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 of 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 ($NODF_R$) 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. A.1.b-c), or altered arbitrarily when secondary attributes uncorrelated with marginal totals, or also contain ties (Fig. A.1.d). 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).
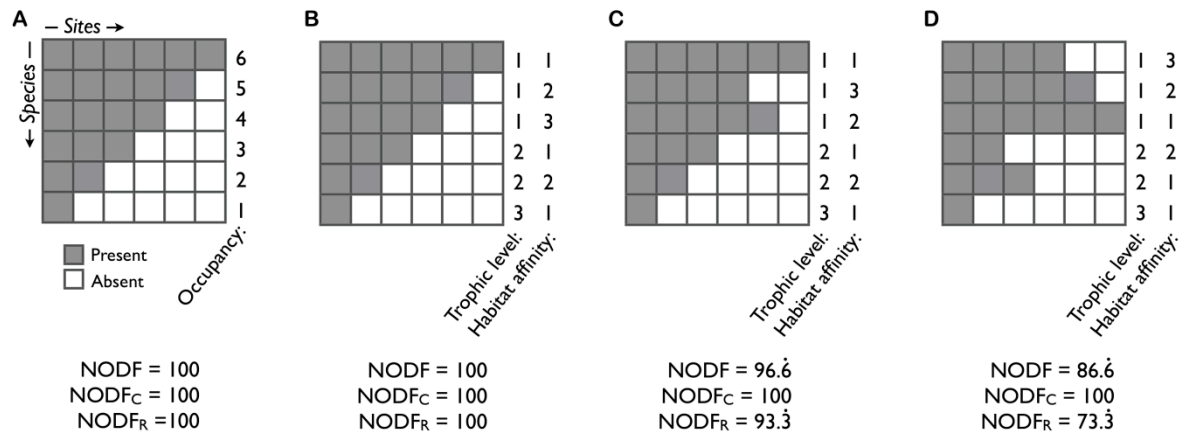
Figure A.1. (a) Matrix nestedness is maximized by ordering rows and columns by their marginal totals. In a one-dimensional univariate gradient analysis, the site (or species) attributes most correlated with these marginal totals are inferred to be the dominant drivers of nestedness. However, when attributes 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.
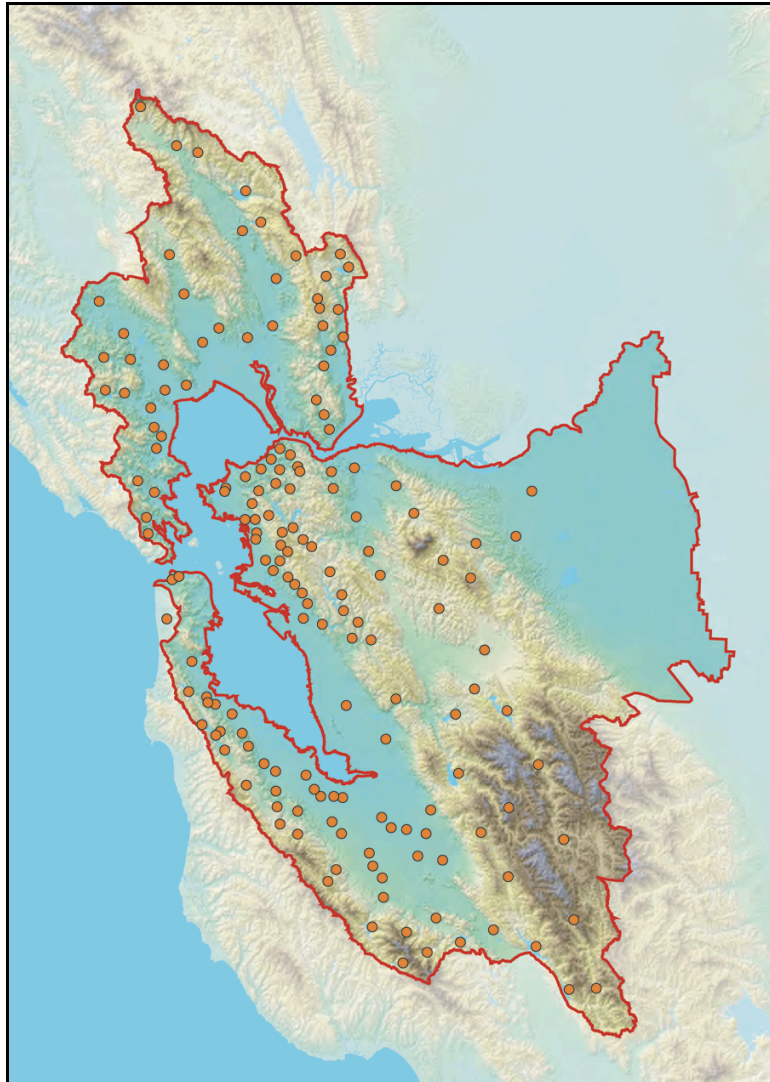
Figure B.1. Approximate locations of the 275 sites surveyed in the watersheds of the San Francisco Bay. See Leidy (2007) for further details.

