158	11.589	< 0.001
	50	6.288	6.672	< 0.001
	100	3.978	4.279	< 0.001
	200	2.828	2.993	< 0.001
	500	2.238	2.285	< 0.001
0.3	25	9.595	10.751	< 0.001
	50	5.543	6.317	< 0.001
	100	3.670	4.087	< 0.001
	200	2.805	3.008	< 0.001
	500	2.315	2.354	< 0.001
0.5	25	8.561	9.576	< 0.001
	50	4.755	5.521	< 0.001
	100	3.228	3.720	< 0.001
	200	2.630	2.882	< 0.001
	500	2.248	2.313	< 0.001
0.7	25	7.612	8.923	< 0.001
	50	4.435	5.37	< 0.001
	100	3.082	3.457	< 0.001
	200	2.466	2.638	< 0.001
	500	2.041	2.092	< 0.001
0.9	25	6.882	7.856	< 0.001
	50	4.354	4.978	< 0.001
	100	3.027	3.330	< 0.001
	200	2.505	2.697	< 0.001
	500	2.155	2.211	< 0.001

Figure A1: Comparison of prey-specific bias for the spatial and non-spatial estimators at bandwidths (A) 0.1 decimal degrees (DD), (B) 0.3 DD, (C) 0.5 DD, (D) 0.7 DD, and (E) 0.9 DD with 50 diet samples and $n = 1,000$ Monte Carlo simulations. Comparisons with a * are significant at the $p < 0.05$ level, ** at the $p < 0.01$ level, and *** at the $p < 0.001$ level for a Wilcoxon Rank Sum test.

57
58 Figure A2: Comparison of prey-specific bias for the spatial and non-spatial estimators for bandwidths (A) 0.1 decimal degrees (DD),
59 (B) 0.3 DD, (C) 0.5 DD, (D) 0.7 DD, and (E) 0.9 DD with 50 diet samples when the diet data were generated ($n = 1,000$) from a
60 Uniform distribution. Comparisons with an * indicates a p value < 0.05 for a Wilcoxon Rank Sum test. Note difference in y-axis
61 scale from Figure A1.

62

Figure A3: Comparison of the overall MSE for the spatial and non-spatial estimators at diet sample sizes of 25, 50, 100, 200, and 500 ($n = 500$) at bandwidths (A) 0.1 decimal degrees (DD), (B) 0.3 DD, (C) 0.5 DD, (D) 0.7 DD, and (E) 0.9 DD. Comparisons with *** indicates the p value for a Wilcoxon Rank Sum test was < 0.001 .

69

70 Figure A4: Comparison of prey-specific MSE for the spatial and non-spatial estimators for bandwidth 0.1 decimal degrees ($n = 500$)
71 for (A) 25, (B) 50, (C) 100, (D) 200, and (E) 500 diet sample sizes. Comparisons with a * are significant at the $p < 0.5$ level, ** at the
72 $p < 0.01$ level, and *** at the $p < 0.001$ level for a Wicoxon Rank Sum test.

Figure A5: Comparison of prey-specific MSE for the spatial and non-spatial estimators for bandwidth 0.3 decimal degrees ($n = 500$) for (A) 25, (B) 50, (C) 100, (D) 200, and (E) 500 diet sample sizes. Comparisons with a * are significant at the $p < 0.5$ level, ** at the $p < 0.01$ level, and *** at the $p < 0.001$ level for a Wilcoxon Rank Sum test.

78

79 Figure A6: Comparison of prey-specific MSE for the spatial and non-spatial estimators for bandwidth 0.5 decimal degrees ($n = 500$)
80 for (A) 25, (B) 50, (C) 100, (D) 200, and (E) 500 diet sample sizes. Comparisons with a * are significant at the $p < 0.5$ level, ** at the
81 $p < 0.01$ level, and *** at the $p < 0.001$ level for a Wilcoxon Rank Sum test.

Figure A7: Comparison of prey-specific MSE for the spatial and non-spatial estimators for bandwidth 0.7 decimal degrees ($n = 500$) for (A) 25, (B) 50, (C) 100, (D) 200, and (E) 500 diet sample sizes. Comparisons with a * are significant at the $p < 0.5$ level, ** at the $p < 0.01$ level, and *** at the $p < 0.001$ level for a Wilcoxon Rank Sum test.

86

87 Figure A8: Comparison of prey-specific MSE for the spatial and non-spatial estimators for bandwidth 0.9 decimal degrees ($n = 500$)
88 for (A) 25, (B) 50, (C) 100, (D) 200, and (E) 500 diet sample sizes. Comparisons with a * are significant at the $p < 0.05$ level, ** at the
89 $p < 0.01$ level, and *** at the $p < 0.001$ level for a Wilcoxon Rank Sum test

Figure A9: Comparison of prey-specific bias for the spatial and non-spatial estimators for bandwidth 0.1 decimal degrees ($n = 500$) for (A) 25, (B) 50, (C) 100, (D) 200, and (E) 500 diet sample sizes. Comparisons with a * are significant at the $p < 0.5$ level for a Wicoxon Rank Sum test.

95

96 Figure A10: Comparison of prey-specific bias for the spatial and non-spatial estimators for bandwidth 0.3 decimal degrees ($n = 500$)
97 for (A) 25, (B) 50, (C) 100, (D) 200, and (E) 500 diet sample sizes. Comparisons with a * are significant at the $p < 0.5$ level for a
98 Wilcoxon Rank Sum test.

99

Figure A11: Comparison of prey-specific bias for the spatial and non-spatial estimators for bandwidth 0.5 decimal degrees ($n = 500$) for (A) 25, (B) 50, (C) 100, (D) 200, and (E) 500 diet sample sizes. Comparisons with a * are significant at the $p < 0.05$ level and ** at the $p < 0.01$ level for a Wilcoxon Rank Sum test.

103

104 Figure A12: Comparison of prey-specific bias for the spatial and non-spatial estimators for bandwidth 0.7 decimal degrees ($n = 500$)
105 for (A) 25, (B) 50, (C) 100, (D) 200, and (E) 500 diet sample sizes. Comparisons with a * are significant at the $p < 0.05$ level, ** at the
106 $p < 0.01$ level, and *** at the $p < 0.001$ level for a Wilcoxon Rank Sum test.

107

Figure A13: Comparison of prey-specific bias for the spatial and non-spatial estimators for bandwidth 0.9 decimal degrees ($n = 500$) for (A) 25, (B) 50, (C) 100, (D) 200, and (E) 500 diet sample sizes. Comparisons with a * are significant at the $p < 0.05$ level, ** at the $p < 0.01$ level, and *** at the $p < 0.001$ level for a Wilcoxon Rank Sum test.

115

116 Figure A14: Comparison of the proportional diet composition by weight for lizardfish estimated using prediction grids overlapping

117 the entire sampling area or truncated to mirror lizardfish distribution.