2018

ASSESSING THE RELATIVE INFLUENCES OF ABIOTIC AND BIOTIC FACTORS ON A SPECIES' DISTRIBUTION USING PSEUDO-ABSENCE AND FUNCTIONAL TRAIT DATA: A CASE STUDY WITH THE AMERICAN EEL (Anguilla rostrata)

Taylor E. Woods

ASSESSING THE RELATIVE INFLUENCES OF ABIOTIC AND BIOTIC FACTORS ON A

SPECIES' DISTRIBUTION USING PSEUDO-ABSENCE AND FUNCTIONAL TRAIT

DATA: A CASE STUDY WITH THE AMERICAN EEL (*Anguilla rostrata*)

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

at Virginia Commonwealth University

by

Taylor Woods

Master of Science, Virginia Commonwealth University 2018

Bachelor of Science, Tulane University 2011

Director: Daniel J. McGarvey, Ph.D.

Associate Professor

Center for Environmental Studies

Virginia Commonwealth University
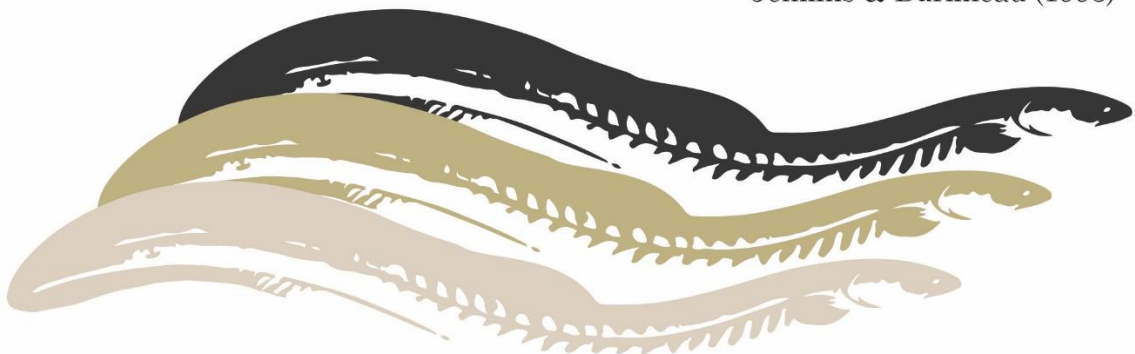
Richmond, Virginia

February 2018
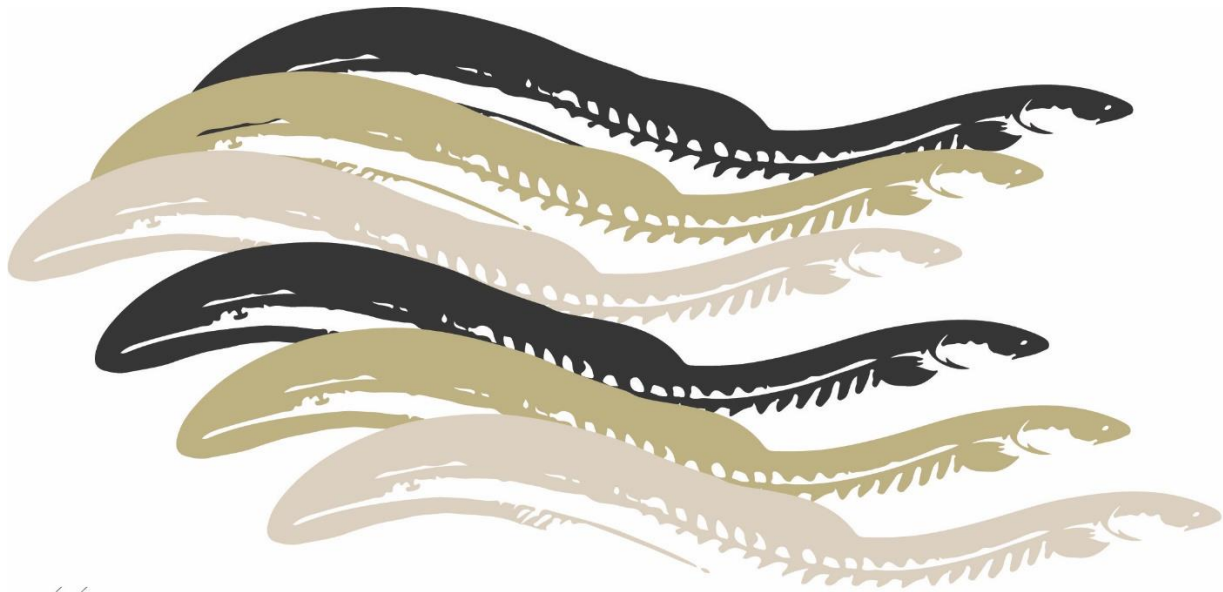
" *Anguilla rostrata* is an awesome eel.....

As its story is long, so too are the coastward drift of the frail larvae, and the trek made by adults to the oceanic spawning ground. "
-- Jenkins & Burkhead (1993)

Name. --Anguill-, from *Anguilla*, a Latin word for eel.
--*Rostrata* means "snout."

Table of Contents

List of Tables and Figures

Abstract

ASSESSING THE RELATIVE INFLUENCES OF ABIOTIC AND BIOTIC FACTORS ON A

SPECIES' DISTRIBUTION USING PSEUDO-ABSENCE AND FUNCTIONAL TRAIT

DATA: A CASE STUDY WITH THE AMERICAN EEL (*Anguilla rostrata*)

By Taylor Woods, Bachelor of Science

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

at Virginia Commonwealth University.

Virginia Commonwealth University, 2018

Major Director: Daniel J. McGarvey, Ph.D., Center for Environmental Studies

Species' distributions are influenced by abiotic and biotic factors but direct comparison of their relative importance is difficult, particularly when working with complex, multi-species datasets. Here, we present a flexible method to compare abiotic and biotic influences at common scales. First, data representing abiotic and biotic factors are collected using a combination of geographic information system, remotely sensed, and species' functional trait data. Next, the relative influences of each predictor variable on the occurrence of a focal species are compared. Specifically, 'sample' data from sites of known occurrence are compared with 'background' data (i.e. pseudo-absence data collected at sites where occurrence is unknown, combined with sample data). Predictor variables that may have the strongest influence on the focal species are identified as those where sample data are clearly distinct from the corresponding background distribution. To demonstrate the method, effects of hydrology, physical habitat, and co-occurring fish

functional traits are assessed relative to the contemporary (1950 – 1990) distribution of the American Eel (*Anguilla rostrata*) in six Mid-Atlantic (USA) rivers. We find that Eel distribution has likely been influenced by the functional characteristics of co-occurring fishes and by local dam density, but not by other physical habitat or hydrologic factors.

Introduction

        Characterizing the effects of biotic and abiotic influences on species' distributions and community structure is a long-standing challenge in ecology (Hutchinson 1957, Chase and Leibold 2003). To make this endeavor more tractable, abiotic and biotic effects have often been conceptualized as distinct levels in a series of nested, hierarchal filters that sort regional species pools into local communities (Poff 1997, Jackson et al. 2001). Notably, abiotic effects tend to comprise the higher, large-scale levels of the hierarchal filter model, while biotic influences comprise the lower, local-scale levels. In this way, species from a regional pool must navigate a series of relatively coarse-scale filters, such as physiological thermal constraints, before they can potentially take up residence at a particular locality. Then, to become a successful colonist, a species must navigate the final local-scale filter: coexistence within the previously established community.

        Perhaps due to the inherent complexity of characterizing biotic interactions within natural communities, evidence for the role of abiotic filtering in nature has accumulated more quickly than for biotic filtering, particularly at landscape or regional scales (Lawton 1999, Morales-Castilla et al. 2015). But with global change and other large-scale perturbations rearranging species' distributions and creating novel communities, ecologists are now earnestly working to enhance understanding of biotic influences (Parmesan and Yohe 2003, Gallardo and Aldridge 2013). For instance, improved methods to incorporate biotic interactions in species distribution models are an active area of inquiry (Guisan and Thuiller 2005, Wisz et al. 2013). Yet despite this progress, two key developments remain that would greatly benefit basic and applied research on biotic filtering: (1) improved methods to efficiently and systematically characterize biotic

influences within complex multi-species communities; and (2) a framework to quantify and directly compare the relative importance of abiotic and biotic factors.

