ASSESSING THE EFFECTS OF SINGLE VERSUS MULTIPLE FLOW PATH DIRECTION METHODS ON SPECIES-ENVIRONMENTAL RELATIONSHIPS AND ESTIMATED SPECIES DISTRIBUTIONS

Gregory B. Anderson

INTRODUCTION

Species distributional models (SDMs) are powerful tools for making inferences about relationships between species and their environments (Franklin 2009). By relating spatial distributional patterns of species (e.g., presence or absence of a species) to environmental predictor variables (e.g., landuse), SDMs provide a framework to predict the response of a species to environmental covariates and to predict the occurrence or abundance of a species throughout a landscape (Guissan and Thuiller 2005). These environmental covariates essentially represent the state of a locality and/or system at a particular time (and possibly multiple times) and are thought to represent features of the resource selection of the species of interest.

With an increase in the availability of data collected from Geographic Information Systems (GIS) and remote sensing, environmental covariates used to model species distributions are commonly remotely derived. For example, SDM strategies such as Genetic Algorithm for Rule Set Production (GARP; Stockwell and Peters 1999) and Maximum Entropy (MaxEnt; Phillips et al. 2006) rely exclusively on data available from/for a GIS. Although studies have demonstrated that remotely derived data can be more predictive of species distributional patterns than field data (Netwon-Cross 2007), there is a large amount of uncertainty that is associated with remotely derived data. For example, there are more than 10 different algorithms to derive slope using elevation data within a GIS (Weih 1991). This uncertainty in selecting the best approach for a procedure, in addition to the existing error within the original data, has rarely been acknowledged within SDMs. However, many of the variables used to predict species distributions (e.g., elevation, landuse, slope, etc.) come from similar spatial analysis procedures.

In order to evaluate the effect of different spatial analyst procedures available on the estimation of species environmental relationships (e.g., regression coefficients) and the eventual prediction of the species distribution across the landscape, I investigated the sensitivity associated with using single versus multiple flow path direction methods with both planimetric and surface area for automatic watershed delineation in a GIS. Watershed characteristics such as watershed area, watershed landuse, and watershed shape have been shown to be good predictors of freshwater fish

distributions (Walters et al. 2003, Wenger et al. 2008, Roy et al. 2007, Anderson et al. 2012); however, all of these parameters depend upon reliable delineation methods of upstream catchments of sample sites. To examine the potential effect of different watershed delineation procedures, I investigated the sensitivity associated with 1) single versus multiple flow direction methods, 2) planimetric versus surface area, and 3) cell resolution (10 meter versus 30 meter) on estimated species-environmental relationships and predicted species distributions within a dataset of fluvial fish occurrences from the upper Etowah Watershed in northern Georgia. More specifically, my objectives were to: 1) create eight upstream catchments for each sample site based on i) a single directional method (i.e., flow goes only to a single cell downstream; Jenson and Domingue 1988) and a multiple directional model (i.e., flow can be proportioned among neighboring downslope cells; Tarboton 1997), ii) planimetric and surface area, and iii) 10 and 30 meter resolution digital elevation models ; 2) use catchments to derive watershed characteristics commonly used within fluvial species distribution models (e.g., percent forest); 3) use watershed characteristics to develop species distribution models for multiple species; 4) predict the occurrence of all species throughout the study area; and 5) compare the sensitivity of watershed characteristics, model support and fit, and projected species distributions for each delineation procedure.

METHODS

STUDY AREA AND FIELD SAMPLING METHODS

The Etowah River system is a highly diverse aquatic network within the Coosa River system of Georgia, Tennessee, and Alabama. High levels of species imperilment have resulted mostly from extensive loss of habitat and continuity, restricting survival of formerly more widespread assemblages to fragmented headwater systems (Burkhead et al. 1997). In order to investigate the patterns of occurrence for several species, 40 sites within the headwaters of the Etowah River system were selected using a stratified random approach and sampled for riffle dwelling fish. In total 38 species were captured, representing 8 families and 20 genera. For a more detailed description of study area, sites, and field sampling methods see Anderson et al. (2012).

GEOGRAPHIC INFORMATION SYSTEMS METHODS

In order to investigate the effect of cell size on watershed delineation procedures, two elevation resolutions, 1/3 arc second (~10 meter) and 1 arc second (~30 meter) were downloaded from the National Elevation Dataset (NED;

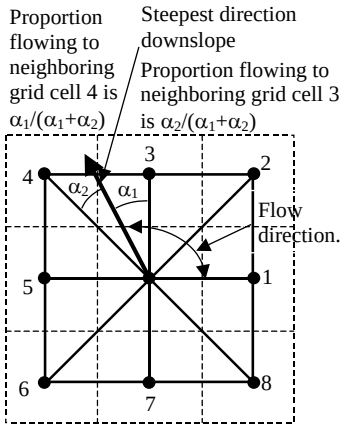


FIGURE 1. Method of determining flow direction in the D∞ flow direction model. Flow direction is determined by the steepest downward slope based on planar triangular facets on a centered grid (Tarboton 1997). Figure taken from the TauDEM Tools help in ArcGIS 10.x.

Gesch et al. 2002) for the study area. Multiple panels for each dataset were mosaiced together in ArcGIS 10.x and projected using cubic convolution resampling in the universal transverse mercator projection for Zone 16 North. Each elevation grid was registered to a common point to ensure grid alignment. In addition to NED data, the 2006 National Landcover Dataset (NLCD; Fry et al. 2011) was also downloaded, projected, and registered using the same procedure; however, nearest neighbor was used for the resampling technique.

