Nestedness patterns and the dual nature of community reassembly in California streams: a multivariate permutation-based approach

MARK NOVAK*, JONATHAN W. MOORE*† and ROBERT A. LEIDY‡
*Department of Ecology and Evolutionary Biology, Long Marine Laboratory, University of California, 100 Shaffer Rd., Santa Cruz, CA, 95060, USA, †Department of Biology, Simon Fraser, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada, ‡U.S. Environmental Protection Agency, 75 Hawthorne Street, San Francisco, CA, 94105, USA

Abstract
Many factors contribute to the nonrandom processes of extinctions and invasions that are changing the structure of ecological communities worldwide. These factors include the attributes of the species, their abiotic environment, and the interactions and feedbacks between them. The relative importance of these factors has been difficult to quantify. We used nested subset theory and a novel permutation-based extension of gradient analysis to disentangle the direct and indirect pathways by which these factors affect the metacommunity structure of freshwater fishes inhabiting the streams tributary to the San Francisco Bay. Our analyses provide quantitative measures of how species and stream attributes may influence extinction vulnerability and invasion risk, highlight the need for considering the multiple interacting drivers of community change concurrently, and indicate that the ongoing disassembly and assembly of Bay Area freshwater fish communities are not fully symmetric processes. Fish communities are being taken apart and put back together in only partially analogous trajectories of extinction and invasion for which no single explanatory hypothesis is sufficient. Our study thereby contributes to the forecasting of continued community change and its effects on the functioning of freshwater ecosystems.

Keywords: body size, conservation, disturbance, extinction risk, freshwater stream fishes, invasion risk, metacommunity structure, multivariate gradient analysis, trait-based biogeography

Received 13 January 2011 and accepted 28 May 2011

Introduction
Natural and human-caused changes in the richness and composition of ecological communities are occurring by both species extinctions and species invasions, processes collectively referred to as community reassembly. Native community disassembly and nonnative community assembly are often nonrandom, but there remains great uncertainty as to the importance of their diverse potential drivers. Most hypotheses relate invasion success or extinction vulnerability either to the attributes of the species (e.g., the ‘reckless invader’ hypothesis) or their environment (e.g., the ‘biotic acceptance’ hypothesis), but few studies have considered these components of community change together (Catford et al., 2009). Furthermore, although many hypotheses consider the interactions between natives and nonnatives to be crucial determinants of community invasibility (e.g., the ‘biotic resistance’ hypothesis), few studies have investigated the potentially shared biotic and abiotic drivers of nonnative success and native vulnerability in unison (García-Berthou, 2007; Catford et al., 2009). Evidence suggests, however, that community reassembly is typically driven by the direct and indirect effects of multiple mechanisms acting simultaneously (e.g., Light & Marchetti, 2007).

Understanding the potentially complicated processes of community reassembly is a critical goal. Effective prioritization of conservation and restoration efforts requires disentangling the pathways by which altered environments and species assemblages interact to drive further community change (Olden et al., 2010). Furthermore, distinguishing the factors that influence extinction vulnerability and invasion success is a key for moving beyond random assembly experiments to illuminate the ecosystem consequences of ongoing biodiversity change (Gross et al., 2005; McIntyre et al., 2007; Zavaleta et al., 2009).

Case in point are the accelerating changes exhibited by freshwater fish communities (Dudgeon et al., 2006; Jelks et al., 2008) whose ecosystem-wide effects on primary productivity, food web structure, and nutrient dynamics are well known (Power, 1990; Schindler et al., 1997; Vanni, 2003). For example, previous analyses of local case histories in California suggest that nonnative establishment is
driven primarily by the suitable matching of species physiological attributes and abiotic conditions, with success or failure being largely independent of the recipient community (Moyle & Light, 1996a,b). State-wide watershed-scale patterns of positively correlated native and nonnative richness have supported this inference (Marchetti et al., 2004a,b). However, watershed-scale patterns also have suggested that the effects of altered abiotic conditions on native species are largely indirect, mediated less by environmental change per se than by the tendency of altered habitats to support nonnative species that consume or compete with natives (Light & Marchetti, 2007).

We employed nested subset theory to investigate the importance of the varied factors driving fish community structure in the streams tributary to the estuary of San Francisco Bay. This theory has long been recognized as a useful framework for identifying the mechanisms affecting changes in the composition of local communities (Patterson & Atmar, 1986) and has seen extensive application in the contexts of biogeography and conservation science (Fleishman et al., 2007). We developed a novel permutation-based extension of nested subset gradient analysis that allowed us to disentangle the direct and indirect drivers of community composition by overcoming two limitations of previous analyses. Specifically, our method circumvents problems associated with rank order ties in explanatory variables and permits the explicit consideration of causal collinearity among putative drivers. It thereby enables the partitioning of the direct and indirect pathways by which both species and site attributes contribute to metacommunity structure, providing relative measures of their putative influence on the risks of extinction and invasion. Application of our approach to the fish communities of San Francisco Bay freshwater streams highlights the need for considering the multiple interacting drivers of community composition explicitly and concurrently and provides insight into the dual nature of community reassembly in this region.

Materials and methods

Nested subset theory

Nested metacommunity patterns occur when species present at species-poor sites are a subset of the species present at species-rich sites (site nestedness) or when the occurrences of species occupying few sites are a subset of the occurrences of species occupying many sites (species nestedness; Fig. S1). The presence of such patterns is interpreted as evidence for species- or site-specific variation in rates of colonization (invasion) and extinction (Lomolino, 1996; Taylor & Warren, 2001), with sites harboring fewer species, or species occupying fewer sites, inferred to experience higher rates of extinction or lower rates of colonization (Atmar & Patterson, 1993).

The typical goal of nestedness analysis is to not only establish whether a nested metacommunity pattern exists, but also to infer its potential drivers. This is done using gradient analysis, where the sites or species of the metacommunity matrix are ordered by each of many one-dimensional variables (or some multivariate ordination thereof) hypothesized to affect colonization or extinction rates (Ulrich et al., 2009). The site or species trait that maximizes nestedness is considered the dominant driver of extinction or colonization (Ulrich et al., 2009). For some metrics of nestedness (e.g., nestedness metric based on paired overlap and decreasing fill (NODF), Almeida-Neto et al., 2008; see next), this procedure is equivalent to determining the site (or species) attribute exhibiting the strongest correlation to the species richness of sites (or the site occupancy of species). For other metrics, more indirect correlative inferences are required (e.g., matrix temperature; see Almeida-Neto et al., 2007). Of course, other causes such as sampling intensity and cross-site differences in habitat quality also can contribute to nestedness (Cam et al., 2000; Ulrich et al., 2009) and may be particularly probable when sites cannot be treated as independent replicates of one another or when species occurrences are phylogenetically influenced. Useful methods for dealing with such issues have recently been proposed (e.g., Leibold et al., 2010; Peres-Neto & Legendre, 2010), although not in the context of nested subset theory.