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.

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

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

```
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<-t(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='')
rownames(C)<-letters[1:nrow(C)]
colnames(C)<-paste('Site',seq(1:m),sep='')

rshuffle.spear(M,R,C,order=TRUE)
rshuffle.spear(M,R,C,order=FALSE,printall=FALSE)
```
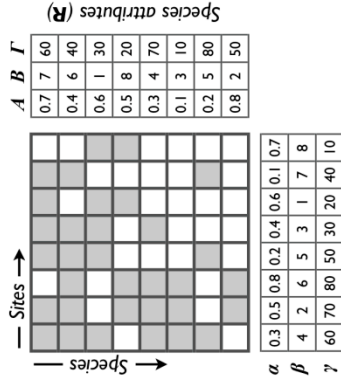
Rank order correlation with Θ

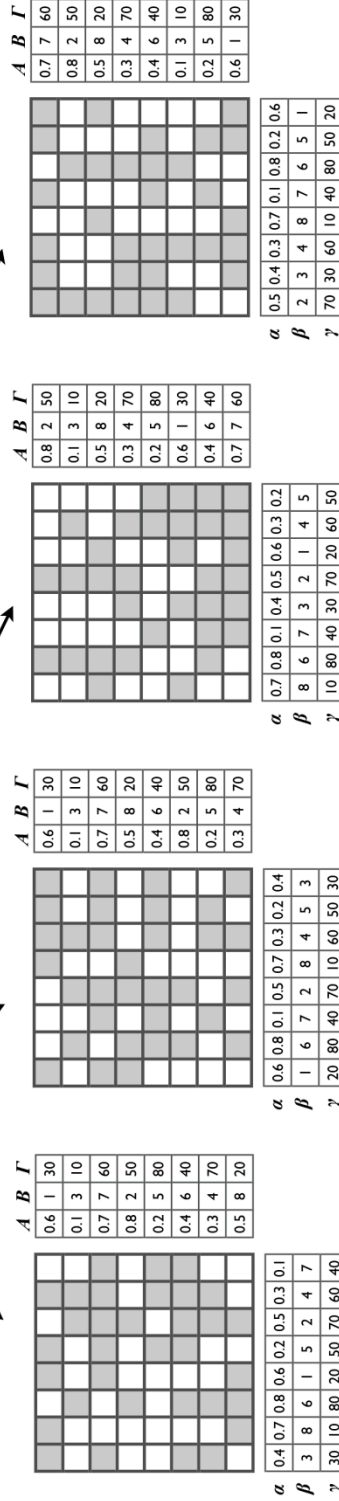| Permutation	Incidence matrix			Species attributes			Site attributes		
	NODF	NODF$_R$	NODF$_C$	A	B	Γ	α	β	γ
1	22.62	25.89	19.35	0.14	-0.64	-0.24	0.64	0.0	-0.38
2	28.72	34.82	22.62	0.37	-0.17	-0.74	0.43	-0.07	0.10
3	8.18	1.19	15.18	-0.12	-0.36	-0.31	0.52	0.60	-0.19
4	35.27	31.49	39.05	0.57	0.48	0.14	-0.07	-0.10	0.19
...	...	...	...	...	...	...	...	...	...
10,000	23.78	25.77	21.79	-0.21	-0.33	-0.07	-0.71	0.1	-0.26