For each site, eight different procedures were used to delineate the upstream catchment draining to the site. First, at the 10 meter resolution, upstream catchments were delineated using the single flow directional method (Jenson and Domingue 1988), also known as the D8 flow method, available in the Hydrologic Toolbox of the Spatial Analyst Extension in ArcGIS 10.x. Second, as an examination of the effect of surface area on the amount of upstream catchment area, upstream catchments were also delineated using a surface area grid as a weight raster in the flow accumulation process. Surface area was estimated using DEM Surface Tools (Jenness 2012), for the reason that it can provide a more realistic estimate of the cell area than the standard planimetric area. Third, upstream catchments were delineated using the multiple flow path procedure developed by Tarboton (1997) that is currently available in the ArcGIS extension TauDEM (Figure 1). Known as the D∞ flow method, this algorithm assigns proportions of flow from each grid cell among downslope cell neighbors in an effort to better approximate flow across surfaces (Tarbotom 1997). Similar to the single flow delineations, two catchments were created for each site: one using planimetric area and one using surface area. Finally, this process was repeated at the 30 meter resolution for a total of eight different estimates of watershed area for each site.

For each catchment, landuse was summarized using the 2006 NLCD (Fry et al. 2011). First, the NLCD land classes were reclassified into binary grids representing developed land (NLCD land classes 21, 22, 23, and 24), forested land (land classes 41, 42, and 43) and agricultural land (land classes 81 and 82; Table 1). To incorporate

TABLE 1. Watershed characteristics derived from each watershed delineation procedure.

Metric	Description
Watershed Area	Total area of landscape draining to the site (i.e., pour point)
Watershed Area^2	Polynomial term for Watershed Area
Percent Forested	Percent of watershed in land class 41, 42, or 43 according to the NLCD 2006
Percent Agriculture	Percent of watershed in land class 81 or 82 according to the NLCD 2006
Percent Developed	Percent of watershed in land class 21, 22, 23, or 24 according to the NCLD 2006

surface area, these grids were then multiplied by the surface area grid created from DEM surface tools. Summarizations were then made for both planimetric area and surface area by using the land class grid (binary for planimetric and integer for surface) as a weight grid within the flow accumulation process for the D8 flow method or the contributing area process for the D∞ method. The resulting grid was therefore an estimate of either the total number of cell belonging to the land class draining to each cell (planimetric) or the total amount of surface area of the land class draining to each cell. This grid was then divided by either the total number of cells draining to each cell or the total amount of surface area draining to each cell, depending on whether planimetric area or surface area was used, to derive an estimate of the proportion of upstream area developed, forested, and agricultural.

STATISTICAL ANALYSIS

Using a condensed set of 23 species collected across the 40 sites in the Upper Etowah River system, I modeled the probability of presence with a multi-species logistic regression model (Ovaskainen and Soininen 2011). In this process regression coefficients for each species are estimated as random effects from a common normal distribution with an estimated mean and variance. Prior to the analysis, I visually tested for spatial autocorrelation (SAC) among residuals from a global model with all covariates included using a semivariogram with Euclidian distances between pairs of sites. Upon inspection of the semivariogram, I detected evidence of a SAC pattern with a spatial lag between 1.5 and 8 km, depending on the species (Figure 2). To quantitatively test for SAC, I calculated Moran's I for each species with a neighborhood of sites equal to 6 km. This value was then compared to 1000 randomizations of the residuals to calculate an empirical p-value for the hypothesis of no spatial autocorrelation. To better account for any SAC, I used a spatial filtering approach with eigenfunction analysis from Principle Coordinates of Neighborhood Matrices (PCNM; Griffith and Peres-Neto 2006). In this approach spatial eigenvectors are created from the spectral decomposition of the following matrix:

$$\left[\mathbf{I} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\right]\mathbf{W}\left[\mathbf{I} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\right]$$

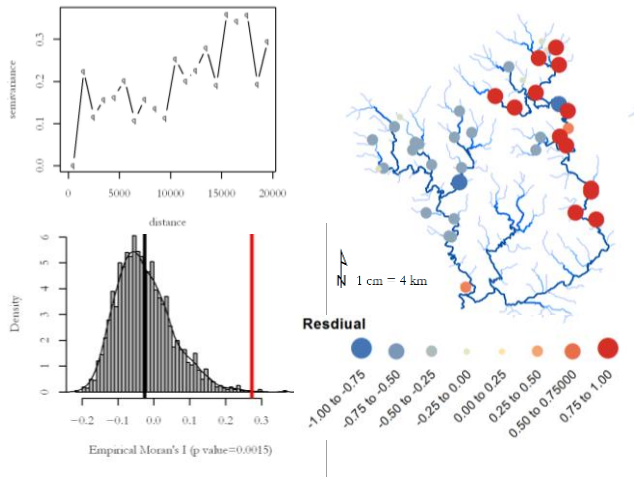


FIGURE 2. Example of spatial autocorrelation present for *Notropis lutipinnis* (yellowfin shiner). Top left panel demonstrates the spatial lag on semivariance between pairwise sites. Bottom left panel demonstrates value of Moran's I (red) compared to 1000 randomized samples (expected value is shown in black). Right panel demonstrates the spatial distribution of residuals for *Notropis lutipinnis* (yellowfin shiner) from the global model (i.e., model containing all covariates)

where $\mathbf{X}$ is a $n \times (p+1)$ matrix of the p predictor variables (e.g., watershed area) with an additional column of ones, $\mathbf{I}$ is an $n \times n$ identity matrix, and $\mathbf{W}$ is a binary connectivity matrix representing site neighbors. In this process site neighbors are defined as sites within a distance that maintains that all sampling localities are connected using a minimum spanning tree algorithm (Legendre and Legendre 1998). Because fish cannot disperse across the landscape in all directions and are limited to the flow network, I tested the amount of support of using a hydrologic distance to define $\mathbf{W}$ versus the more commonly used Euclidian distance. Pairwise hydrologic distances were derived in ArcGIS using a binary flow network derived from the D8 flow method. This binary flow network was used as a cost raster within the Cost Distance function in the Spatial Analyst Toolbox. For each distance matrix, I extracted the positive eigenvectors and performed a permutation test to determine if Moran's I was significant for the eigenvector (α=0.05). Using only significant eigenvectors, I fit a model for each distance matrix with no other covariates except for the significant positive eigenvectors to estimate the weight of evidence for the distance matrix using Akaike's Information Criterion (AIC; Akaike 1973; Burnham and Anderson 2002) as corrected for small sample sizes (AICc; Hurvich and Tasi 1989). Additionally, I evaluated the ability for each model to predict into withheld data using 5-fold cross-validation. Predictability of the model was evaluated using the threshold independent metric area under the curve (AUC) of the receiver operating characteristic (ROC) plot (Hanley and McNeil 1982). The model with the lowest AICc and highest AUC was then used for spatial filtering for all candidate models.