A significant limitation of the univariate approach to gradient analysis is an assumption that putative explanatory traits are themselves either uncorrelated or causally independent. Collinearity among predictors may be alleviated in part by calculating partial or semipartial correlations between predictor and response variables. However, predictors still are assumed to be causally independent. In many circumstances, such an assumption is unwarranted. For example, the oft-observed relationship between vulnerability and body size may be offset to an unknown degree by the counteracting relationship between body size and fecundity that may decrease a species’ vulnerability (Reynolds et al., 2005). Disentangling such correlations is facilitated by the explicit consideration of hypothesized causal relationships among all predictor and response variables.

A second limitation is introduced by the occurrence of ties in the rank ordering of an attribute. When such ties are present, as is commonly the case for species attributes such as trophic level or habitat affinity, the inferred explanatory power of the attribute may be considerably altered by the ordering of secondary attributes (Appendix S1). The effects of these secondary attributes remain hidden by univariate methods, potentially leading to incorrect inferences regarding the importance of putative reassembly drivers.

The method described next circumvents these limitations and enabled us to disentangle the manner by which a suite of species- and site attributes can explain the structure of a threatened freshwater fish community.

California fish communities

The region surrounding San Francisco Bay is a juxtaposition of degradation and diversity, exhibiting a rich fauna of fishes and high human population densities. The composition of
stream fish assemblages was characterized at 275 sites within 23 watersheds tributary to the estuary from 1993 to 2004 (94% prior to 1998; Appendix S2). Sites were stratified to maximize the diversity of representative habitat types (i.e., riffle, run, pool) in different geomorphic settings (high-elevation, high-gradient, bedrock to low-elevation, and low-gradient unconsolidated substrate). Fishes were sampled primarily by single-pass electro-fishing, although in deeper (>1 m) or shallower (<5 cm) habitats, respectively, seine or dip nets and visual surveys also were employed. A minimum section of 30 m stream length was sampled at all sites. This typically included at least two pool-riffle-run units. Sections of >30 m length were surveyed when few or no fishes were collected within the first 30 m. Habitats within a reach were sampled with equal effort, although those immediately adjacent to stream banks often received more intensive sampling as they typically provided the most heterogeneous habitat (see Leidy, 2007 for further details).

Site attributes

Information on a total of 18 site-specific variables was obtained for each site. These included measures of physical, biotic, and water quality conditions (Leidy, 2007). We focused on 10 of these variables (Table 1, Fig. 1), discarding others due to their high correlation with included variables (e.g., water vs. air temperature), or because they were not measured at all sites, or because no specific hypothesis regarding their effect on species occupancy was proposed. Each site also was subjectively rated on the extent to which human activities had visibly altered the form and physical structure of the stream channel, water quality, and the riparian habitat. More specifically, this rating included assessments of the intactness of the riparian habitat, levels of siltation, turbidity and apparent pollution, changes in substrate, and the degree of stream channelization (see Leidy & Fiedler, 1985 for details).

Species attributes

We assembled information on 10 different species attributes hypothesized to affect extinction vulnerability and invasion success from the literature and public databases (Moyle, 2002; http://www.fishbase.org). These included attributes indicative of their ecology, reproductive potential, and physiological tolerance, and whether or not a species was native or nonnative to the streams of the estuary (Table 1, Fig. 1).

Nested subset analysis

To estimate metacommunity nestedness, we used the metric based on paired overlap and decreasing fill, NODF (Almeida-Neto et al., 2008; Oksanen et al., 2010). This metric is calculated from the percentage of species occurring in sites having lower richness that overlap with the species occurring at sites having higher richness and the percentage of sites occupied by species occupying few sites that overlap with the sites occupied by species occupying many sites. Thus, unlike other metrics, nestedness may not only be calculated for the whole incidence matrix (NODF\(_\text{S}\)), but for species (NODF\(_\text{S}\)) and sites (NODF\(_\text{C}\)) individually as well. The NODF metric is also less sensitive to matrix size and shape, and less prone to Type I error, than other commonly used metrics (Almeida-Neto et al., 2009). Sites or species with equal species richness or occupancy do not contribute to the nestedness measured by NODF (Almeida-Neto et al., 2008). Based on the NODF metric, nestedness is maximized when the columns and rows of the incidence matrix are ordered by descending marginal totals (i.e., decreasing richness and occupancy).

We also estimated the degree to which species exhibited checkerboard patterns in their co-occurrence using the C-score metric (Stone & Roberts, 1990; Almeida-Neto et al., 2007; Oksanen et al., 2010) because some observed levels of nestedness were less than expected by chance (see Results). Checkerboard patterns occur when two or more species occupy mutually exclusive sites and are considered indicative of negative interactions (e.g., competition or predation) when sites are equally suitable for all species. The C-score is the average number of checkerboards exhibited across all species pairs and is invariant with respect to matrix sorting.

We used the null model algorithm based on constrained fixed marginal totals implemented in Almeida-Neto & Ulrich (2011) with a sequential swap algorithm to determine the probability that nestedness and checkerboard estimates could be obtained by chance (Gotelli, 2000; Ulrich et al., 2009). Simulations have suggested that this is the best-performing, although most conservative, of the available algorithms (Ulrich & Gotelli, 2007; Ulrich et al., 2009). We used 100 000 swaps to minimize Type I and Type II error rates (Fayle & Mancina, 2010) and estimated probabilities using 10 000 simulated matrices (Almeida-Neto & Ulrich, 2011).

Permutation gradient analysis

Our method for inferring the effects of species and site attributes on the nestedness of the metacommunity is founded on a permutation approach (Appendix S3). Let \( R \) denote the matrix of all species attributes and \( C \) the matrix of all site attributes. The rows of \( R \) are the species and those of \( C \) the sites, with their columns containing the respective attributes. The rows (\( r \), species) and columns (\( c \), sites) of the metacommunity incidence matrix are then randomly shuffled and the rows of \( R \) and \( C \) ordered correspondingly. This is repeated a large number of times (10 000 permutations). After each shuffle, we estimated \( \text{NODF}_R \), \( \text{NODF}_C \), \( \text{NODF}_s \), \( \text{NODF}_c \) of the incidence matrix, as well as the rank order correlation of each attribute (as ordered in \( R \) or \( C \)) with a number sequence of length equal to the number of species (or sites) in descending order. Rows and columns may be shuffled simultaneously because shuffling columns (or rows) has no effect on species (or site) nestedness. With a sufficient number of permutations, the resulting distributions of attribute-specific rank order correlation coefficients becomes normally distributed regardless of the data-types or distributions of the original attribute values by virtue of the central limit theorem.

The relative contributions (\( b \)) of species and site nestedness to the nestedness of the overall metacommunity may be esti-
mated from the resulting permutations by multiple regression
\( \text{NODF}_M \sim 0 + \beta_R \cdot \text{NODF}_R + \beta_C \cdot \text{NODF}_C \). This is equivalent to
\[
\beta_i = \frac{i(i-1)}{c(c-1) + r(r-1)},
\]
where \( r \) is the number of species, \( c \) is the number of sites, and \( i \) is either \( r \) or \( c \) for estimating \( \beta_R \) or \( \beta_C \), respectively.