Methods to characterize biotic interactions are often based upon records of pairwise species' co-occurrences (Connor and Simberloff 1979, Gotelli and McCabe 2002). For example, species distribution models seeking to incorporate biotic influences on the distribution of a focal species have made use of heterospecific occurrence records, abundances, and co-occurrence-based indices as biotic predictor variables (e.g. Leathwick and Austin 2001, Meier et al. 2011). Unfortunately, inferences based on co-occurrence data may be compromised by the fact that multiple processes can sometimes lead to the same co-occurrence patterns (Peres-Neto 2004, Cazelles et al. 2016). Direct observational evidence of pairwise interactions (positive, negative, or otherwise) can overcome these limitations, but such empirical data are difficult to obtain and only available for a relatively small number of species pairs (Bascompte and Jordano 2007, Connor et al. 2013). Furthermore, biotic influences may manifest as direct or indirect associations between many species that are not well-described by pairwise scenarios (Wooton 1994). Methods to efficiently characterize biotic interactions within complex, multi-species datasets (the 'biotic milieu' of McGill et al. 2006) are therefore needed (Tylianakis et al. 2008, Gallien et al. 2017).

The second development – a framework to assess the relative importance of abiotic and biotic factors – is necessary because differing scales are often used to characterize abiotic and biotic influences (Wiens 2011, Staniczenko et al. 2017), which makes direct comparison difficult. Abiotic factors are commonly represented by coarse-scale environmental surveys or standardized data from a geographic information system, while biotic interactions are evaluated at finer scales, often through experimental studies (Jackson et al. 2001, Ovaskainen et al. 2017).

Also hindering direct comparisons are the assumptions and logistical hurdles that must be navigated when scaling up from the fine-scale biotic data (e.g. point observations) to the larger scales at which abiotic data are most often available (Araújo and Luoto 2007, Funk et al. 2016). As noted above, a hierarchal filtering framework that *a priori* assumes abiotic habitat filtering is dominant at coarser scales while biotic influences are paramount at finer scales is often used to integrate abiotic and biotic data in a single, regional analysis (Pearson and Dawson 2003, Boulangeat et al. 2012). But large-scale species distribution and co-occurrence patterns may arise from habitat filtering, biotic interactions, or a combination of both processes (HilleRisLambers et al. 2012, Cadotte and Tucker 2017). Thus, new tools are needed to facilitate direct comparisons of abiotic and biotic influences at common scales.

In this study, we demonstrate a method to quantify then compare abiotic and biotic influences on the distribution of a focal species. The method uses a combination of physical habitat and functional trait data to first quantify abiotic and biotic conditions across a large landscape, then uses species' presence and 'pseudo-absence' samples (see next paragraph) to assess the relative importance of abiotic and biotic influences on the focal species' known occurrences. Functional trait analyses emphasize species' physiological, morphological, and behavioral characteristics, rather than their taxonomic identities (McGill et al. 2006, Frimpong and Angermeier 2010). They are ideal for assessing biotic influences in multi-species datasets because traits often regulate biotic interactions (e.g. resource competition among species with similar feeding behaviors; MacArthur and Levins 1967). Functional trait data can also be aggregated into assemblage- or community-level summaries of biotic influences (Carmona et al. 2016), then contrasted with abiotic variables.

Pseudo-absence samples are a solution to the 'problem' of presence-only data (Pearce and Boyce 2006). When presence-absence data are available, a traditional discrimination method such as logistic regression can be used to assess the effect of a given variable on a focal species' distribution (Guisan and Zimmermann 2000). However, presence-only data require alternative approaches, one of which is to use pseudo-absence samples to characterize the entire range of conditions that are potentially available to a focal species, then compare this range with conditions at sites of known occurrence (Hirzel et al. 2002). This is the basic method applied in Maximum Entropy species distribution modeling (Phillips et al. 2006) and, more generally, point process generalized linear models (Renner et al. 2015): a 'sample' distribution representing conditions at verified presence locations is compared with a random 'background' sample, representing conditions throughout the potential range (Elith et al. 2011). By comparing sample and background distributions, it is possible to determine whether the focal species is responsive to a given predictor variable. If so, it will occur in only a subset of the available conditions, as indicated by sample and background distributions that exhibit modest or minimal overlap (see Fig. 1 of Merow et al. 2013).

Here, we apply the logic of pseudo-absence sampling by using the degree of sample-background overlap (SBO) to assess, at a common spatial scale, the relative effects of abiotic and biotic variables on a focal species' distribution. Notably, while species' presence and pseudo-absence samples have most often been compiled for abiotic variables (for use in species distribution models), the basic SBO approach is applicable to any variable that can be 'mapped' to the area of interest (Hirzel et al. 2002). However, our goal is not to model the distribution of a focal species *per se*. Rather, by comparing SBO for suites of abiotic and biotic variables, we seek

to systematically infer whether the focal species' distribution is more closely associated with abiotic or biotic influences.

As an empirical context, we focus on the contemporary (records collected between 1950 – 1990) distribution of the American Eel (*Anguilla rostrata*) in six Mid-Atlantic (USA) rivers. The American Eel has a unique life history that makes it an ideal study organism for a comparison of abiotic and biotic influences. Eels are catadromous (ocean spawning) fishes with tremendous migratory capabilities and they are not believed to undergo a juvenile imprinting process (Gagnaire et al. 2012). Thus, the range of freshwater habitats that an individual Eel can potentially colonize is very large. Furthermore, Eels are semelparous and panmictic; adults spawn only once and do so within large aggregations of conspecifics. This 'all eggs in one basket' strategy minimizes the probability that selective forces will drive diversification in local habitat use among distinct populations or lineages (Oliveira 1999). Together, these life history characteristics suggest that freshwater habitat selection by upstream migrating elvers (juvenile Eels) is a 'real-time' function of the Eel's perception of local habitat conditions or assemblage structure (of resident fishes), rather than an imprinted or inherited response.

Using a combination of contemporary Eel occurrence records with data on local hydrology and physical habitat (abiotic influences), as well as functional trait data for local fish assemblages (biotic influences), our objectives are: (1) to use a combination of species' occurrence and functional trait data to quantify potential biotic influences within regional stream networks; and (2) assess the relative importance of abiotic and biotic influences on the regional distribution of the American Eel.

Material and methods

Study area and spatial framework

      This study includes six of the major Mid-Atlantic (USA) river basins within the native range of the American Eel (Jenkins and Burkhead 1993): the Potomac, Rappahannock, York, James, Chowan, and Roanoke rivers (Fig. 1). To facilitate direct comparisons among variables, all hydrologic, physical habitat, and fish data were aggregated within 'subwatersheds' or 12-digit Hydrologic Units from the US Geological Survey (2012) Watershed Boundary Dataset. Subwatersheds are currently the smallest spatial units included in the Watershed Boundary Dataset; a total of 1,407 subwatersheds exist within the six major river basins, with a mean surface area of 89.3 km^2 (SD = 40.5 km^2). They were chosen for this study because they maximized our ability to match Eel occurrence records with independent data on hydrology and habitat data at local spatial scales.
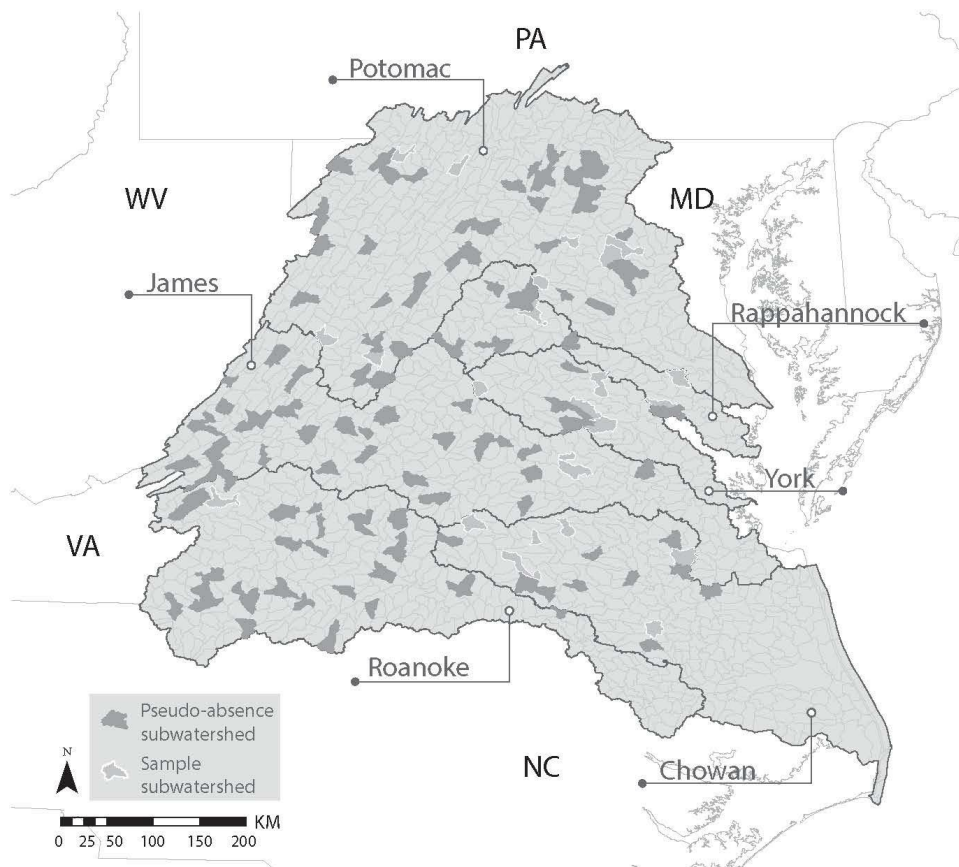
Figure 1. Map of the six Mid-Atlantic, USA drainages that defined our study region (i.e. total landscape for this study). The extent of each drainage is bounded by dark grey lines and the smaller subdivisions within each drainage show all potential subwatersheds. Also shown are the locations of sample (representing conditions at sites where the American Eel is known to occur) and pseudo-absence (representing a random sample of landscape conditions) subwatersheds that were determined to provide adequate data and were therefore included in the analysis (see text).