For each set of watershed characteristics derived from the different flow path algorithms (D8 and D∞), area estimation method (plannimetric and surface), and resolution (10m and 30m), 24 different models representing all possible additive combinations of the five covariates described in Table 1 were fitted, although the polynomial watershed area covariate was only included in the model if watershed area was there as well. Each model was built in

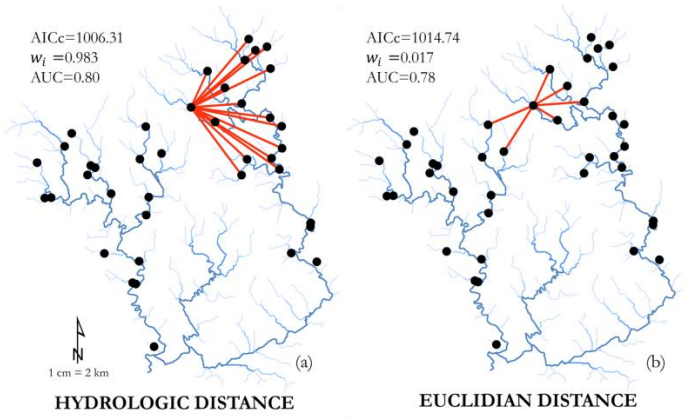


FIGURE 3. Spatial neighborhoods for sample site 4 used to account for spatial autocorrelation. (a) Hydrologic distance neighborhood was defined as sites within 31,651 stream km. (b) Euclidian distance neighborhood was defined as sites within 5,937 air km. Neighborhoods are shown with Akaike's Information Criterion (AICc) values, models weights (w_i), and the Area Under the Curve (AUC) based on the Receiver Operating Characteristic.

R using the package lme4 (Bates et al. 2011) which fits generalized linear mixed models using restricted maximum likelihood estimation. Models were then compared using AICc and verified using the threshold independent metric AUC based on the ROC.

The regression coefficients of the best supported model from each of the candidate sets were used to predict the distribution of each taxon in all stream segments across the study area in the upper Etowah River system. In order to estimate the total length of streams each species occupied, sites with a probability greater than the sample prevalence for each species were designated as occupied. Occupied stream segments were then created using raster functions and converted to a vector polyline to estimate the length geometry.

ESTIMATION OF SENSITIVITY TO FLOW DIRECTION PROCEDURE

Sensitivity to the type of flow direction procedure was examined at three levels. At the first level, I examined the changes in summary statistics of the watershed characteristics listed in Table 1. At the second level, I examined changes in model ranking and estimated regression coefficients of each characteristic for all species considered. Finally, at the third level, I examined the change in estimated length of streams occupied by each species.

RESULTS

WATERSHED CHARACTERISTICS

Changes in estimated watershed characteristics were variable (Tables 2-3). For watershed area there was surprisingly little change in estimated areas between different resolutions using the same flow path procedure (i.e., single or multiple) and area estimation method (i.e., planimetric or surface). This change was slightly higher for pairwise comparisons of resolutions using surface area (19.1 ha for D8 and 24.7 ha for D∞) than planimetric area (2 ha for D8 and 3.6 ha for D∞). Likewise, relatively small changes (~20-40 ha) were observed between single and multiple path methods of the same area estimation method; however, these differences increased dramatically when comparing area derived by planimetric versus surface area (148-214 ha). Although these discrepancies were equally

TABLE 2. Summary statistics of 10m cell watershed characteristics for catchments delineated with the single flow path and multiple flow path algorithm using planimetric and surface area.

	Single Flow Path (D8)				Multiple Flow Path (D∞)			
	Planimetric Area		Surface Area		Planimetric Area		Surface Area	
Variable	**Mean (SD)**	**Min-Max**	**Mean (SD)**	**Min-Max**	**Mean (SD)**	**Min-Max**	**Mean (SD)**	**Min-Max**
Area (ha)	3972.5(6670.5)	155.5 - 22348.8	4152.8(6958.3)	160.9 - 23190.9	4004.8(6733.4)	159.7 - 23092.7	4185.7(7022)	165.1 - 23949.7
Forested Area (ha)	3562.5(5934.7)	89.3 - 19295.5	3736.3(6214)	90.3 - 20209.9	3588(5981.6)	90.5 - 19609.2	3762.4(6261.7)	91.5 - 20416.6
Developed Area (ha)	150.6(275.4)	4.5 - 1124.7	153.6(280.6)	4.5 - 1146.3	153.6(282)	4.5 - 1189.4	156.6(287.4)	4.6 - 1211.9
Agricultural Area (ha)	128(302.5)	0 - 1314	129.1(305)	0 - 1325.2	129.5(306.1)	0 - 1346.6	130.6(308.7)	0 - 1358.1
Percent Forested	88.6(10.4)	53.6 - 98.7	88.7(10.4)	53.7 - 98.7	88.5(10.4)	53.4 - 98.7	88.7(10.4)	53.4 - 98.7
Percent Developed	4.5(3.4)	1.3 - 16.3	4.4(3.4)	1.3 - 16.3	4.6(3.4)	1.3 - 16.2	4.5(3.4)	1.3 - 16.2
Percent Agriculture	2.8(4.6)	0 - 24.6	2.8(4.6)	0 - 24.6	2.8(4.7)	0 - 24.9	2.8(4.6)	0 - 24.9

TABLE 3. Summary statistics of 30m cell watershed characteristics for catchments delineated with the single flow path and multiple flow path algorithm using planimetric and surface area.