Under an assumption of independence among attributes the effect of each species- (or site-) attribute on species (or site) nestedness may be estimated from the permutations by standard multiple regression (e.g., \( \text{NODF}_R \sim \beta_0 + \sum \beta_i x_i \)). This assumption was not defensible for our dataset as many attributes were expected to have causal relationships (e.g., body size, diet, and fecundity). We therefore used structural equation models (path analysis) to explicitly account for these dependencies and thereby tease apart from the total effect the direct and indirect effects that each attribute had on nestedness (Fig. 1; Rosseel, 2010). Applied to the permutations, the resulting unstandardized direct effect coefficients (\( \beta_D \)) represent the maximum change in species (site) nestedness expected when the species (sites) of the incidence matrix are ordered by a focal attribute if there were no ties and the effects of all other attributes are removed. More specifically, they represent the change in nestedness expected when an attribute is reordered from being randomly ordered (\( \rho = 0 \)) to fully ordered, with all other attributes held constant. Total effect coefficient (\( \beta_T \)) represents the change in nestedness expected when all attributes, excluding those downstream of the focal attribute’s causal pathway, are held constant. For instance, positive coefficients reflect increases in nestedness when an attribute is ordered in descending order.

Native vs. nonnative gradient analysis
We repeated all analyses for native and nonnative species separately. To facilitate the comparison of checkerboard patterns exhibited by native and nonnative species, we standardized their C-scores by the number of sites each species-pair occupied,
\[
\text{std. C-score} = \frac{\sum_{ij} ((c_i - c_{ij})(c_j - c_{ij}))/c_{ij} + (c_i - c_{ij})}{r(r-1)/2},
\]
where \( c_i \) and \( c_j \) are the number of species \( i \) and \( j \)’s occurrences and \( c_{ij} \) is their number of co-occurrences. For the gradient analysis, we removed native status from the list of potential species attributes, and added the richness of natives (nonnatives) to the site attributes in the analysis of nonnative (native) species. The addition of the other group’s richness as a putative explanatory variable allowed us to quantify the support for the biotic resistance and environmental acceptance hypotheses (Catford et al., 2009).

Results
A total of 33 species were identified at 256 sites. Eight surveyed sites at which no fishes were documented, repeat surveys of the same site, those with missing attributes, and those performed in saline sloughs were removed prior to analysis. An unidentifiable sunfish individual also was removed prior to analysis. The resulting incidence matrix was composed of 15 native species observed at 253 sites and 18 nonnative species observed at 71 sites (Appendix S2).
Species of the metacommunity exhibited a stronger degree of nestedness than expected by chance (Table 2; NODFR = 26.4, P < 0.001) indicating that infrequently observed species were present at a nonrandom subset of the sites occupied by species observed at many sites. In contrast, sites were less nested than expected by chance (NODFC = 42.4, P = 0.03). As a result of the large ratio of surveyed sites to the total number of observed species, site nestedness dominated the signal of the whole incidence matrix (βR = 0.016, βC = 0.984) which therefore also exhibited less nestedness than expected by chance (NODFM = 42.1, P = 0.04). Both species and sites exhibited significantly more checkerboarding than expected by chance (Table 2).

Our hypothesized causal model of species attributes explained 46.1% of the permutation-derived variation in species nestedness, whereas the model of site attributes explained 31.4% of the variation in site nestedness (Fig. 2). The attribute with the largest total effect on metacommunity nestedness was the distinction between native and nonnative species (βT = 9.8); that is, ordering the rows of the metacommunity with native species above nonnative species increased nestedness by 9.8 units. Ordering of species by their temperature tolerance from highest to lowest (βT = 7.9) and ordering of sites by their stream order from highest to lowest (βT = 6.7) had the next largest total effect. As a result of indirect effects, the total effects of stream order and of body size (βT = −4.2) were more than twice as large as their direct effects (βD = 2.8 and −1.8, respectively). The latter occurred because body size exhibited positive relationships with a species’ fecundity (βD = 0.59) and dominant

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Site- and species-specific attributes included in the analysis of fish metacommunities in streams tributary to the San Francisco Bay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute</td>
<td>Description</td>
</tr>
<tr>
<td><strong>Site (column) attributes</strong></td>
<td></td>
</tr>
<tr>
<td>Native/ nonnative richness</td>
<td>Continuous; number of observed species</td>
</tr>
<tr>
<td>Elevation</td>
<td>Continuous; from digitized USGS 7.5′ scale topographic maps (m)</td>
</tr>
<tr>
<td>Stream order</td>
<td>Ordinal</td>
</tr>
<tr>
<td>Disturbance</td>
<td>Ordinal; integrated visual rating of direct anthropogenic impact (0. Pristine–5. Channelized); see Leidy &amp; Fiedler (1985)</td>
</tr>
<tr>
<td>Stream width</td>
<td>Continuous; wetted channel (m)*</td>
</tr>
<tr>
<td>Stream depth</td>
<td>Continuous; mean water depth (cm)†</td>
</tr>
<tr>
<td>Dominant habitat</td>
<td>Nominal; 1. pool/2. pool-riffle/3. riffle **†‡</td>
</tr>
<tr>
<td>Flow rate</td>
<td>Continuous; discharge (cfs)</td>
</tr>
<tr>
<td>Fine sediment</td>
<td>Continuous; % substrate silt/clay/mud according to Wentworth particle-size scale*†</td>
</tr>
<tr>
<td>Riparian shade</td>
<td>Continuous; % wetted channel covered by a vertical projection of the riparian vegetation</td>
</tr>
<tr>
<td>Water temperature</td>
<td>Continuous (°C)</td>
</tr>
<tr>
<td>Conductivity</td>
<td>Continuous (μmho)</td>
</tr>
<tr>
<td><strong>Species (row) attributes</strong></td>
<td></td>
</tr>
<tr>
<td>Native status</td>
<td>Binary; native or nonnative to San Francisco Bay tributaries§</td>
</tr>
<tr>
<td>Body size</td>
<td>Continuous; maximum body length (cm)¶</td>
</tr>
<tr>
<td>Dominant prey</td>
<td>Ordinal; 1. vertebrates/2. vertebrates–invertebrates/3. invertebrates/4. invertebrates–algae/5. algae–detritus §</td>
</tr>
<tr>
<td>Habitat affinity</td>
<td>Nominal; primary water-column position: 1. benthic/2. benthic–suspension/3. suspension §</td>
</tr>
<tr>
<td>Fecundity</td>
<td>Continuous; maximum eggs/individual§</td>
</tr>
<tr>
<td>Growth rate</td>
<td>Continuous; maximum growth in first year (mm)§</td>
</tr>
<tr>
<td>Maturation rate</td>
<td>Continuous; minimum age at first reproduction (years)§</td>
</tr>
<tr>
<td>Temperature tolerance</td>
<td>Continuous; maximum habitat temperature (°C)§</td>
</tr>
<tr>
<td>Salinity tolerance</td>
<td>Continuous; maximum habitat salinity (ppt)§</td>
</tr>
</tbody>
</table>