Hydrologic data and subwatershed selection

Selection of subwatersheds for use in our analyses began with a survey of available hydrology data, as these data were the most limiting of the three data classes that we incorporated (hydrologic, physical habitat, and biotic). Hydrologic metrics were calculated from daily discharge records, downloaded for all stream gauges located within the six study basins via the US Geological Survey's Water Data for the Nation website (<https://waterdata.usgs.gov/nwis>). Within each subwatershed that contained at least one

stream gauge, we used three criteria to determine which gauge data would be used to represent

local hydrology (Fig. 2). First, we queried gauges with a minimum of 20 years of nearly

continuous (periodic instances of several days or weeks of missing data were acceptable)

discharge records between the 1955 – 1985 water years (i.e. 1 October – 30 September; $n = 158$

gauges with sufficient records). This ensured that our hydrologic metrics would be broadly

representative of contemporary flow conditions (Gan et al. 1991) and temporally consistent with

the majority of fish occurrence records (collected between 1950 – 1990 in our study basins; see

Huang and Frimpong 2016 and 'Fish and biotic/functional trait data' section below). Second, we

identified subwatersheds with more than one internal gauge ($n = 5$), determined which gauge was

closest to the downstream terminus or 'pour point' of each subwatershed, then removed gauges

that were further upstream. This process narrowed the pool of suitable Hydrologic Units with

representative flow data to 153 subwatersheds.
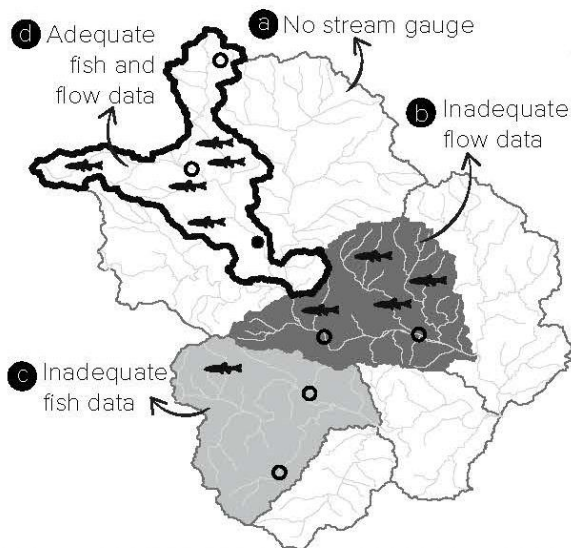
< Figure 2. Map illustrating the process to select subwatersheds used in this study. Four hypothetical scenarios are shown for subwatersheds that were: (a) not considered due to absence of an internal stream gauge (white, thin grey outline); (b) eliminated from the analysis due to inadequate flow data (e.g. internal gauge supplied flow data outside of water years 1960 – 1980; see text); (c) eliminated from further analysis due to inadequate fish sampling effort (≤ 10th percentile of sample densities; see text); and (d) retained for further analysis, with adequate fish and flow data (white with heavy black outline).

Hydrology in each of the remaining subwatersheds was then characterized with a subset of the Indicators of Hydrologic Alteration (IHA) metrics (Richter et al. 1996). Sixteen IHA metrics were selected to represent the five primary flow regime components: magnitude, frequency, duration, timing, and rate of change (Olden and Poff 2003). IHA metrics were calculated with the 'IHA' package (Law 2013) in R (R Core Team 2017), then appended to a subwatershed $\times$ IHA metrics matrix. From this matrix, we calculated Pearson's correlation coefficients and removed correlated metrics ($|r| \geq 0.6$), taking care to retain at least one variable from each flow regime component (see Table 1).

Table 1. Hydrologic, physical habitat, and functional trait variables used in the analysis following elimination of collinear variables. Each of the 27 predictor variables is described including units and statistics used for summaries at the subwatershed (CV = coefficient of variation, MD = median, MN = mean, %TRT = trait modality proportion). For hydrologic variables, the flow regime component represented by the predictor variable is shown in parentheses.

Variable Code	Description	SumStat
Hydrology		
Apr	April stream flow (ft^3/s) (*Magnitude*)	MD, CV
Bsflw	Baseflow index (ft^3/s) (*Magnitude*)	CV
DtMn	Julian date of annual minimum flow (*Timing*)	MD
HFlwCnt	Count of high flow periods above the 75th percentile (*Frequency*)	CV
HFlwDur	Duration of high flow periods (*Duration*)	CV
Rev	Number of reversals from rising to falling rates (*Rate of change*)	CV
Physical habitat		
Area	Catchment surface area (km^2)	MN
DmDns	Density of dams per catchment area (dams/ km^2)	MN

DmStr	Volume of dams per catchment surface area (m^3/km^2)	MN
Elev	Catchment elevation (m)	MN, CV
Rnff	Catchment runoff (mm)	MN, CV
Functional traits		
Fec	Fecundity (# eggs) per female per spawning season	MN
IncbTm	Incubation time from embryo fertilization to hatch (hours)	MN
SeasLen	Length of spawning season (# months)	MN
Loc	Locomotion mode: metric indicating morphology and locomotion	
LocAccl	Accelerators: ambush predators with large caudal fins	%TRT
LocAng	Anguilliform swimmers with eel-like bodies	%TRT
LocCrp	Creepers: bottom rovers with subterminal mouths	%TRT
LocCrsr	Cruisers: active swimmers with streamlined body forms	%TRT
LocHgr	Huggers: benthic fishes adapted to cling to substratum	%TRT
LocMnvr	Maneuverers with laterally compressed body forms	%TRT
LocStbl	Mode unique to Paddlefish (*Polyodon spathula*)	%TRT

Physical habitat data

Stream habitat within each of the identified subwatersheds was represented by the 1:100,000 scale National Hydrography Dataset version 2 ('NHDv2'; McKay et al. 2012). All NHDv2 stream segments within each of the 153 selected subwatersheds (see above) were queried then appended with physical habitat characteristics from the original NHDv2 attribute tables and the StreamCat database (Hill et al. 2016). StreamCat variables represent local habitat conditions, including both natural and anthropogenic variables (e.g. land cover, urban land use

and geologic data), and are distributed at both catchment (i.e. the landscape that is immediately adjacent to a given stream segment, exclusive of landscapes further upstream) and watershed scales (i.e. the entire, cumulative landscape that is upstream of and contributing flow to a given segment; see Hill et al. 2016). From StreamCat and NHDv2, we selected a suite of 22 catchment-scale physical habitat metrics to represent instream fish habitat, such as stream order and channel slope (Schlosser 1991). Catchment-scale metrics from StreamCat were selected, rather than watershed-scale, to better emphasize local conditions. For each of these physical habitat metrics, we calculated subwatershed means, medians, maximums, and coefficients of variation (CV), based upon the entire population of stream segments within a given subwatershed. Subwatershed summary statistics for each physical habitat metric were then appended to a subwatershed $\times$ habitat variable matrix and correlated variables ($\left| r \right| \geq 0.6$) were removed (see Table 1).