	Single Flow Path (D8)				Multiple Flow Path (D∞)			
	Planimetric Area		Surface Area		Planimetric Area		Surface Area	
Variable	**Mean (SD)**	**Min-Max**	**Mean (SD)**	**Min-Max**	**Mean (SD)**	**Min-Max**	**Mean (SD)**	**Min-Max**
Area (ha)	3974.5(6670.5)	157.5 - 22350.3	4133.7(6923.6)	161.6 - 23089.6	4001.2(6728.2)	160.1 - 23076.1	4161(6982.1)	164.8 - 23828.3
Forested Area (ha)	3564.4(5934.7)	85.6 - 19297.4	3718(6180.6)	86.3 - 20100.5	3585.6(5977.7)	87.1 - 19598.6	3739.8(6224.3)	87.9 - 20307.5
Developed Area (ha)	150.7(275.2)	4.5 - 1125.1	153.3(279.8)	4.5 - 1144.2	152.9(281.3)	4.5 - 1186.2	155.5(285.9)	4.5 - 1206.1
Agricultural Area (ha)	128(302.4)	0 - 1313.6	128.9(304.6)	0 - 1323.3	129.1(305.7)	0 - 1344.8	130.1(307.9)	0 - 1354.8
Percent Forested	88.6(10.4)	53.4 - 98.6	88.7(10.4)	53.4 - 98.6	88.5(10.4)	53 - 98.6	88.7(10.4)	53 - 98.6
Percent Developed	4.5(3.3)	1.3 - 15.8	4.4(3.3)	1.3 - 15.8	4.5(3.4)	1.3 - 16.2	4.5(3.4)	1.3 - 16.2
Percent Agriculture	2.8(4.7)	0 - 25.2	2.8(4.7)	0 - 25.2	2.8(4.7)	0 - 25.3	2.8(4.7)	0 - 25.3

TABLE 4. Log likelihood (LN $\mathcal{L}$), number of parameters (K), AICc, delta AICc (Δ_i), model weight (w_i), evidence ratio (w_1/w_i), and area under the curve (AUC) based on within-sample prediction for the best supported model for each cell size, flow path algorithm, and area approximation method.

Cell Size	Flow Path	Area	LN $\mathcal{L}$	K	AICc	Δ_i	w_i	Evidence Ratio	AUC
10	D∞	Planimetric	-415.62	15	861.78	0	0.21897845	1	0.929
10	D∞	Surface	-415.77	15	862.07	0.29	0.18942124	1.15603957	0.928
30	D∞	Planimetric	-415.87	15	862.26	0.48	0.17225455	1.27124915	0.928
30	D∞	Surface	-416.01	15	862.54	0.76	0.14975091	1.46228459	0.928
10	D8	Planimetric	-416.65	15	863.82	2.04	0.07896252	2.77319476	0.928
10	D8	Surface	-416.8	15	864.13	2.35	0.0676247	3.23814294	0.927
30	D8	Planimetric	-416.83	15	864.18	2.4	0.06595504	3.32011692	0.927
30	D8	Surface	-416.97	15	864.47	2.69	0.05705258	3.83818654	0.927

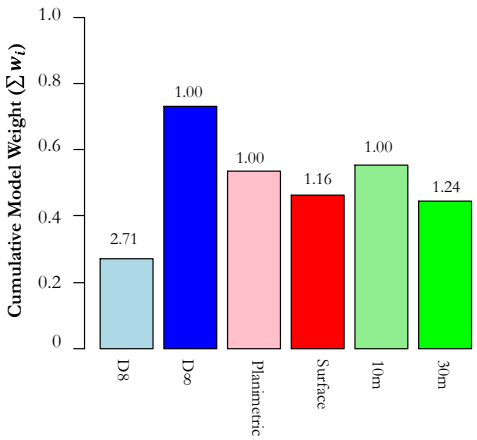
large for estimated areas of land-use classes (Tables 2-3), the estimated percentages of each land-class differed substantially less. This suggests that while estimates differ between planimetric and surface areas for land-use, they differed proportionally to the total area.

MULTI-SPECIES DISTRIBUTIONAL MODELS

Tests for spatial autocorrelation were significant for several species within the dataset (Figure 2). Of the two neighborhood matrices (i.e., Euclidian and Hydrologic) used to account for spatial autocorrelation (Figure 3), the hydrologic neighborhood had the greatest support, carrying 98.27% of the weight. Additionally, this neighborhood had better out-of-sample prediction, with an AUC of 0.80, a sensitivity of 0.49 and a specificity of 0.87 (Table 4). Using the Hydrologic network, three positive eigenvectors from the Principle Coordinate of Neighborhood Matrices analysis were found to be significant (α=.05) based on permutation tests using Moran's I. These three eigenvectors were recalculated for each candidate set of models to reduce multicollinearity between the predictor variables and the spatial eigenvectors. The eigenvectors for each candidate set of covariates were subsequently included in all models within the analysis to deal with the present spatial autocorrelation.

Within each candidate set of models there was great support for a single top model (w_i>0.59); however, that support differed depending on which flow path algorithm, area approximation method, or resolution was used. The uncertainty of selecting a top model was greatest for the 10m single flow path planimetric area candidate set, with one model holding 59.5% of the total model weight, and was smallest for the 30m multiple flow path surface area candidate set, with one model holding 64.2% of the total model weight. No matter which candidate set of covariates was used, the best-supported model included the same covariates: watershed area, watershed area^2, and percent forested.