*Mean of 3–5 transects placed perpendicular to stream flow.
†Mean of 9–15 point estimates taken equidistantly along replicate transects.
‡Estimated visually within the 1 m² quadrat surrounding each sampling point.
deviation, respectively, of 10,000 simulated matrices. Positive
ecies nestedness (NODF
matrix-wide nestedness than expected by chance
ering than expected by chance (Table 2). In con-
NODF
nestedness for native and nonnatives, respectively,
and nonnatives, respectively (Fig. 3). Species attributes
had considerably larger effects on native species than
on nonnatives, with the ordering of native species by
their temperature tolerance (βT = 19.1), fecundity
(βT = −15.3), body size (βT = −9.36), and time to hatch
(βT = 7.6) having the largest total effects. The total effect
of body size on native species nestedness was opposite
in sign to its direct effect (βD = 2.51) due to its indirect
effects via dominant prey and habitat affinity. The
ordering of sites by their stream order (βT = 7.8), con-
ductivity (βT = −4.9), and disturbance (βT = −4.8) had
the largest total effects on native nestedness. The ordering
of sites by conductivity (βT = 6.2) and disturbance
(βT = 4.1) also had large total effects on nonnative nest-
edness, but their effect was positive rather than nega-
tive. Stream depth had a large and positive total effect
on both natives (βT = 4.2) and nonnatives (βT = 4.8).
Richness of natives and nonnatives had only weak
effects on the other’s nestedness (βT = 0.2 and 1.7,
respectively). While few site attributes had their direct
effects on nestedness reversed via indirect effects, some
attributes (e.g., stream order and elevation) had total
effects that were almost twice as large, or twice as
small, as their direct effects.

Our results were robust to seasonal variation in
community structure and potential biases associated
with <100% detection probabilities (Appendix S4). The
quantitative importance of attributes as inferred by our

prey (βD = 0.02), which also had negative effects on spe-
cies nestedness (βD = −4.01 and −2.09, respectively).

The degree of species nestedness observed among
native species had a 16% probability of occurring by
chance alone (Table 2; NODF_R = 35.5, \( P = 0.16 \)).
Native-only sites, however, were significantly less
nested than expected by chance alone (NODF_C = 44.9,
\( P < 0.001 \)). Again the large ratio of occupied sites to the
total number of observed native species caused site
nestedness to dominate the signal of the whole inci-
dence matrix (βK = 0.003, βC = 0.995). Thus, the matrix
was also less nested than expected by chance
(NODF_M = 44.8, \( P < 0.001 \)). Both native species and the
sites they occupied exhibited significantly more check-
erboxing than expected by chance (Table 2). In con-
trast, nonnative species exhibited significantly stronger
matrix-wide nestedness than expected by chance
(Table 2; NODF_M = 25.5, \( P = 0.03 \)) because both species
(NODF_R = 24.7, \( P < 0.05 \)) and their occupied sites
(NODF_C = 25.5, \( P = 0.10 \)) exhibited nestedness
(βR = 0.058, βC = 0.942). Neither the nonnative species
nor their sites exhibited significant levels of checker-
boxing (Table 2). Nevertheless, when standardized
by the number of occupied sites, nonnative species
exhibited considerably more checkerboxing among
themselves than they did with native species and did
native species among themselves (Table 2).

Our hypothesized causal model of species attributes
explained 61.4% and 20.6% of the variation in species
nestedness for native and nonnatives, respectively,
whereas our model of site attributes explained 32.7%
and 40.1% of the variation in species nestedness for natives
and nonnatives, respectively (Fig. 3). Species attributes
had considerably larger effects on native species than
on nonnatives, with the ordering of native species by
their temperature tolerance (βT = 19.1), fecundity
(βT = −15.3), body size (βT = −9.36), and time to hatch
(βT = 7.6) having the largest total effects. The total effect
of body size on native species nestedness was opposite
in sign to its direct effect (βD = 2.51) due to its indirect
effects via dominant prey and habitat affinity. The
ordering of sites by their stream order (βT = 7.8), con-
ductivity (βT = −4.9), and disturbance (βT = −4.8) had
the largest total effects on native nestedness. The ordering
of sites by conductivity (βT = 6.2) and disturbance
(βT = 4.1) also had large total effects on nonnative nest-
edness, but their effect was positive rather than nega-
tive. Stream depth had a large and positive total effect
on both natives (βT = 4.2) and nonnatives (βT = 4.8).
Richness of natives and nonnatives had only weak
effects on the other’s nestedness (βT = 0.2 and 1.7,
respectively). While few site attributes had their direct
effects on nestedness reversed via indirect effects, some
attributes (e.g., stream order and elevation) had total
effects that were almost twice as large, or twice as
small, as their direct effects.

Our results were robust to seasonal variation in
community structure and potential biases associated
with <100% detection probabilities (Appendix S4). The
quantitative importance of attributes as inferred by our

Table 2  Descriptive and null-model-derived statistics of the nested and checkerboard patterns exhibited by freshwater fishes in
streams tributary to the San Francisco Bay

<table>
<thead>
<tr>
<th></th>
<th>Fill Contribution</th>
<th>NODF</th>
<th>P-value</th>
<th>Z-score</th>
<th>C-score</th>
<th>P-value</th>
<th>Z-score</th>
<th>Standardized C-score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All species</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matrix</td>
<td>0.099</td>
<td>–</td>
<td>42.14</td>
<td>0.04</td>
<td>–1.80</td>
<td>130.536</td>
<td>5 7.40</td>
<td>37.70</td>
</tr>
<tr>
<td>Sites</td>
<td>–</td>
<td>0.984</td>
<td>42.40</td>
<td>0.03</td>
<td>–1.86</td>
<td>247.23</td>
<td>&lt;0.001</td>
<td>13.70</td>
</tr>
<tr>
<td>Species</td>
<td>–</td>
<td>0.016</td>
<td>26.44</td>
<td>&lt;0.001</td>
<td>3.51</td>
<td>4.00</td>
<td>&lt;0.001</td>
<td>14.07</td>
</tr>
<tr>
<td><strong>Natives only</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matrix</td>
<td>0.186</td>
<td>–</td>
<td>44.84</td>
<td>&lt;0.001</td>
<td>–3.61</td>
<td>73.324</td>
<td>5 7.40</td>
<td>139.50</td>
</tr>
<tr>
<td>Sites</td>
<td>–</td>
<td>0.997</td>
<td>44.87</td>
<td>&lt;0.001</td>
<td>–3.64</td>
<td>698.32</td>
<td>&lt;0.001</td>
<td>8.08</td>
</tr>
<tr>
<td>Species</td>
<td>–</td>
<td>0.003</td>
<td>35.47</td>
<td>0.16</td>
<td>1.00</td>
<td>2.30</td>
<td>&lt;0.001</td>
<td>7.57</td>
</tr>
<tr>
<td><strong>Nonnatives only</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matrix</td>
<td>0.103</td>
<td>–</td>
<td>25.48</td>
<td>0.03</td>
<td>1.47</td>
<td>4481.5</td>
<td>5 7.40</td>
<td>7.57</td>
</tr>
<tr>
<td>Sites</td>
<td>–</td>
<td>0.942</td>
<td>25.52</td>
<td>0.10</td>
<td>1.05</td>
<td>29.29</td>
<td>0.19</td>
<td>0.90</td>
</tr>
<tr>
<td>Species</td>
<td>–</td>
<td>0.058</td>
<td>24.74</td>
<td>0.05</td>
<td>1.64</td>
<td>1.80</td>
<td>0.18</td>
<td>0.89</td>
</tr>
</tbody>
</table>