Fish and functional trait data

Fish occurrence records were obtained from the IchthyMaps database (Frimpong et al. 2015, 2016) for each of the 153 subwatersheds that satisfied our hydrologic criteria (see above). IchthyMaps is a compilation of contemporary fish records (collected between 1950 – 1990) that were assembled from multiple sources, then geo-referenced to their respective digital stream segments in the NHDv2. Within the study region, IchthyMaps provided a total of 32,463 fish occurrence records (distributed among 139 species), 559 of which were American Eel records. To account for potential sampling bias in the aggregated IchthyMaps data (i.e. under-sampled subwatersheds), we screened subwatersheds that contained low numbers of fish samples following McGarvey et al. (*in press*). Briefly, we calculated sample density as the total number of IchthyMaps observations (i.e. total number of occurrence records, regardless of species'

identity) within a given subwatershed, divided by the total length of all stream segments within that subwatershed. This helped to detect subwatersheds that were vulnerable to under-sampling bias. Subwatersheds with fish sample densities below the 10th percentile (i.e. ≤ 3.0 occurrences per km of stream channel; median fish sample density = 11.3, CV = 1.3) were removed ($n = 16$) from the dataset.

Fish species' functional traits were then used to quantify biotic influences within each of the remaining subwatersheds ($n = 137$). We first compiled all IchthyMaps records within each subwatershed and converted the species list into a master species' presence matrix (subwatershed $\times$ species). Next, we collected functional trait data for each species through an extensive literature review, inclusive of the four data sources reported in Mims et al. (2010): (1) regional fish atlases (i.e. 'The fishes of…' texts); (2) primary and secondary literature publications; (3) the online FishBase (<www.fishbase.org>) and FishTraits (<www.fishtraits.info>) databases; and (4) keyword internet searches. For species' traits with multiple published values, we retained observations that were reported from localities that were closest to our study area. For instance, functional trait descriptions from the *Freshwater Fishes of Virginia* (Jenkins and Burkhead 1993) were prioritized over other sources for many of our trait values. Functional trait data were then compiled into a species $\times$ trait matrix and categorical variables were re-coded as binary dummy variables.

Finally, functional trait summaries were calculated for each subwatershed. Mean, median, and maximum values were calculated for numeric traits while proportions (% trait modality) were calculated for categorical traits. Importantly, we omitted American Eel traits from the subwatershed trait summaries because our objective was to detect interspecific trait influences on contemporary Eel occurrences; we sought only to determine which traits of co-occurring species

may affect Eel habitat select, not to define whole assemblage trait profiles or to assess

intraspecific effects. We then compiled subwatershed trait summary values into a subwatershed

× trait matrix and removed highly correlated variables ($|r| \geq 0.7$; see Table 1).

Comparing sample and background distributions

After each of the above screening and data preparation steps was complete, our final

dataset consisted of 24 subwatersheds with confirmed American Eel presences and 113 pseudo-

absence (i.e. Eel presence was unknown) subwatersheds (Fig. 3, step 1). Thus, we had 24 sample

units and 137 background units (i.e., the combined landscape of 113 pseudo-absences and 24

samples; see Hirzel et al. 2002, Elith et al. 2011) to use in our analyses. We then implemented a

random permutation algorithm to assess similarities between the sample and background

distributions for each of the physical habitat, hydrologic, and functional trait variables.

Permutations were used instead of direct, static comparisons of the complete sample and

background distributions for each variable because we did not wish to overfit our results, thereby

leading to low transferability or generality (Thomas and Bovee 1993, Chatfield 1995).
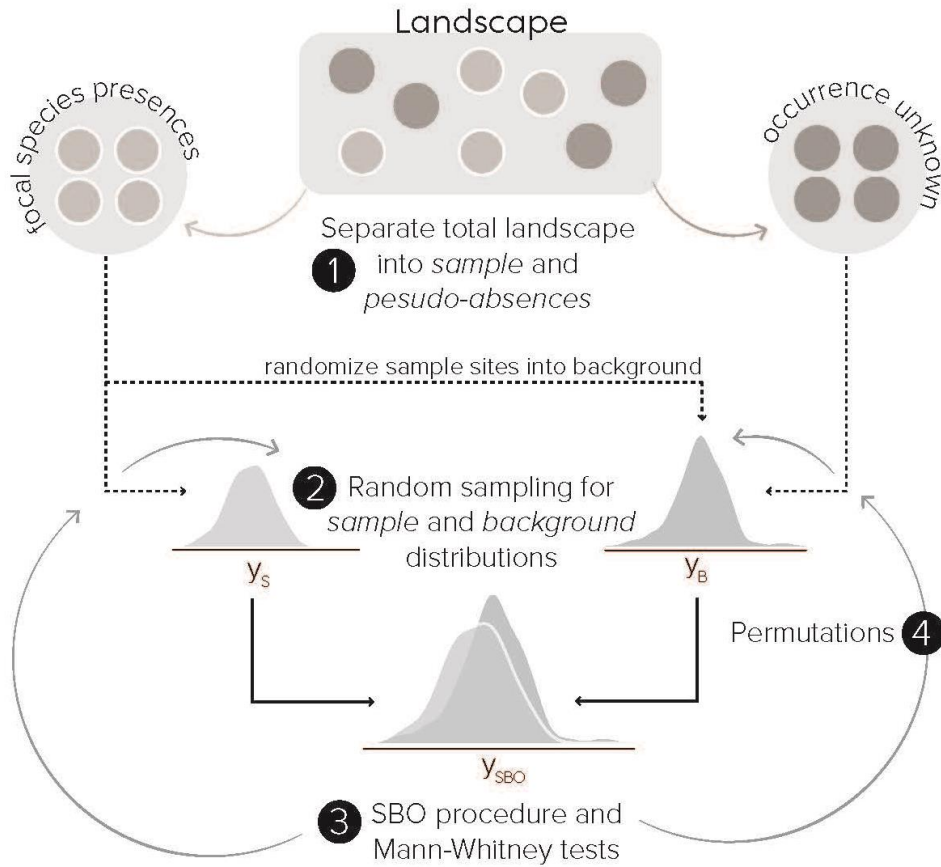
Figure 3. Workflow diagram of the procedure to compute permuted sample-background overlap (SBO) and Mann-Whitney U statistics for each of the predictor variables included in this study. Major steps in the process are identified with numbers and explained in the Materials and Methods (main text). Here, the total landscape consisted of all subwatersheds within the six study river basins.

In each of 1000 permutations, we randomly selected (without replacement) 12 of the 24 sample units and 60 of the background units (Fig. 3, step 2). These permuted sample sizes preserved the original, approximate ratio of sample:pseudo-absence units (~ 1:5) and ensured that our results would not be biased by unbalanced representation of the sample or pseudo-absence data in any given permutation. We then compared the randomized sample and background distributions for each of the physical habitat, hydrologic, and functional trait variables with a 2-step process (Fig. 3, step 3). First, we quantified sample-background overlap (SBO) as the literal area of overlap between continuous sample and background distributions

(i.e. superimposed probability distribution functions), with potential values ranging from 1 (perfect overlap) to 0 (no overlap). SBO values close to 1 suggest that sample and background units are both random samples from the complete landscape (i.e. sample values for the predictor variable do not differ with respect to the background values, H_o). Alternatively, SBO values much smaller than 1 imply that sample units are nonrandom with respect to the landscape (i.e. the sample values encompass a relatively small subset of the complete range of background values, indicative of a selective process, H_A; see Hirzel et al. 2002).

For each permutation and all predictor variables, SBO was calculated with equation 4 of Mouillot et al. (2005); their equation 4 was conceived as a metric of overlap in the trait densities of two co-occurring species, using kernel density functions (see also Mason et al. 2008, 2011). Our application of Mouillot et al. (2005) was, however, fundamentally different; while the original authors interpreted a high degree of overlap among kernel functions as evidence of potential competition, we sought to identify variables with *minimal* overlap as evidence of a strong, non-random influence on Eel distribution. Because many of the variables considered here were non-normal and/or contained zero values (e.g. dam density), we applied a ln $(x + 1)$ transformation to all continuous variables. All kernel density estimates were calculated with the *density* function (Gaussian kernels and default bandwidth and n settings) in the stats package in R.