FIGURE 4. Cumulative model weights and respective evidence ratios for each watershed calculation method. Evidence ratios were derived as the quotient of the method with the highest support divided by the method with the lowest support. Cumulative model weights and evidence ratios were only calculated using the methods of different flow paths (shown within different shades of blue), area estimation procedures (red), or cell size resolutions (green). For example, the evidence ratio for the D8 Flow path algorithm is 2.71 times less likely than the D∞ flow path algorithm.



In comparison between the eight best-supported models from the eight different candidate sets (Table 4), the 10m multiple flow path planimetric area model held the greatest amount of support (w_i=0.22; Table 4); however, any other candidate sets derived using the D∞ flow method were nearly equally as likely (Evidence Ratios < 1.5). The cumulative model weight ($\sum w_i$) for the models derived using the D∞ flow method was 0.73 (Figure 4). In contrast, models derived using the D8 flow method were greater than 2.7 times less likely to be the best supported model and carried a cumulative model weight of 0.27. Differences in support were far less when comparing models with different area approximation methods (surface area models were 1.16 times less likely) or difference cell size resolutions (30m models were 1.24 times less likely).

The comparisons of the estimated distributions of the regression coefficients demonstrated only minor changes in distributional means and variances with the multiple flow direction method having slightly higher regression coefficients (Table 5). Estimated variances appeared to be dependent on both the flow direction method and the resolution used to delineate catchments.

Although the regression coefficients did not change considerably among the best supported models, the projected

TABLE 5. Estimated means, standard errors (SE), and variances for species regression coefficients from the best-supported models for each candidate set. Best-supported models are shown with the AICc model weights (w_i) from Table 4. Regression coefficients were estimated as random effects in a generalized linear mixed model.

				Watershed Area		Watershed Area^2		Percent Forest	
Cell Size	Flow Path	Area	w_i	Mean (SE)	Variance	Mean (SE)	Variance	Mean (SE)	Variance
10	D∞	Planimetric	0.219	2.66 (0.65)	6.62	-0.85 (0.22)	0.51	-0.42 (0.10)	0.02
10	D∞	Surface	0.1894	2.66 (0.65)	6.60	-0.85 (0.22)	0.52	-0.43 (0.10)	0.01
30	D∞	Planimetric	0.172	2.66 (0.65)	6.69	-0.85 (0.22)	0.52	-0.42 (0.10)	0.01
30	D∞	Surface	0.150	2.65 (0.65)	6.65	-0.85 (0.22)	0.53	-0.42 (0.10)	0.01
10	D8	Planimetric	0.079	2.57 (0.64)	6.52	-0.81 (0.22)	0.51	-0.41 (0.10)	0.02
10	D8	Surface	0.068	2.55 (0.52)	6.50	-0.81 (0.22)	0.52	-0.41 (0.10)	0.01
30	D8	Planimetric	0.066	2.57 (0.65)	6.57	-0.81 (0.22)	0.52	-0.40 (0.10)	0.01
30	D8	Surface	0.057	2.55 (0.64)	6.53	-0.80 (0.22)	0.53	-0.41 (0.10)	0.01