*The proportion of whole matrix nestedness (NODF_M) explained by nestedness among sites (NODF_C) and species (NODF_R).
\( Z\text{-score} = (x - \bar{x}_\text{simul})/\sigma_\text{simul} \), where \( x \) is the observed NODF or C-score value, and \( \bar{x}_\text{simul} \) and \( \sigma_\text{simul} \) are the mean and standard deviation, respectively, of 10,000 simulated matrices. Positive \( Z\text{-scores} \) indicate higher \( x \) than expected by chance, whereas negative \( Z\text{-scores} \) indicate lower \( x \) than expected by chance.
†Site C-scores calculated on the transposed incidence matrix.
§C-score standardized by sites occupied (see main text).
¶Checkerboard units; the total number of 2-by-2 checkerboards occurring in the matrix (Gotelli, 2000).
approach corresponded little with those of standard univariate analyses (Appendix S5)

Discussion

Attributes of extinction and invasion risk

Our application of the permutation-based gradient analysis to the fish communities of streams tributary to the San Francisco Bay underscores how an integrated view of community structure, and by inference of species extinctions (community disassembly) and invasions (community assembly), is necessary. Of all the hypotheses reviewed by Catford et al. (2009) and Olden et al. (2010), no single hypothesis was consistent with all the important drivers of extinction or invasion risk inferred by our analysis (Fig. 3).

For example, supporting the 'biotic acceptance' hypothesis that posits that the same environmental factors that reduce native vulnerability promote nonnative invasion, elevation and stream depth both had similarly strong associations with natives and nonnatives. This is to say that lower elevation and deeper stream reaches corresponded to both decreased native vulnerability and increased nonnative invasion risk, given our hypothesized causal model. However, while high order streams were also associated with reduced native vulnerability and increased nonnative invasion risk, given our hypothesized causal model. However, while high order streams were also associated with reduced native vulnerability and increased nonnative invasion risk, given our hypothesized causal model. However, while high order streams were also associated with reduced native vulnerability and increased nonnative invasion risk, given our hypothesized causal model.
nonnatives nestedness were of opposite sign – more disturbed reaches and those with higher conductivity were associated with increased native vulnerability and increased nonnative invasion risk, and those with higher temperatures were associated with decreased native vulnerability and decreased nonnative invasion risk.

Patterns of site nestedness suggest no support for the ‘biotic resistance’ hypothesis which posits that communities with many native species show reduced risk of nonnative invasion. Indeed the opposite was true, with sites having higher native richness showing increased nonnative nestedness than sites with lower native richness, consistent with other studies demonstrating positive spatial correlations in the richness of native and nonnative fishes (Marchetti et al., 2004a). The observation that standardized checkerboard scores were lower between natives and nonnatives than within nonnatives themselves (Table 2, Fig. S3) similarly suggests a relatively low influence of negative species interactions (Baltz & Moyle, 1993; Moyle & Light, 1996b). Among the species’ attributes contributing to metacommunity patterns, fecundity and temperature tolerance exhibited the strongest associations with natives (Fig. 3a). Higher temperature tolerance thus corresponded to decreased native vulnerability, consistent with the effect of stream temperature on their site nestedness. Temperature tolerance was not an important attribute for nonnative species.

The weak effect of fecundity on nonnative species suggests little support for the ‘propagule pressure’ hypothesis (Catford et al., 2009). In fact, despite fecundity’s large inferred effect on natives, this hypothesis’ underlying principle – which posits that a high supply or frequency of offspring increases invasion success – also was not supported for native species. Higher fecundity was associated with an increase in the inferred vulnerability of natives, presumably due to an unconsidered correlated trait. The inferred strength of fecundity’s negative effect on vulnerability affected a net reversal of body size’s direct effect to decrease vulnerability. The strong positive association of body size with fecundity meant that its indirect effect on nestedness was negative and stronger than its positive direct effect. Only because of this indirect effect was large body size inferred to exhibit a positive association with native vulnerability, as is typically observed (Reynolds et al., 2005).

More generally, given our assumed causal model of between-attribute relationships (Fig. 1), the power of species attributes to explain the metacommunity structure of native and nonnative species differed remarkably. Nonnative metacommunity structure was weak (Table 2) and far less well explained by their species attributes than was the structure of natives (Fig. 3a). This observation is consistent with the idiosyncratic nature of characteristics associated with successfully invading nonnative fishes in California (Marchetti et al., 2004a,c) and around the globe (Blanchet et al., 2009), and suggests that nonnative life-histories may play less of a role in their success than do other factors influencing the choices of humans in their selective introduction of fishes (see also Dill & Cordone, 1997; Moyle, 2002). The high standardized checkerboard scores observed among nonnatives (Table 2, Fig. S3) may similarly have more to do with the consequences of these human-made choices than with the competitive interactions between species. Understanding the control and establishment of nonnative fishes will thus entail integrating sociology with the study of fish biology.

Implication for nestedness analysis

Our analyses highlight the necessity of distinguishing the species patterns of site occurrence from the site patterns of species richness in the study of metacommunity structure. The shape of the metacommunity (the number of species observed vs. the number of sites surveyed) exhibited strong control over the inferred contribution of species and sites to the nestedness of the entire metacommunity. For example, the strong nested pattern observed among species of the entire fish metacommunity would have been masked by the lower than expected degree of nestedness observed among sites due to the high ratio of sites to species in this dataset (Table 2). It is therefore unlikely that a gradient analysis utilizing a metric of matrix-wide metacommunity nestedness such as matrix temperature (Atmar & Patterson, 1993) would have discerned the importance of distinguishing native and nonnative species in explaining metacommunity structure (Fig. 2a and b). The importance of disturbance, conductivity, and days-to-hatch would thus have remained obscured by the counteracting sign of their effects on native and nonnative species (Figs 2 and 3). We suggest that future analyses of metacommunity structure continue to use measures of nestedness, such as the NODF metric, which can distinguish between the row- and column-nestedness of a metacommunity.

Assumptions and limitations

Our analysis did not account explicitly for potential autocorrelation in species and site attributes associated with their shared evolutionary histories or lack of spatial independence. It is possible, for example, that phylogenetically controlled traits not included in our analysis are driving the counter-intuitive effect of fecundity on native nestedness, although visual inspection of the data suggests that this is not the case. Furthermore, our use of the NODF metric affected a discounting of site pairs with matching species sets.
(Almeida-Neto et al., 2008). This served to reduce the influence of first-order spatial autocorrelation potentially introduced by the nonrandom sampling of the region’s environments. We therefore consider these potential sources of biased inference to be weak.