In the second step, nonparametric Mann-Whitney U-statistics were used to determine whether permuted sample values were consistently higher or lower than permuted background values for each of the physical habitat, hydrologic, and functional trait variables. Two-sided statistics were used because we did not expect *a priori* that Eel sample values would consistently be higher or lower than background values for most of the predictor variables; we sought only to

document whether the sample and background distributions were consistently different. In each

permutation, we calculated and recorded the *U*-statistic from a 2-sided Mann-Whitney test for

each variable, using the *wilcox.test* function in R. *U*-statistic values were then interpreted relative

to the magnitude of deviation from an equal ranking of sample and background values (i.e. *U* =

360). *U* values closer to zero indicated that sample distribution values were consistently ranked

higher or were larger than background values, whereas *U* values approaching the maximum (i.e.

U = 720) indicated that sample values were consistently ranked lower or were smaller than

background values for a given variable.

As a final step, we used a multivariate SBO approach to assess the collective effect of

each of the three classes of variables on Eel distributions throughout the study rivers. For each

class of predictor variable, nonmetric multidimensional scaling (nMDS) was used to build a 2-

dimensional ordination of the sample and background data. Ninety-five percent confidence

ellipses, or 'hypervolumes' (Blonder et al. 2014), were interpolated for the sample points and

background points in each ordination plot, with the regions defined by the background ellipses

indicating the expected hypervolume for a random sample of the complete landscape. We then

visually assessed the degree of overlap between sample and background hypervolumes to

determine whether Eel occurrences were nonrandom with respect to the background

hypervolume for each of the three predictor variable classes. We did not, however, use

permutations to perform repeated comparisons of the sample and background hypervolumes.

Instead, we used the complete sample and background data (*n* = 24 and 137, respectively) to

perform a single multivariate SBO analysis for each of the three classes of predictor variables.

For each predictor class, we calculated a Gower dissimilarity matrix (Gower 1971) that included

all variables within the class (i.e. three independent subwatershed × subwatershed dissimilarity

matrices), the used the dissimilarity matrices to perform nMDS. Gower dissimilarities were

calculated with the 'FD' package (Laliberte et al. 2014), nMDS was performed with the 'vegan'

package (Oksanen et al. 2017), and confidence ellipses were plotted with 'ggplot2' (Wickham

2009) in R.

Data deposition

The raw data and code for analyses are found on figshare:

<https://doi.org/10.6084/m9.figshare.5481205.v3> (Woods and McGarvey 2017).

Results

Permuted statistics for individual predictor variables

Of the three predictor variable classes, hydrologic variables generally exhibited the

highest SBO values (median SBO = 0.83, CV = 0.06; Fig 4a), suggesting that contemporary

occurrences of the American Eel are not strongly associated with specific hydrologic conditions.

Only bsflwCV had low permuted SBO values (< 0.75; Fig. 4a) and *U* values that strongly

deviated from the line of equality, with sample values consistently larger than background values

(Fig. 4b). Sample and background distributions differed minimally with respect to the remaining

hydrologic variables. In comparison, physical habitat variables appeared to have greater

influence on the contemporary Eel distribution (median SBO = 0.76, CV = 0.16; Fig 4a). In the

physical habitat class, low permuted SBO values were observed for dmDnsMN and elevMN

(Fig. 4a). *U* values showed that sample observations were consistently larger than background

observations for dmDnsMN, but the opposite trend was observed for elevMN (Fig. 4b). Overall,

the functional trait data class had the lowest SBO values (median SBO = 0.69, CV = 0.32), with

particularly low permutation values for locMnvr, locCrsr, incbTm, and seasLen (Fig. 4a).

Sample distribution values of locMnvr and seasLen were higher than background, whereas

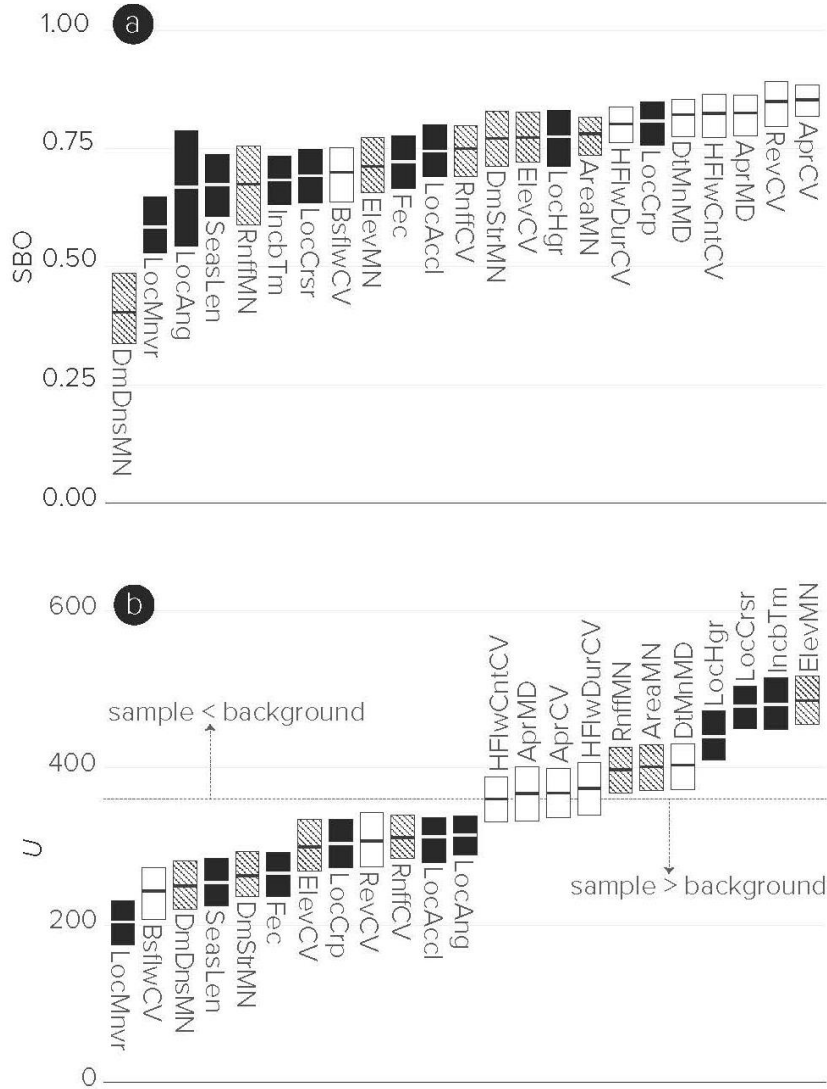locCrsr and incbTm background values exceeded sample values (Fig. 4b).

Figure 4. Boxplots (25th, 50th, and 75th quartiles) showing permutation results as sample-background overlap (SBO; panel a) and Mann-Whitney U-statistic (panel b) values. Boxplot colors denote the three classes of predictor variables: hydrologic (white, black outline), physical habitat (grey hatched, black outline), and functional traits (black, white outline). In panel b, the horizontal line at U = 360 represents rank equivalence between sample and background distribution values. Below the rank equivalence line, sample distribution values are, on average, larger than background values for the variable in question. Above the rank equivalence line, sample distribution values are smaller than background values for the variable in question.