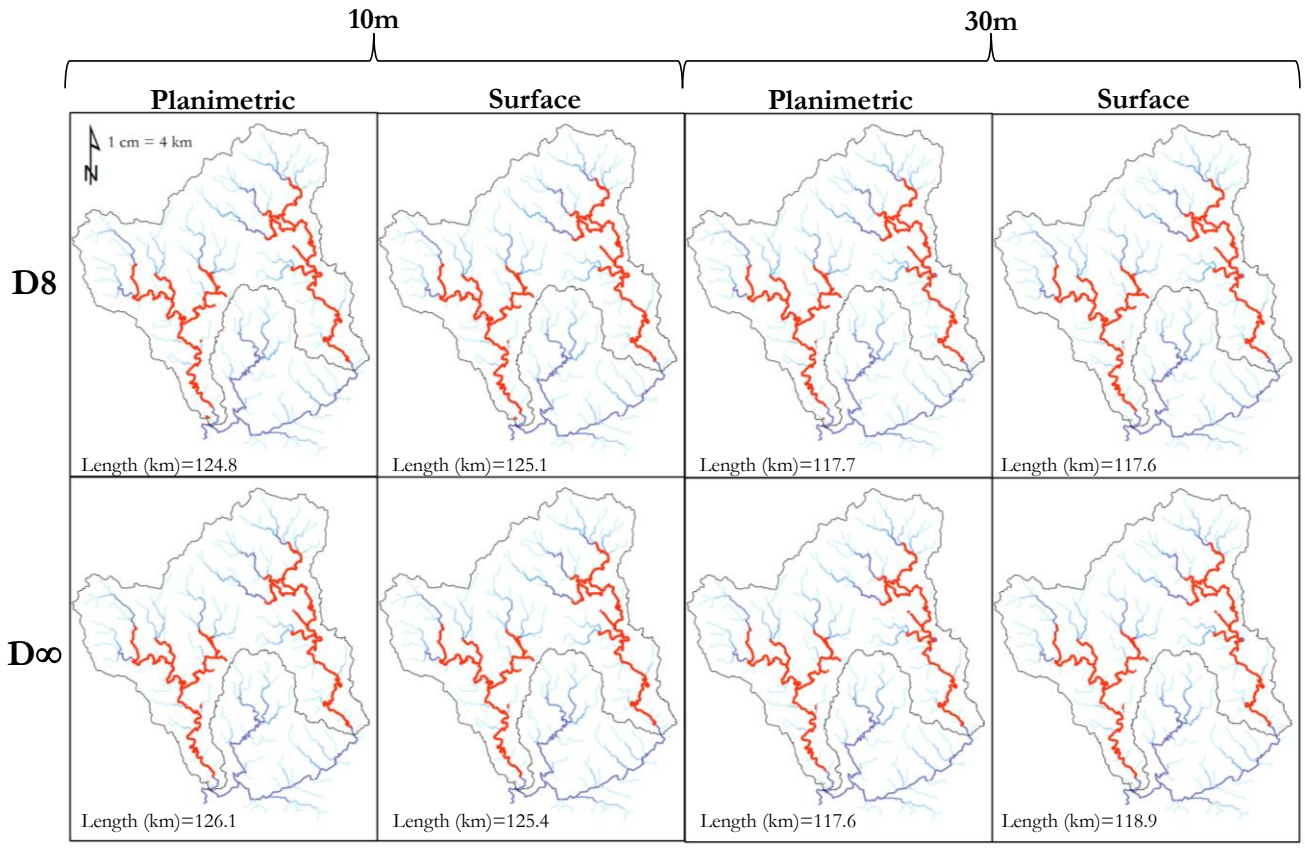


FIGURE 5. Projected streams occupied (red) for *Etheostoma brevirostrum* (holiday darter) using different flow path algorithms, surface area approximations, and cell resolutions. Streams (blue) shown were derived with a flow network from the D8 algorithm with a 10 meter elevation grid.

stream lengths occupied by some species within the dataset differed tremendously depending on which flow direction procedure, area approximation method, and cell resolution was used. Generally, only small differences were observed between the different methods when the same cell resolution was used ($\bar{x}$=2.08 and 2.15 km for 10 and 30m resolutions, respectively; Figure 5). However, for some headwater species, differences in estimated stream lengths differed substantially, even when the same resolution was used (max=6.77 km; Figure 6 and Appendix A). Finally, when estimated stream lengths were compared across resolutions, large differences were observed ($\bar{x}$=8.47 km), indicating that the resolution used can drastically change the estimate derived.

DISCUSSION

The use of remotely derived watershed characteristics within fluvial species distribution models has become a common practice. However, the assumptions and uncertainties of the spatially derived datasets are not often explored. In this study I investigated the effect of using single (D8) versus multiple flow direction methods (D∞),

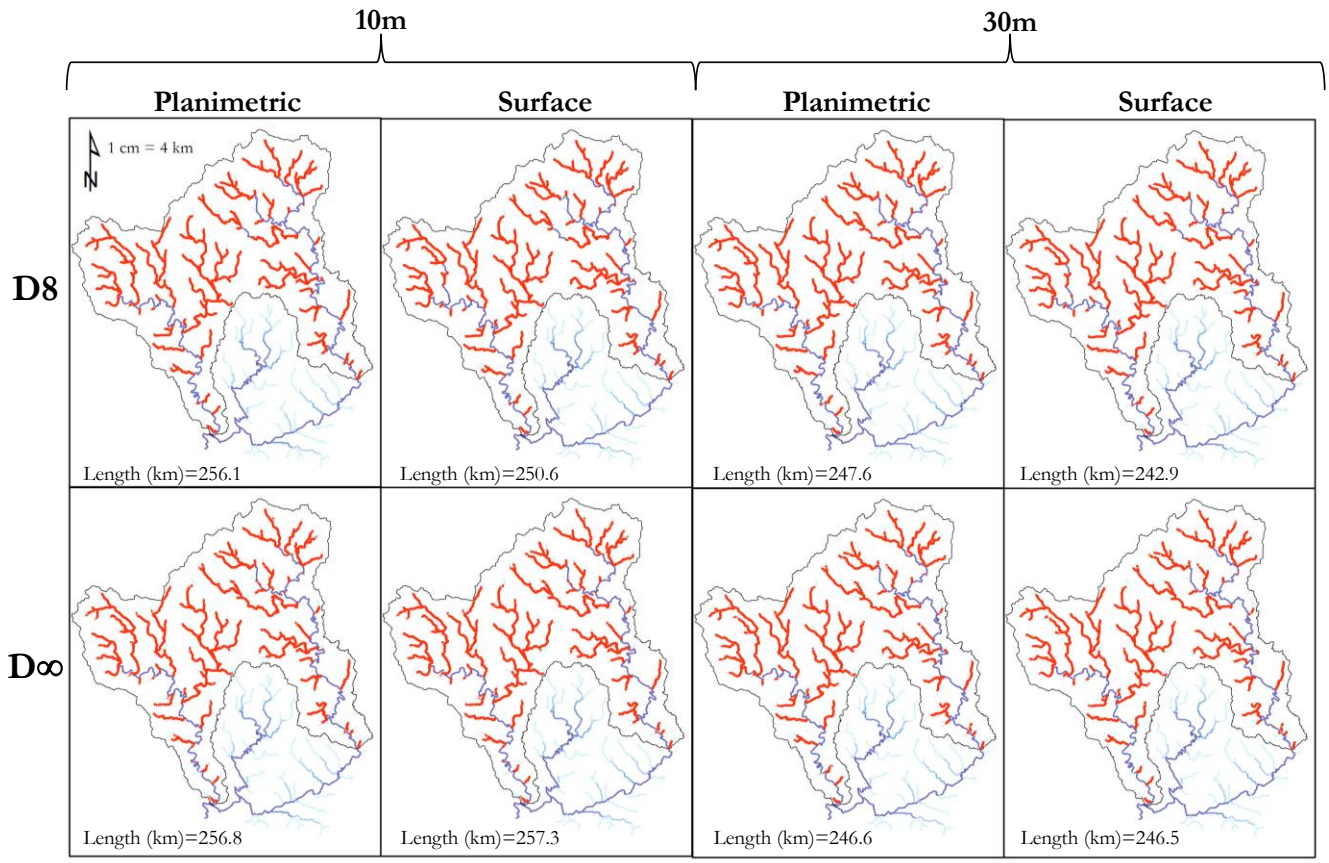


FIGURE 6. Projected streams occupied (red) for *Semotilus atromaculatus* (creek chub) using different flow path algorithms, surface area approximations, and cell resolutions. Streams (blue) shown were derived with a flow network from the D8 algorithm with a 10 meter elevation grid.