A further nontrivial limitation of our permutation-based approach to gradient analysis also remains. The interpretation of permutation-based attribute effects is less intuitive than the interpretation of rank order correlation coefficients between traits and the incidence matrix’s marginal totals (richness and occupancy). Both types of approaches rely on an assumption that species and site attributes exhibit monotonic relationships to nestedness in the case of the permutation-based approach, or to richness and occupancy in the case of simple correlative approach, although for either approach these relationships may be nonlinear. The counteracting benefits of the permutation-based approach, in addition to those described above, are that (i) traits are related to nestedness structure directly rather than via richness, occupancy, or the species order of the maximally packed matrix that are themselves only correlates of nestedness, (ii) it requires no post hoc adjustment to control for Type I errors introduced by multiple comparisons of the same data, and that (iii) putative explanatory traits need not be continuous variables and may exhibit a wide variety of nonnormal distributions. Correlations between nestedness and the ordering of each trait derived from the permutation-based approach are normally distributed regardless of the trait’s individual distribution and thereby satisfy the assumptions of subsequent parametric analysis (e.g., structural equation modeling) without additional transformation or the use of generalized linear models.

Ultimately, of course, the utility of nestedness theory rests on statistical correlations. Many underlying assumptions are often not met (Atmar & Patterson, 1993). Indeed, the most critical assumption unaddressed by our study is that spatial patterns of species occurrences do in fact reflect the intrinsically temporal processes of community reassembly. Analyses relying on snapshot surveys accumulated across sites may fail to reflect the true temporal dynamics of species or the traits to which they are compared (Donlan et al., 2005). As a consequence, the true underlying drivers and consequences of community change can remain obscured (White & Kerr, 2006). The robustness of the relationships between nestedness patterns and temporal rates of extinction and invasion has itself only recently seen some support in fish and plant communities (Taylor & Warren, 2001; Elmendorf & Harrison, 2009); additional temporal tests of nested subset theory are needed (Donlan et al., 2005). Procedures to incorporate temporal perspectives in the analysis of biogeography and macroecology are increasingly available, but their application remains largely rare due to limited temporal data (Kerr et al., 2007; Fisher et al., 2010). Thus, as in most nonexperimental studies, our inferred effects of species and site attributes on extinction and invasion risk should be cautiously treated as hypotheses with which to refine further investigations (Fleishman et al., 2007).

Conclusions

In their synthesis of biogeography’s relevance to the conservation of freshwater fishes, Olden et al. (2010) advised for the trait-based quantification of extinction and invasion risk and a recognition of the interactive effects of multiple stressors in freshwater ecosystems. Methods that advance our ability to address these needs are equally necessary in other ecosystems and taxonomic groups (Catford et al., 2009; Sutherland et al., 2009). Our permutation-based analyses allow such an integration of the multiple potential drivers of community structure and, by inference, community reassembly. By allowing the explicit consideration of causal collinearity among putative drivers, our method enables the partitioning of the direct and indirect pathways by which both species traits and environmental variables contribute to metacommunity structure. It also provides standardized inferences regarding each attribute’s influence on this structure. Such measures are needed to progress beyond the forecasting of best- and worst-case scenarios of biodiversity change and how they will affect the functioning of ecosystems (Zavaleta et al., 2009). Our analysis of San Francisco Bay freshwater stream fishes suggests that their communities are being taken apart and put back together with both shared and nonshared rules of disassembly and assembly. Future work is needed to determine how strongly the attributes inferred to confer native vulnerability and nonnative invasion risk correlate with the attributes that contribute to the varied ecosystem processes that fishes affect (Gross et al., 2005).

Acknowledgements

We thank P. Guimaraes Jr. and J. Byrnes for their insights, R. Thill and R. Terry for assistance in compiling the databases. P. L. Fiedler and three anonymous reviewers for comments on the manuscript, and the numerous individuals who assisted with the field surveys. MN acknowledges the support of the USGS Western Ecological Research Center and the freshwater ecology research group at UCSC.

References


Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Effects of ties on the inference of attribute importance.

Appendix S2. Site locations and species-specific site-occupancies.

Appendix S3. R-code and simple example of the permutation-based gradient analysis.

Appendix S4. Robustness of results to false absences and the structure of the summer-only metacommunity.

Appendix S5. Contrast to univariate correlation and partial-correlation approaches.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.
Appendix A. Effects of ties on the inference of attribute importance

We illustrate here how the occurrence of ties in the rank ordering of an attribute may introduce considerable error in the inferred explanatory power of an attribute to matrix nestedness when rank order correlations are used. Such ties are particularly common among species attributes such as trophic level, dominant prey or habitat affinity, or among site attributes such as habitat type or stream order (see Table 1).

The two components of the NODF metric of matrix nestedness, the nestedness of species (NODF_s) and the nestedness of sites (NODF_c), are maximized when rows and columns are ordered by their marginal totals, occupancy and richness respectively (Fig. A1.1a, Almeida-Neto et al., 2008). In gradient analysis, the maximum explanatory power of species (site) attributes is determined by the attribute most correlated (either positively or negatively) with occupancy (richness). That is, the species (site) attribute whose ordering maximizes row (column) nestedness is considered the dominant driver of the species (site) associated metacommunity structure.

However, in the presence of rank order ties in a focal attribute (e.g., trophic level), the ordering of secondary attributes (e.g., habitat affinity) also contributes to nestedness. Indeed, the inferred importance of a focal attribute may be increased or decreased when secondary attributes are strongly correlated with marginal totals (Fig. A1.1b-c), or altered arbitrarily when secondary attributes uncorrelated with marginal totals, or also contain ties (Fig. A1.1d). The permutation method described in the main text resolves this issue by considering a large number of random orderings of rows and columns (or all possible orderings given a sufficiently small matrix).
Figure A.1. (a) Matrix nestedness is maximized by ordering rows and columns by their marginal totals. In a one-dimensional univariate gradient analysis, the site (or species) attributes most correlated with these marginal totals are inferred to be the dominant drivers of nestedness. However, when attributes exhibit ties in the rank order (e.g., trophic level), their inferred importance may effectively be (b) maximized, (c) minimized, or (d) changed arbitrarily by the appropriate ordering of secondary attributes (e.g., habitat affinity). Without explicit consideration of such secondary orderings, the true importance of focal attributes remains obscured.
Appendix B. Site locations and species-specific site-occupancies.

Figure B.1. Approximate locations of the 275 sites surveyed in the watersheds of the San Francisco Bay. See Leidy (2007) for further details.
Table B.1. Frequency by which native and non-native species were observed across the 256 sites included in the analysis, as well as species-specific mean paired nestedness values (see Almeido-Neto et al., 2008) and their associated probabilities for the All species, Native only, and Non-native only incidence matrices.