24
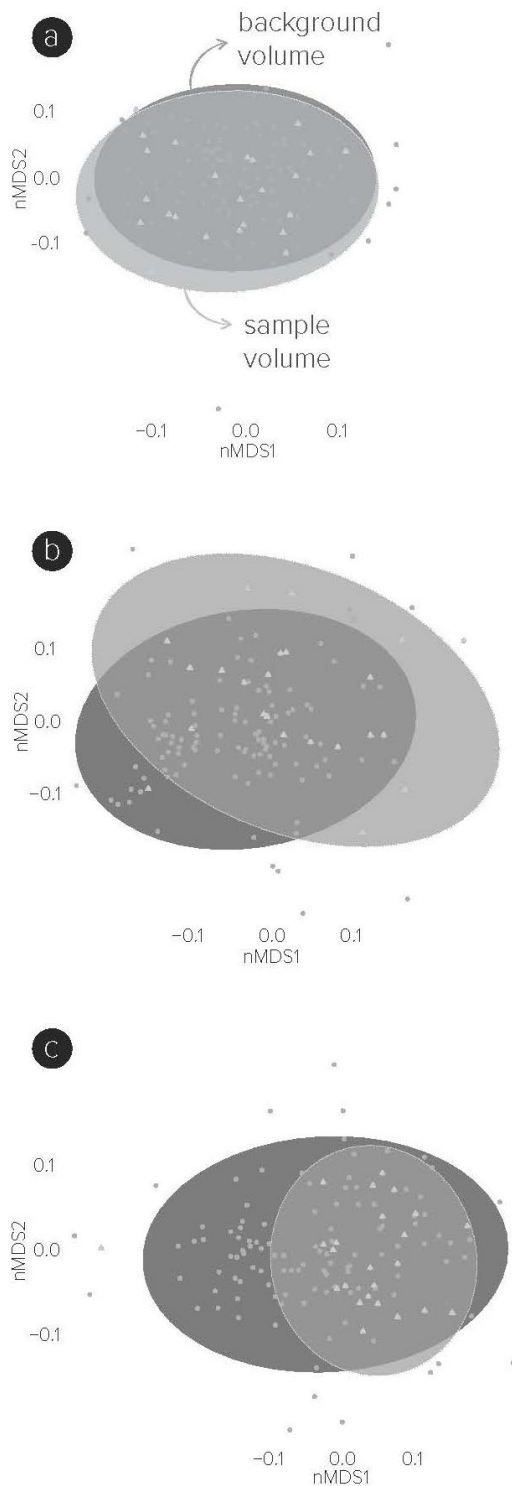
Multivariate hypervolumes

Consistent with permutation results for individual variables, comparisons of the multivariate sample and background hypervolumes for the three data classes indicated that functional traits may have the overall strongest influence on Eel occurrence. Sample and background hypervolumes exhibited extensive overlap for hydrologic variables (Fig. 5a). Overlap between sample and background hypervolumes was intermediate for physical habitat variables (Fig. 5b). Hypervolume overlap was smallest for the functional trait data class, with approximately 50% overlap observed between the sample and background hypervolumes (Fig. 5c).

< Figure 5. Nonmetric multidimensional scaling (nMDS) ordination plots for the hydrologic (a), physical habitat (b), and functional trait (c) data classes. Hypervolumes are shown as 95% confidence ellipses for sample (light grey) and background (dark grey) data. Points represent individual subwatersheds from the sample distribution (light grey triangles) and the background distribution (dark grey circles).

Discussion

Abiotic and biotic influence on American Eel distribution

Within six Mid-Atlantic rivers, we compared abiotic and biotic conditions at confirmed American Eel presence sites (the sample distribution) against a representative sample of conditions across the entire landscape where Eel occurrence was unknown (the background distribution). In general, results indicated that the sample distribution was more closely associated with physical habitat and the functional traits of co-occurring fishes than with hydrology.

Dam density appeared to have one of the strongest influences on Eel distribution. This result seemed intuitive because it is well-known that dams prevent migratory Eels from reaching upstream tributaries (Wiley et al. 2004, Machut et al. 2007) and can influence Eel distributions more than other local scale, abiotic factors (Hitt and Roberts 2012). However, the direction of influence of dam density values was surprising: rather than associating with low dam densities, the Eel sample distribution suggested that Eels may be associated with relatively high dam densities. This counterintuitive result may be explained by a tendency for Eels to aggregate near dams. Prior studies have shown that Eel densities increase near dams and suggested that this may result from large congregations of Eels that gather downstream of impassable barriers (Machut et al. 2007). We mapped our Eel presence records against georeferenced dam locations and found that the sample distribution may have included multiple sites that were downstream of an impassible dam (these sites would have high dam density values) and therefore represented the upstream limits of potential Eel migration (see Appendix, Fig. A1). If so, the strong association between Eel presence and high dam density may be an artifact of higher probability of Eel

capture at these sites, rather than a preferential selection of subwatersheds with high dam densities *per se*.

We found that local hydrology may play a minimal role in determining Eel presence. The only variable within this class that showed an effect on Eel distribution (bsflwCV) seems to suggest that the sample distribution may be associated with habitat characterized by variable or unpredictable flows. These results are surprising because Eels are migratory and might therefore be expected to associate with specific, predictable flow conditions (Bunn and Arthington 2002). One possible explanation for the lack of a strong, consistent Eel-flow relationship is that the set of hydrologic predictor variables did not include the flow components that are truly of greatest importance to migratory Eels. Nonetheless, these results are consistent with studies that failed to identify strong associations between Eel populations and local abiotic variables (Smogor et al. 1995, Wiley et al. 2004).

Overall, functional trait characteristics tended to be important in differentiating sample and background distributions. To explain these results, we group the trait variables into two categories: reproductive behavior and locomotion. Reproductive trait results indicate that Eels may associate with heterospecific fishes that have high reproductive capacities shown through shorter incubation times and longer spawning seasons. Locomotion traits suggest that Eels are highly associated with heterospecifics that exhibit maneuvering locomotion modes, but not with cruising species. Taken together, these functional trait results may suggest that Eels locally coexist with a suite of traits that provide lowered resource competition and increased prey availability. Eels are thought to compete with cruising species whose active streamlined body forms make them superior predators to comparatively slow anguilliform species (Sinha and Jones 1967) and our results suggest Eels may select sites with lower proportions of these species.

Alternatively, Eels may favor coexistence with maneuvering species because these comparatively sluggish locomotion modes make them less adept competitors and easier prey.

Functional traits as proxies for biotic interactions

Pairwise (species × species) approaches have traditionally been used to study biotic interactions, with methods ranging from basic null models to more recent developments in network analysis. However, these approaches can be computationally impractical because the number of species pairs that may potentially interact grows quickly as the number of locally co-occurring species increases (Morales-Castilla et al. 2015). In the present context, information on 138 pairwise associations would be needed to fully account for all direct biotic interactions between the American Eel and locally co-occurring heterospecifics. More efficient methods to account for biotic interactions are clearly needed.

Using functional traits to generalize processes that are logistically difficult to quantify on a species × species basis is currently an active area of research. For example, body size may be used to estimate dispersal in metacommunity frameworks (De Bie et al. 2012), multidimensional trait classifications can be used to link species to their ecosystem functions (Winemiller et al. 2015), and knowledge of trait performance in relation to the environment may be used to predict community responses to abiotic filtering (Webb et al. 2010). Here, we demonstrated a general, flexible method to quantifying biotic interactions using community functional traits. The question now is whether the observed sample-background trait differences truly convey the influence of biotic interactions. Direct, empirical evidence linking traits to interaction effects through measurable changes in population growth rates is currently lacking (Alexander et al. 2016). But obtaining this empirical evidence may be difficult or impossible, given that relatively

28

few species pairs exhibit strong interactions (e.g. predation or obligate mutualism), while weak interactions (e.g. size-dependent competitive encounters) may be common. Therefore, less specific methods that use functional traits to infer species' ecosystem roles may be a better way to operationalize traits-based approaches to biotic interactions (McGill et al. 2006).

Our study demonstrates two benefits of a such a generalized, traits-based approach. First, as noted above, community trait analyses are more computationally feasible and less data-limited than biotic filtering studies that rely upon large numbers of pairwise species contrasts. Second, removing taxonomy from the study of biotic interactions makes results transferable between disparate regions and at coarser spatial scales. Therefore, our traits-based results are potentially relevant to anguillids globally, whereas taxonomic analyses will necessarily change with turnover in the composition of local fish assemblages. Anguillids demonstrate this benefit particularly well because freshwater Eels worldwide exhibit similar life histories and results derived from one species of *Anguilla* may be transferrable to congeners (Haro et al. 2000). For example, our results for American Eel distribution may apply to European Eel (*A. anguilla*) or New Zealand Longfin Eel (*A. dieffenbachia*) distributions.

SBO logic as a flexible tool to compare abiotic and biotic influences

Hierarchal frameworks have traditionally presumed that biotic filtering is a local-scale process. However, it is now recognized that biotic filtering can also drive species distribution patterns at large spatial scales (Kraft et al. 2015). A need for methods that directly compare biotic and abiotic influences at common scales has therefore emerged and we conclude by discussing some limitations and benefits of our method to make such comparisons.

One potential caveat in our results is the possibility that inferred biotic effects (i.e. functional traits with low permuted SBO values and high magnitude U) could be artifacts of abiotic filtering processes that regulate species' presences and, by extension, perceived functional trait patterns. For instance, our observation that Eels are highly associated with the prevalence of maneuvering (locomotion mode) heterospecifics could be artefactual if dams have a strong influence on the presence of maneuvering fishes. In this case, we do not believe that our biotic filtering results are spurious. Using linear regression, we found no evidence that that the proportion of maneuvering fishes was indicative of species composition downstream of dams ($F_{1,111} = 1.19$, $p = 0.278$). However, we acknowledge that post-hoc tests to confirm that a given functional trait result is not an artifact of another abiotic variable will often be necessary, either for discrete pairs of variables (e.g. linear regression) or entire networks of variables (e.g. structural equation modeling).