planimetric versus surface area approximations, and 10 versus 30 meter resolution elevation models on the characterization of upstream catchments to be used in species distribution models. I found that the estimated catchment area and landuse areas can differ dramatically depending on the delineation procedure used, with the difference being the greatest between planimetric and surface area approximations. However, for this dataset, the increase in area of each of the land classes did not dramatically affect the estimated percentages of each land-class within each upstream catchment (i.e., areas were proportional). I also found that using a multiple-flow path algorithm (D∞) versus a single-flow path algorithm (D8) can increase the fit of a species distribution model, but for this dataset the improvement was quite small, especially when evaluated using prediction characteristics (i.e.., AUC). This small effect was also seen within the estimated regression coefficients, with the means and variances of the coefficient

distribution being nearly identical. However, as demonstrated in Figures 5 and 6, even small changes in these estimates can drastically alter the estimate of potential habitat for some species.

This analysis used a spatial eigenfunction approach for dealing with spatial autocorrelation present within the dataset. Similar to the findings of Landeiro et al. (2011), I found that the use of the hydrologic distance matrix within the PCNM analysis was better supported than Euclidian distances. This was substantiated by the better AUC estimate from the k-fold cross-validation approach as well. However, an important assumption of this analysis and the one performed by Landeiro et al. (2011) is that the neighborhood matrix is symmetric. In reality, the spatial dependence within stream networks might be much more complex. For example, in areas with high velocity or dramatic changes in elevation (e.g., waterfalls) dispersal might only occur in the downstream direction. Although it is feasible to incorporate an asymmetric matrix in the PCNM analysis, it is very likely that the asymmetry would vary spatially (e.g., greater in areas with higher slope). An interesting option for dealing with this heterogeneity would be the use of a mixture model. Thus, two separate models could be used in a joint likelihood to 1) identify neighbors (i.e., process model) and 2) incorporate the spatial eigenvectors into the species distribution model (i.e., state model).

The study area in this analysis consists of an area of moderate to high change in elevation. The upper-most portion of the catchment is within the Blue Ridge physiographic providence, while the remainder of the catchment lies within the Piedmont. In this type of landscape, uncertainty of flow direction is not a large issue because the steepness of the terrain makes flow paths more definable. The results from the analysis would likely be drastically different in areas of flatter topography such as the coastal plain. Conversely, in flatter topography the effect of using planimetric area versus surface area would likely decrease because the surface to area ratio would converge to 1 (i.e., planimetric area=surface area). The finding in this study that using different area approximation methods can alter the estimate of watershed area and land-use area is very important. In the case of this analysis, the changes in land-use were proportional to the change in over-all catchment area. However, in many regions this may not be the case.

CONCLUSION

This study demonstrates the effect of using different watershed delineation procedures to estimate the resource selection, habitat uses, and distributional characteristics of multiple species. Although the analysis in this study only found substantive evidence that using watershed delineation procedures that account for the uncertainty of flow direction can improve estimates of species distribution models, in other geographic areas (e.g., flatter topography) this

difference might be much larger. There appears to be little consequence, except a limited increased in flow dispersion, to using the D∞ method over the widely used D8 method. Both approaches can incorporate weight rasters and land-class summarizations can be made using both methods. Additionally, although I did not find a large amount of support favoring the use of surface area over planimetric area in this SDM approach, surface area is a much more realistic approximation of the true area of the landscape. Therefore, it would be advisable to use both surface area and multiple-flow path methods in the automatic watershed delineation procedure for nearly any analysis.

REFERENCES

Akaike, H. 1978. A Bayesian analysis of the minimum AIC procedure. Annals of the Institute of Statistical Mathematics 30:9-14.

Anderson, G.B., M.C. Freeman, M.M. Hagler, and B.J. Freeman. 2012. Occupancy modeling and estimatin of the holiday darter species complex within the Etowah River system. Transactions of the American Fisheries Society 141: 34-45.

Bates, D., M. Maechler and B. Bolker. 2011. lme4: linear mixed-effects models using S4 classes. R package version 0.999375-42. http://CRAN.R-project.org/package=lme4

Burnham, K.P. and D.R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York.

Franklin, J. 2009. Mapping Species Distributions: Spatial Inference and Prediction. Cambridge University Press, Cambridge, MA.

Fry, J.A, G. Xian, S. Jin, J.A. Dewitz, C.G. Homer, L. Yang, C.A. Barnes, N.D. Herold, and J.D. Wickham. 2011. Completion of the 2006 National Land Cover Database for the Conterminous United States. Photogrammetric Engineering and Remote Sensing 77: 858-864.

Gesch, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, and D. Tyler. 2002. The national elevation dataset. Photogrammetric Engineering and Remote Sensing 68(1): 5-11.

Griffith, D.A., and P.R. Peres-Neto. 2006. Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. Ecology 87: 2603-2613.

Guisan, A. and W. Thuiller. 2005. Predicting species distributions: offering more than simple habitat models. Ecology Letters 8: 993-1009.

Hurvich, C.M., and C.L. Tsai. 1989. Regression and time series model selection in small samples. Biometrika 76: 297-307.

Jenness, J. 2012. DEM Surface Tools for ArcGIS (surface_area.exe). Jenness Enterprises. Available at: http://www.jennessent.com/arcgis/surface_area.htm.

Jenson, S.K. and J.O. Domingue. 1988. Extracting topographic structure from digital elevation data for Geographic Information System analysis. Photogrammetric Engineering and Remote Sensing 54: 1593-16000.