<table>
<thead>
<tr>
<th>Species</th>
<th>Status</th>
<th>Occupancy</th>
<th>All NODF</th>
<th>P</th>
<th>Native NODF</th>
<th>P</th>
<th>Non-native NODF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lavinia symmetricus</td>
<td>Native</td>
<td>154</td>
<td>12.0</td>
<td>0.00</td>
<td>6.3</td>
<td>0.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Oncorhynchus mykiss</td>
<td>Native</td>
<td>131</td>
<td>65.1</td>
<td>0.00</td>
<td>49.6</td>
<td>0.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Catostomus occidentalis</td>
<td>Native</td>
<td>119</td>
<td>16.0</td>
<td>0.00</td>
<td>11.0</td>
<td>0.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gasterosteus aculeatus</td>
<td>Native</td>
<td>100</td>
<td>40.5</td>
<td>0.01</td>
<td>37.6</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cottus asper</td>
<td>Native</td>
<td>76</td>
<td>32.5</td>
<td>0.34</td>
<td>35.5</td>
<td>0.21</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cottus gulosus</td>
<td>Native</td>
<td>41</td>
<td>59.0</td>
<td>0.02</td>
<td>40.8</td>
<td>0.38</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ptychocheilus grandis</td>
<td>Native</td>
<td>40</td>
<td>20.0</td>
<td>0.00</td>
<td>15.0</td>
<td>0.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lavinia exilicauda</td>
<td>Native</td>
<td>23</td>
<td>39.2</td>
<td>0.42</td>
<td>41.4</td>
<td>0.47</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hysterocarpus traskii</td>
<td>Native</td>
<td>6</td>
<td>34.0</td>
<td>0.16</td>
<td>12.3</td>
<td>0.03</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Leptocottus armatus</td>
<td>Native</td>
<td>4</td>
<td>66.7</td>
<td>0.28</td>
<td>73.3</td>
<td>0.17</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lampetra tridentata</td>
<td>Native</td>
<td>3</td>
<td>12.7</td>
<td>0.00</td>
<td>45.7</td>
<td>0.48</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mylopharodon conocephalus</td>
<td>Native</td>
<td>3</td>
<td>25.7</td>
<td>0.09</td>
<td>70.7</td>
<td>0.23</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Orthodon microlepidotus</td>
<td>Native</td>
<td>3</td>
<td>19.4</td>
<td>0.05</td>
<td>0.0</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Oncorhynchus tshawytscha</td>
<td>Native</td>
<td>1</td>
<td>93.4</td>
<td>0.02</td>
<td>76.0</td>
<td>0.35</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pogonichthys macrolepidotus</td>
<td>Native</td>
<td>1</td>
<td>64.4</td>
<td>0.20</td>
<td>7.0</td>
<td>0.35</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lepomis cyanellus</td>
<td>Non-native</td>
<td>34</td>
<td>13.8</td>
<td>0.06</td>
<td>-</td>
<td>-</td>
<td>4.8</td>
<td>0.43</td>
</tr>
<tr>
<td>Gambusia affinis</td>
<td>Non-native</td>
<td>23</td>
<td>25.4</td>
<td>0.02</td>
<td>-</td>
<td>-</td>
<td>6.7</td>
<td>0.49</td>
</tr>
<tr>
<td>Cyprinus carpio</td>
<td>Non-native</td>
<td>12</td>
<td>17.2</td>
<td>0.10</td>
<td>-</td>
<td>-</td>
<td>8.3</td>
<td>0.49</td>
</tr>
<tr>
<td>Lepomis machrochirrus</td>
<td>Non-native</td>
<td>11</td>
<td>6.8</td>
<td>0.05</td>
<td>-</td>
<td>-</td>
<td>9.7</td>
<td>0.34</td>
</tr>
<tr>
<td>Menidia beryllina</td>
<td>Non-native</td>
<td>9</td>
<td>61.5</td>
<td>0.33</td>
<td>-</td>
<td>-</td>
<td>6.9</td>
<td>0.29</td>
</tr>
<tr>
<td>Lucania parva</td>
<td>Non-native</td>
<td>8</td>
<td>56.7</td>
<td>0.35</td>
<td>-</td>
<td>-</td>
<td>19.6</td>
<td>0.00</td>
</tr>
<tr>
<td>Micropterus salmoides</td>
<td>Non-native</td>
<td>7</td>
<td>0.0</td>
<td>0.06</td>
<td>-</td>
<td>-</td>
<td>12.0</td>
<td>0.23</td>
</tr>
<tr>
<td>Acanthogobius flavimanus</td>
<td>Non-native</td>
<td>5</td>
<td>0.0</td>
<td>0.07</td>
<td>-</td>
<td>-</td>
<td>4.2</td>
<td>0.15</td>
</tr>
<tr>
<td>Morone saxatilis</td>
<td>Non-native</td>
<td>5</td>
<td>33.0</td>
<td>0.45</td>
<td>-</td>
<td>-</td>
<td>6.4</td>
<td>0.34</td>
</tr>
<tr>
<td>Lepomis gibbosus</td>
<td>Non-native</td>
<td>4</td>
<td>30.5</td>
<td>0.08</td>
<td>-</td>
<td>-</td>
<td>7.3</td>
<td>0.42</td>
</tr>
<tr>
<td>Lepomis microleophus</td>
<td>Non-native</td>
<td>3</td>
<td>14.5</td>
<td>0.27</td>
<td>-</td>
<td>-</td>
<td>14.0</td>
<td>0.28</td>
</tr>
<tr>
<td>Micropterus dolomieu</td>
<td>Non-native</td>
<td>3</td>
<td>0.0</td>
<td>0.21</td>
<td>-</td>
<td>-</td>
<td>8.7</td>
<td>0.50</td>
</tr>
<tr>
<td>Ameiurus melas</td>
<td>Non-native</td>
<td>2</td>
<td>0.0</td>
<td>0.50</td>
<td>-</td>
<td>-</td>
<td>0.0</td>
<td>0.26</td>
</tr>
<tr>
<td>Carassius auratus</td>
<td>Non-native</td>
<td>2</td>
<td>0.0</td>
<td>0.50</td>
<td>-</td>
<td>-</td>
<td>8.0</td>
<td>0.50</td>
</tr>
<tr>
<td>Notemigonus crysoleucus</td>
<td>Non-native</td>
<td>1</td>
<td>0.0</td>
<td>0.42</td>
<td>-</td>
<td>-</td>
<td>0.0</td>
<td>0.50</td>
</tr>
<tr>
<td>Percina macrolepidra</td>
<td>Non-native</td>
<td>1</td>
<td>0.0</td>
<td>0.43</td>
<td>-</td>
<td>-</td>
<td>0.0</td>
<td>0.50</td>
</tr>
<tr>
<td>Pimephales promelas</td>
<td>Non-native</td>
<td>1</td>
<td>0.0</td>
<td>0.49</td>
<td>-</td>
<td>-</td>
<td>0.0</td>
<td>0.50</td>
</tr>
<tr>
<td>Tridentiger trignocephalus</td>
<td>Non-native</td>
<td>1</td>
<td>0.0</td>
<td>0.47</td>
<td>-</td>
<td>-</td>
<td>0.0</td>
<td>0.50</td>
</tr>
</tbody>
</table>
library(vegan) # required for NODF calculation

# Define functions

# Spearman correlation of observed order to sorted order
spear< function(x){ cor(seq(1,length(x)),order(x,decreasing=T),method='spearman') }

# NODF function from vegan with friendly output
NODF< function(mat){out<-(as.matrix(nestednodf(mat,order=FALSE)$statistic[c(3,2,1)])); colnames(out)<-c('NODF','NODF.row','NODF.col');out}