Perhaps the greatest benefit of our approach is that the use of a background distribution that incorporates sample and pseudo-absence data leads to stronger inferences than studies based solely on an empirical sample distribution. By evaluating the sample distribution in the context of a regional background distribution, we can begin to identify variables that truly drive filtering processes, whereas methods that omit an additional background context can only summarize observed species-habitat associations. Therefore, this framework may provide a powerful context for selecting covariates in species distribution models. For instance, a logical next step from our analysis would be to develop formal models of American Eel habitat suitability, using results from the SBO procedure to inform the covariate selection process. But the extent of our data (i.e. not a continuous representation of the entire range of the American Eel) limited our ability to conduct a robust distribution modeling analysis. The full potential of our framework may be best

measured by applying more extensive datasets to our framework and selecting the 'top' ranked variables from the SBO procedure for inclusion in the species distribution models.

The SBO framework is flexible and can be applied to numerous species in diverse settings and at varying spatial scales. Thus, we believe this approach has potential to shed new light on the relative roles of abiotic and biotic filtering in community assembly. Importantly, our approach is readily compatible with a growing number of publicly available climate, physical habitat, functional trait, and species' occurrence datasets. To quantify the 'biotic background', we used community mean functional trait values, but the framework is flexible enough to accommodate any number or class of predictor variables. For instance, future studies could include emergent indicators of community traits, such as functional diversity indices (Villéger et al. 2008), or measures of phylogenetic similarity. A better understanding of the role that biotic filtering plays in regulating community assembly will also benefit conservation efforts, particularly when predicting community responses to global change (Blois et al. 2013).

References

Alexander, J. M. et al. 2016. When Climate Reshuffles Competitors: A Call for Experimental Macroecology. – Trends Ecol. Evol. 31: 831–841.

Araújo, M. B. and Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. – Glob. Ecol. Biogeogr. 16: 743–753.

Bascompte, J. and Jordano, P. 2007. Plant-Animal Mutualistic Networks: The Architecture of Biodiversity. – Annu. Rev. Ecol. Evol. Syst. 38: 567–593.

Blois, J. L. et al. 2013. Climate change and the past, present, and future of biotic interactions. – Science 341: 499–504.

Blonder, B. et al. 2014. The n-dimensional hypervolume. – Glob. Ecol. Biogeogr. 23: 595–609.

Boulangeat, I. et al. 2012. Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. – Ecol. Lett. 15: 584–593.

Bunn, S. E. and Arthington, A. H. 2002. Basic Principles and Ecological Consequences of Altered Flow Regimes for Aquatic Biodiversity. – Environ. Manage. 30: 492–507.

Cadotte, M. W. and Tucker, C. M. 2017. Should Environmental Filtering be Abandoned? – Trends Ecol. Evol. 32: 1–9.

Carmona, C. P. et al. 2016. Traits Without Borders: Integrating Functional Diversity Across Scales. – Trends Ecol. Evol. 31: 382–394.

Cazelles, K. et al. 2016. A theory for species co-occurrence in interaction networks. – Theor. Ecol. 9: 39–48.

Chase, J. M. and Leibold, M. A. 2003. Ecological Niches: Linking Classical and Contemporary Approaches. – The University of Chicago Press.

Chatfield, C. 1995. Model uncertainty, data mining and statistical inference. – J. R. Stat. Soc.

Ser. A Stat. Soc. 158: 419–466.

Connor, E. F. and Simberloff, D. 1979. The Assembly of Species Communities: Chance or Competition? – Ecology 60: 1132–1140.

Connor, E. F. et al. 2013. The checkered history of checkerboard distributions. – Ecology 94: 2403–2414.

De Bie, T. et al. 2012. Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. – Ecol. Lett. 15: 740–747.

Elith, J. et al. 2011. A statistical explanation of MaxEnt for ecologists. – Divers. Distrib. 17: 43–57.

Frimpong, E. and Angermeier, P. 2010. Trait-Based Approaches in the Analysis of Stream Fish Communities. – Am. Fish. Soc. Symp. 73: 109–136.

Frimpong, E. et al. 2015. Historical Stream Fish Distribution Database for the Conterminous United States (1950-1990): IchthyMaps. – U.S. Geol. Surv.

Frimpong, E. A. et al. 2016. IchthyMaps: A Database of Historical Distributions of Freshwater Fishes of the United States. – Fisheries 41: 590–599.

Funk, J. L. et al. 2016. Revisiting the Holy Grail: using plant functional traits to understand ecological processes. – Biol. Rev. 92: 1156–1173.

Gagnaire, P. A. et al. 2012. The genetic consequences of spatially varying selection in the panmictic american eel (*Anguilla rostrata*). – Genetics 190: 725–736.

Gallardo, B. and Aldridge, D. C. 2013. Evaluating the combined threat of climate change and biological invasions on endangered species. – Biol. Conserv. 160: 225–233.

Gallien, L. et al. 2017. The effects of intransitive competition on coexistence. – Ecol. Lett. 20: 791–800.

Gan, K. C. et al. 1991. Analysis of periodicity in streamflow and rainfall data by Colwell's

indices. – J. Hydrol. 123: 105–118.

Gotelli, N. J. and McCabe, D. J. 2002. Species co-occurrence: A meta-analysis of J. M.

Diamond's assembly rules model. – Ecology 83: 2091–2096.

Gower, J. 1971. A General Coefficient of Similarity and Some of Its Properties. – Biometrics 27:

857–871.

Guisan, A. and Thuiller, W. 2005. Predicting species distribution: Offering more than simple

habitat models. – Ecol. Lett. 8: 993–1009.

Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. –

Ecol. Model. 135: 147–186.

Haro, A. et al. 2000. Population Decline of the American Eel: Implications for Research and

Management. – Fisheries 25: 7–16.

Hill, R. A. et al. 2016. The Stream-Catchment (StreamCat) Dataset: A Database of Watershed

Metrics for the Conterminous United States. – J. Am. Water Resour. Assoc. 52: 120–128.

HilleRisLambers, J. et al. 2012. Rethinking Community Assembly through the Lens of

Coexistence Theory. – Annu. Rev. Ecol. Evol. Syst. 43: 227–248.

Hirzel, A. H. et al. 2002. Ecological-niche factor analysis: how to compute habitat-suitability

maps without absence data? – Ecology 83: 2027–2036.

Hitt, N. P. and Roberts, J. H. 2012. Hierarchical spatial structure of stream fish colonization and

extinction. – Oikos 121: 127–137.

Huang, J. and Frimpong, E. A. 2016. Limited transferability of stream-fish distribution models

among river catchments: reasons and implications. – Freshw. Biol. 61: 729–744.

Hutchinson, G. E. 1957. Concluding Remarks. – Cold Spring Harb. Symp. Quant. Biol. 22: 415–

427.

Jackson, D. A. et al. 2001. What controls who is where in freshwater fish communities - the roles
of biotic, abiotic, and spatial factors. – Can. J. Fish. Aquat. Sci. 58: 157–170.

Jenkins, R. and Burkhead, N. 1993. Freshwater Fishes of Virginia. – American Fisheries Society.

Kraft, N. J. B. et al. 2015. Plant functional traits and the multidimensional nature of species
coexistence. – PNAS 112: 797–802.

Laliberte, E. et al. 2014. FD: Measuring functional diversity (FD) from multiple traits, and other
tools for functional ecology.

Law, J. 2013. IHA: This package implements The Nature Conservancy's Indicators of
Hydrologic Alteration software in R.

Lawton, J. H. 1999. Are There General Laws in Ecology? – Oikos 84: 177–192.

Leathwick, J. R. and Austin, M. P. 2001. Competitive interactions between tree species in New
Zealand old-growth indigenous forests. – Ecology 82: 2560–2573.

MacArthur, R. and Levins, R. 1967. The Limiting Similarity, Convergence, and Divergence of
Coexisting Species. – Am. Nat. 101: 377–385.

Machut, L. S. et al. 2007. Anthropogenic impacts on American eel demographics in Hudson
River tributaries, New York. – Trans. Am. Fish. Soc. 136: 1699–1713.