Landeiro, V.L., W.E. Magnusson, A.S. Melo, H.M.V. Espirito-Santo, and L.M. Bini. 2011. Spatial eigenfunction analysis in stream networks: do watercourse and overland distances produce different results? Freshwater Biology 56: 1184-1192.

Legendre, P., and L. Legendre. 1998. Numerical Ecology. Elsevier Science, Amsterdam, The Netherlands.

Newton-Cross, G., P.C.L. White, and S. Harris. 2007. Modeling the distribution of badgers *Meles meles*: comparing predictions from field-based and remotely derived habitat data. Mammal Review 37: 54-70.

Ovaskainen, O. and J. Soininen. 2011. Making more out of spare data: hierarchical modeling of species communities. Ecology 92: 289-295.

Phillips, S.J., R.P. Anderson, R.E. Schapire. Maximum entropy modeling of species geographic distributions. Ecological Modeling 190: 231-259.

Roy, A.H., B.J. Freeman, and M.C. Freeman. 2007. Riparian influences on stream fish assemblage structure in urbanizing streams. Landscape Ecology 22: 385-402.

Stockwell, D. and D. Peters. The GARP modeling system: problems and solutions to automated spatial prediction. International Journal of Geographical Information Science 13: 143-158.

Tarboton, D.G. 1997. A new method for the determination of flow directions and contributing areas in grid digital elevation models. Water Resources Research 33: 309-319.

Walters, D.M., D.S. Leigh, and A.B. Bearden. 2003. Urbanization, sedimentation, and the homogenization of fish assemblages in the Etowah River basin, USA. Hydrobiologia 494: 5-10.

Weih, R.C. Jr. 1991. Evaluating Methods for Characterizing Slope Conditions with Polygons. Ph.D. Dissertation. Virginia Polytechnic Institute and State University, Blacksburg, VA.

Wenger, S.J., J.T. Peterson, M.C. Freeman, B.J. Freeman, and D.D. Homans. 2008. Stream fish occurrence in response to impervious cover, historical land use, and hydrogeomorphic factors. Canadian Journal of Fisheries and Aquatic Sciences 65: 1250-1264.

APPENDIX A. Estimated kilometers of stream occupied using the best-supported model for each candidate set. 10m Range and 30m Range represent the maximum estimate minus the minimum estimate of kilometers for the species using 10m resolution data and 30m resolution data, respectively. Total Range represents the maximum estimate minus the minimum estimate of kilometers for all methods and resolutions.

	10m				30m						
	Single Flow Path (D8)		Multiple Flow Path (D∞)		Single Flow Path (D8)		Multiple Flow Path (D∞)				
Species	**Planimetric**	**Surface**	**Planimetric**	**Surface**	**Planimetric**	**Surface**	**Planimetric**	**Surface**	**10m Range**	**30m Range**	**Total Range**
Campostoma oligolepis	146.92	145.04	145.95	144.8	137.8	136.92	137.47	138.13	2.12	1.21	10
Cyprinella callistia	129.53	131.37	128.73	130.35	123.36	124.98	123.35	124.98	2.64	1.63	8.02
Cyprinella trichroistia	135.89	136.84	135.46	136.16	128.13	129.93	128.23	129.91	1.38	1.8	8.71
Nocomis leptocephalus	149.7	147.94	149.15	147.19	140.03	138.68	139.61	140.6	2.51	1.92	11.02
Notropis chrosomus	82.3	80.5	78.01	76.63	77.26	76.41	73.3	73.67	5.67	3.96	9
Notropis lutipinnis	107.45	107.4	106.29	106.69	102.69	102.62	101.15	101.51	1.16	1.54	6.3
Notrpis stilbius	87.92	87.92	88.27	88.29	83.06	83.06	83.43	83.43	0.37	0.37	5.23
Notropis xaenocephalus	121.88	122.72	122.03	123.03	116.29	116.62	116.66	116.81	1.15	0.52	6.74
Luxilus zonistus	97.26	97.22	98.02	98.04	91.39	91.07	92.58	92.71	0.82	1.64	6.97
Semotilus atromaculatus	256.14	250.56	256.82	257.33	247.58	242.93	246.6	246.53	6.77	4.65	14.4
Hypentelium etowanum	159.97	158.25	159.8	158.04	151.59	150.25	151.59	151.08	1.93	1.34	9.72
Oncorhynchus mykiss	105.24	105.5	104.75	105.45	100.59	101.05	100.74	99.88	0.75	1.17	5.62
Cottus carolinae	166.68	167.35	164.96	165.22	159.49	160.16	157.04	157.62	2.39	3.12	10.31
Lepois auritus	144.42	143.87	142.13	141.88	136.47	136.12	135.27	134.74	2.54	1.73	9.68
Lepois machrochirus	89.94	88.15	89.51	87.18	86.53	85.02	85.59	83.74	2.76	2.79	6.2
Micropterus coosae	99.2	97.91	98.69	97.92	92.45	92.2	92.18	92.61	1.29	0.43	7.02
Micropterus punctulatus	93.72	92.34	92.25	90.87	90.75	88.84	88.77	87.43	2.85	3.32	6.29
Micropterus salmodies	144.15	142.59	142.68	141.55	138.57	137.41	135.95	133.96	2.6	4.61	10.19
Etheostoma brevirostrum	124.84	125.09	126.05	125.43	117.73	117.62	117.6	118.99	1.21	1.39	8.45
Etheostoma etowahae	101.59	102.4	101.77	102.51	96.15	96.83	96.64	90.29	0.92	6.54	12.22
Etheostoma stigmaeum	92.02	92.27	91.75	91.96	86.58	86.63	85.66	86.11	0.52	0.97	6.61
Percina nigrofasciata	106.16	104.56	105.71	104.17	98.6	98.01	98.79	99.03	1.99	1.02	8.15
Percina palmaris	135.3	136.36	135.55	136.84	128.92	130.16	129.49	130.72	1.54	1.8	7.92