# Permutation function that spits out NODF and associated spearman correlations
rshuffle.spear< function(M,R,C,order=FALSE,printall=TRUE,reps=3){
  Out< dim(0)
  r< nrow(M)
  c< ncol(M)
  for (p in 1:reps){
    if(order==FALSE){
      temp.row.order< sample(r)
      temp.col.order< sample(c)
    }
    if(order==TRUE){
      temp.row.order< order(apply(M,1,sum),decreasing=TRUE)
      temp.col.order< order(apply(M,2,sum),decreasing=TRUE)
    }
    temp.M< M[temp.row.order,temp.col.order]
    temp.R< R[temp.row.order,]
    temp.C< C[,temp.col.order]
    NODF.out< NODF(temp.M)
    Out< rbind(Out,cbind(NODF.out, t(as.matrix(apply(temp.R,2,spear))), t(as.matrix(apply(temp.C,1,spear)))))
    Out< data.frame(Out)
    colnames(Out)[-c(1:3)]<-(c(colnames(R),rownames(C))
    temp.M< cbind(temp.M,apply(temp.M,1,sum))
    temp.M< rbind(temp.M,apply(temp.M,2,sum))
    if(printall==TRUE){print(temp.M); print(temp.R); print(temp.C)}
  }
  return(Out)
}

# Make up some data & analyze it
m< n< 8
fill< 0.46
M< matrix(rbinom(n*m,1,fill),nrow=n,byrow=T)
M< M[order(apply(M,1,sum),decreasing=TRUE),order(apply(M,2,sum),decreasing=TRUE)]

R< matrix(c(sample(seq(0.1,n/10,0.1)),sample(seq(1,n,1)),sample(seq(10,n*10,10))),nrow=n,byrow=F)
C< matrix(c(sample(seq(0.1,m/10,0.1)),sample(seq(1,m,1)),sample(seq(10,m*10,10))),ncol=n,byrow=T)
colnames(R)< LETTERS[1:ncol(R)]
rownames(R)< paste('Sp',seq(1:n),sep='')
colnames(C)< letters[1:nrow(C)]
rownames(C)< paste('Site',seq(1:m),sep='')

rshuffle.spear(M,R,C,order=TRUE)
rshuffle.spear(M,R,C,order=FALSE,printall=FALSE)
### Maximum Nestedness

<table>
<thead>
<tr>
<th>NODF</th>
<th>NODFc</th>
<th>NODFs</th>
</tr>
</thead>
<tbody>
<tr>
<td>57.29</td>
<td>56.31</td>
<td>58.27</td>
</tr>
</tbody>
</table>

### Incidence Matrix

<table>
<thead>
<tr>
<th>Species</th>
<th>Sites</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>α</th>
<th>β</th>
<th>γ</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.7</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>B</td>
<td>0.4</td>
<td>0.6</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>C</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

### Rank Order Correlation with Θ

<table>
<thead>
<tr>
<th>Permutation</th>
<th>Incidence matrix</th>
<th>Rank order correlation with Θ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Species attributes</td>
<td>Site attributes</td>
</tr>
<tr>
<td></td>
<td>NODF</td>
<td>NODFc</td>
</tr>
<tr>
<td>1</td>
<td>1.00</td>
<td>0.95</td>
</tr>
<tr>
<td>2</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>3</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>4</td>
<td>0.97</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Θ represents a sequence \( \{1, 2, 3, \ldots, n\} \) where \( n \) is the number of species (r) or sites (c). In this example, \( n = r = c = 8 \).
Appendix D. Robustness of results to false absences and the structure of the summer-only metacommunity

As in most biodiversity surveys, it is unlikely that all species present at a site were detected. Indeed, we observed a weakly positive relationship between species richness and log-transformed survey area (section length x stream width, $R^2 = 0.16$). The false absences resulting from <100% detection probabilities can bias inferences of nestedness (Fleishman et al., 2007). To assess the sensitivity of our inferences of metacommunity patterns to incomplete detection we recalculated nestedness and checkerboard metrics after removing all species populations having densities lower than a particular threshold density. This threshold density was successively increased in a stepwise fashion from the lowest to the highest observed population densities. We thereby assumed that detection probabilities were log-linearly dependent on a species’ site-specific density ($m^2$). This approach avoided assumptions inherent in occupancy-modeling, did not require repeated sampling of the same site, and also provided insight into the contribution of rare (low density) versus abundant species to metacommunity patterns.

Patterns of nestedness and checkerboarding were largely insensitive to the effects of hypothetical false absences of species with population densities less than two orders of magnitude higher density than the lowest population density observed (Fig. D.1). However, analyses performed using surveys performed across fewer sites, on more cryptic taxa, or on species exhibiting less variation in densities, are likely to be more sensitive to such false absences in both their metric estimate and its inferred statistical significance. Among non-native fishes, for example, estimates of nestedness and their null model frequencies varied substantially between matrices that included only populations with densities $\geq 0.005 \text{ m}^{-2}$ and those that included only populations with densities $\geq 0.05\text{m}^{-2}$ (Fig. D.1b). Our analyses thus suggest that future studies of metacommunity structure consider more explicitly the sensitivity of their results to incomplete species detections and the effects of low density populations (see also Cam et al., 2000, Fleishman et al., 2007).
To further assess the robustness of our results, we also repeated the gradient analyses after restricting our data to include only those surveys representative of summer fish community structures (161 sites and 30 species, June-September, Moyle, 2002). These provided virtually equivalent results to our primary analysis (Fig. D.2).
Figure D.1. The sensitivity of inferred metacommunity patterns to <100% detection probabilities as assessed by recalculating nestedness and checkerboard metrics after removing populations having densities less than successively higher density thresholds. (a) Frequency distribution of native and non-native site-specific densities observed across all 256 surveys. Changes in (b) matrix nestedness, (c) standardized species C-score (see main text), and (d) the remaining species richness after the removal of populations with lower occupying density for all fishes (circles), natives only (squares), and non-natives only (diamonds). Filled symbols in (b) and (c) indicate estimates having a ≤ 10% probability of occurring in null model simulations.
Figure D.2. As in Figure 1 of the main text, but with analysis restricted to surveys conducted in June – September (161 sites and 30 species).
Appendix E. *Contrast to correlation and partial-correlation approach.*

To contrast our results with the inferences of a standard one-dimensional gradient analysis, we also calculated pairwise Spearman’s rank order correlations ($\rho_{xy}$) and partial Spearman’s rank order correlations ($\rho_{xy|z}$) between all 10 species attributes and their site occupancy, and between all 11 site attributes and their species richness.

Qualitatively, the permutation-based gradient analyses gave similar results to the method of using rank order and partial rank order correlation coefficients to assess the importance of species and site in explaining nestedness; the two methods agreed 60-80% of the time in comparing of the direction (sign) of their inferred attribute-specific effects. Nevertheless, the quantitative importance of attributes as inferred by our permutation-based gradient analyses corresponded little with their quantitative importance as inferred by the rank order and partial rank order correlation approach; Pearson’s correlation coefficients between the two ranged from $r = -0.33$ to 0.06.
Figure E.1. Spearman’s rank order correlations ($\rho_{xy}$) and partial rank order correlations ($\rho_{xy.z}$) of (a) species attributes and their site occupancy, and (b) site attributes and their species richness. *$p < 0.05$, **$p < 0.01$. 

13
14
15
16
17
18