Mason, N. W. H. et al. 2008. Does niche overlap control relative abundance in French lacustrine
fish communities? A new method incorporating functional traits. – J. Anim. Ecol. 77: 661–
669.

Mason, N. W. H. et al. 2011. Niche overlap reveals the effects of competition, disturbance and
contrasting assembly processes in experimental grassland communities. – J. Ecol. 99: 788–
796.

McGarvey, D. J. et al. in press. On the use of climate covariates in aquatic species distribution

    models : are we at risk of throwing out the baby with the bath water? (doi:

    10.1111/ecog.03134)

McGill, B. J. et al. 2006. Rebuilding community ecology from functional traits. – Trends Ecol.

    Evol. 21: 178–185.

McKay, L. et al. 2012. NHDPlus Version 2: User Guide.

Meier, E. S. et al. 2011. Co-occurrence patterns of trees along macro-climatic gradients and their

    potential influence on the present and future distribution of *Fagus sylvatica* L. – J.

    Biogeogr. 38: 371–382.

Merow, C. et al. 2013. A practical guide to MaxEnt for modeling species' distributions: What it

    does, and why inputs and settings matter. – Ecography. 36: 1058–1069.

Mims, M. C. et al. 2010. Life history trait diversity of native freshwater fishes in North America.

    – Ecol. Freshw. Fish 19: 390–400.

Morales-Castilla, I. et al. 2015. Inferring biotic interactions from proxies. – Trends Ecol. Evol.

    30: 347–356.

Mouillot, D. et al. 2005. Niche overlap estimates based on quantitative functional traits: A new

    family of non-parametric indices. – Oecologia 145: 345–353.

Oksanen, J. et al. 2017. Vegan: community ecology package.

Olden, J. D. and Poff, N. L. 2003. Redundancy and the choice of hydrologic indices for

    characterizing streamflow regimes. – River Res. Appl. 19: 101–121.

Oliveira, K. 1999. Life history characteristics and strategies of the American eel, *Anguilla*

    rostrata. – Can. J. Fish. Aquat. 56: 795–802.

Ovaskainen, O. et al. 2017. How to make more out of community data? A conceptual framework

and its implementation as models and software. – Ecol. Lett. 20: 561–576.

Parmesan, C. and Yohe, G. 2003. A globally coherent fingerprint of climate change impacts across natural systems. – Nature 421: 37–42.

Pearce, J. L. and Boyce, M. S. 2006. Modelling distribution and abundance with presence-only data. – J. Appl. Ecol. 43: 405–412.

Pearson, R. G. and Dawson, T. P. 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? – Glob. Ecol. Biogeogr. 12: 361–371.

Peres-Neto, P. R. 2004. Patterns in the co-occurrence of fish species in streams: the role of site suitability, morphology and phylogeny versus species interactions. – Oecologia 140: 352–360.

Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – Ecol. Model. 190: 231–259.

Poff, N. L. 1997. Landscape Filters and Species Traits: Towards Mechanistic Understanding and Prediction in Stream Ecology. – J. North Am. Benthol. Soc. 16: 391–409.

R Core Team 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Renner, I. W. et al. 2015. Point process models for presence-only analysis. – Methods Ecol. Evol. 6: 366–379.

Richter, B. D. et al. 1996. A Method for Assessing Hydrologic Alteration within Ecosystems. – Conserv. Biol. 10: 1163–1174.

Schlosser, I. J. 1991. Stream Fish Ecology: A Landscape Perspective. – Bioscience 41: 704–712.

Sinha, V. R. P. and Jones, J. W. 1967. On the food of the freshwater eels and their feeding

relationship with the salmonids. – J. Zool. 153: 119–137.

Smogor, R. A. et al. 1995. Distribution and Abundance of American Eels in Virginia Streams: Tests of Null Models across Spatial Scales. – Trans. Am. Fish. Soc. Am. Fish. Soc. 124: 789–803.

Staniczenko, P. P. A. et al. 2017. Linking macroecology and community ecology: refining predictions of species distributions using biotic interaction networks. – Ecol. Lett. 20: 693–707.

Thomas, J. A. and Bovee, K. D. 1993. Application and testing of a procedure to evaluate transferability of habitat suitability criteria. – Regul. River. 8: 285–294.

Tylianakis, J. M. et al. 2008. Global change and species interactions in terrestrial ecosystems. – Ecol. Lett. 11: 1351–1363.

US Geological Survey (USGS) 2012. Federal standards and procedures for the National Watershed Boundary Dataset (WBD): Techniques and Methods 11-A3.

Villéger, S. et al. 2008. New multidimensional functional diversity indices for a multifaceted framework in functional ecology. – Ecology 89: 2290–2301.

Webb, C. T. et al. 2010. A structured and dynamic framework to advance traits-based theory and prediction in ecology. – Ecol. Lett. 13: 267–283.

Wickham, H. 2009. ggplot2: Elegant Graphics for Data Analysis.

Wiens, J. J. 2011. The niche, biogeography and species interactions. – Philos. Trans. R. Soc. B Biol. Sci. 366: 2336–2350.

Wiley, D. J. et al. 2004. Relations between Physical Habitat and American Eel Abundance in Five River Basins in Maryland. – Trans. Am. Fish. Soc. 133: 515–526.

Winemiller, K. O. et al. 2015. Functional traits, convergent evolution, and periodic tables of

niches. – Ecol. Lett. 18: 737–751.

Wisz, M. S. et al. 2013. The role of biotic interactions in shaping distributions and realized

assemblages of species: Implications for species distribution modelling. – Biol. Rev. 88:

15–30.

Woods, T. and McGarvey, D. J. 2017. Data and code from: A framework to assess the relative

influences of abiotic and biotic factors on a species' distribution using pseudo-absence and

functional trait data: a case study with the American Eel (*Anguilla rostrata*). – figshare,

<https://doi.org/10.6084/m9.figshare.5481205.v3>.

Wooton, T. 1994. Putting the Pieces Together : Testing the Independence of Interactions among
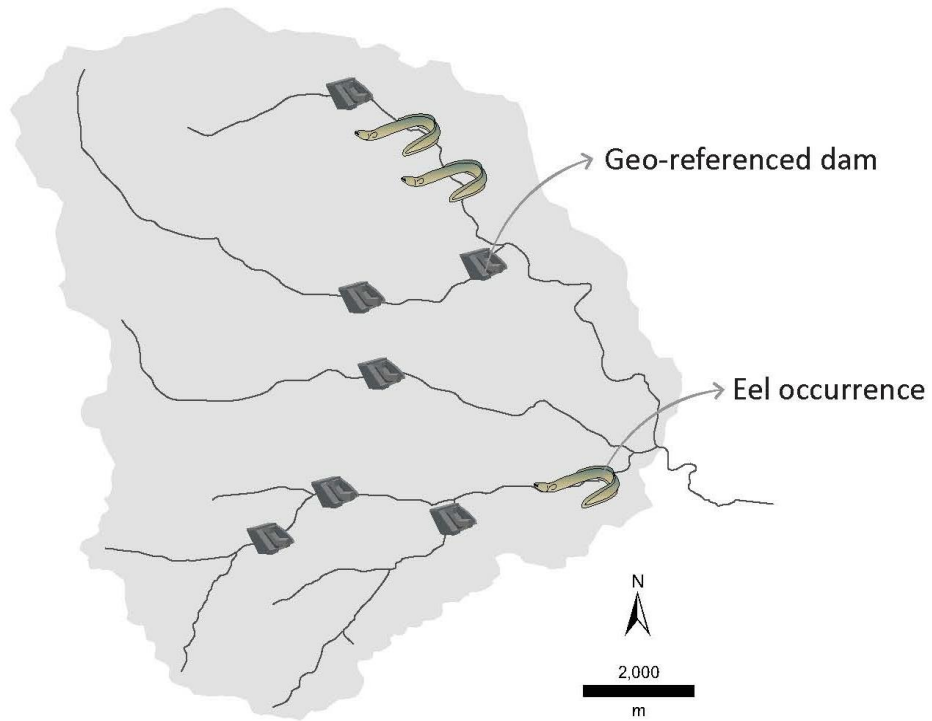
Organisms. – Ecology 75: 1544–1551.

Figure A1. An illustration of a subwatershed exhibiting high dam density, effectively preventing the Eels from moving further upstream. Downstream of large, impassible dams, Eels tend to congregate and are therefore more easily captured during stream surveys.