Θ represents a sequence $\{1, 2, 3, ..., n\}$ where n is the number of species (r) or sites (c). In this example, $n = r = c = 8$.

Maximum Nestedness
NODF = 57.29
NODF$_C$ = 56.31
NODF$_R$ = 58.27

Present
Absent

Species attributes (R)

A B Γ

Sites
Species

Site attributes (C)

α
β
γ

Permutation 1
Permutation 2
Permutation 3
Permutation 4

NODF = 22.62
NODF$_C$ = 19.35
NODF$_R$ = 25.89

NODF = 28.72
NODF$_C$ = 22.62
NODF$_R$ = 34.82

NODF = 23.78
NODF$_C$ = 21.79
NODF$_R$ = 25.77

NODF = 35.27
NODF$_C$ = 39.05
NODF$_R$ = 31.49

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 ≥ 0.005 m^{-2} and those that included only populations with densities ≥ 0.05m^{-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).
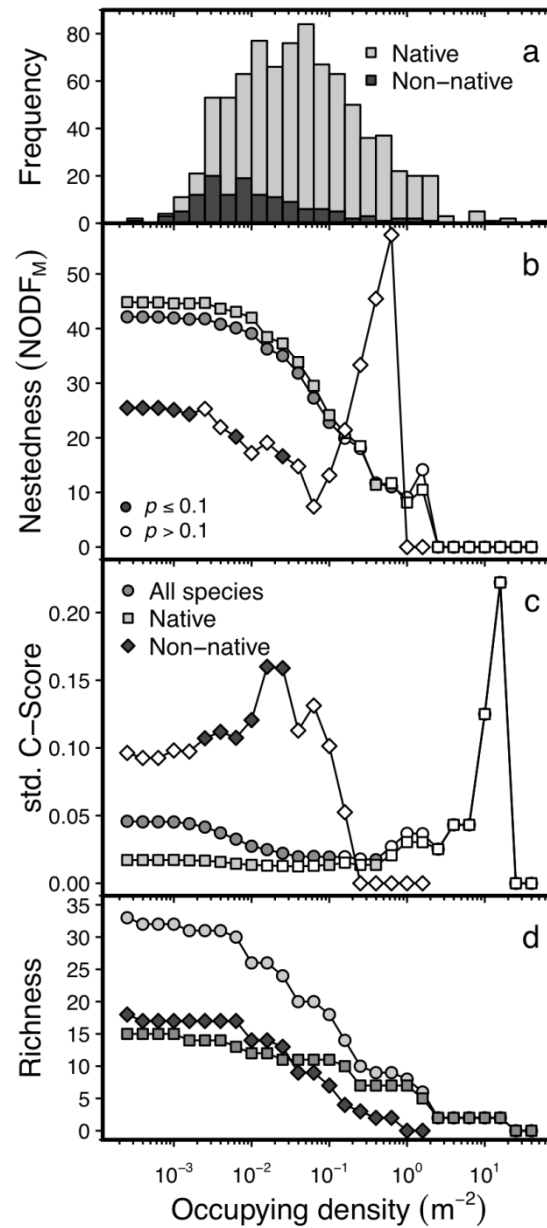
Figure D.1. The sensitivity of inferred metacommunity patterns to <100% detection probabilities as assessed by recalculating nestedness and checkerboard metrics after removing populations having densities less than successively higher density thresholds. (a) Frequency distribution of native and non-native 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.
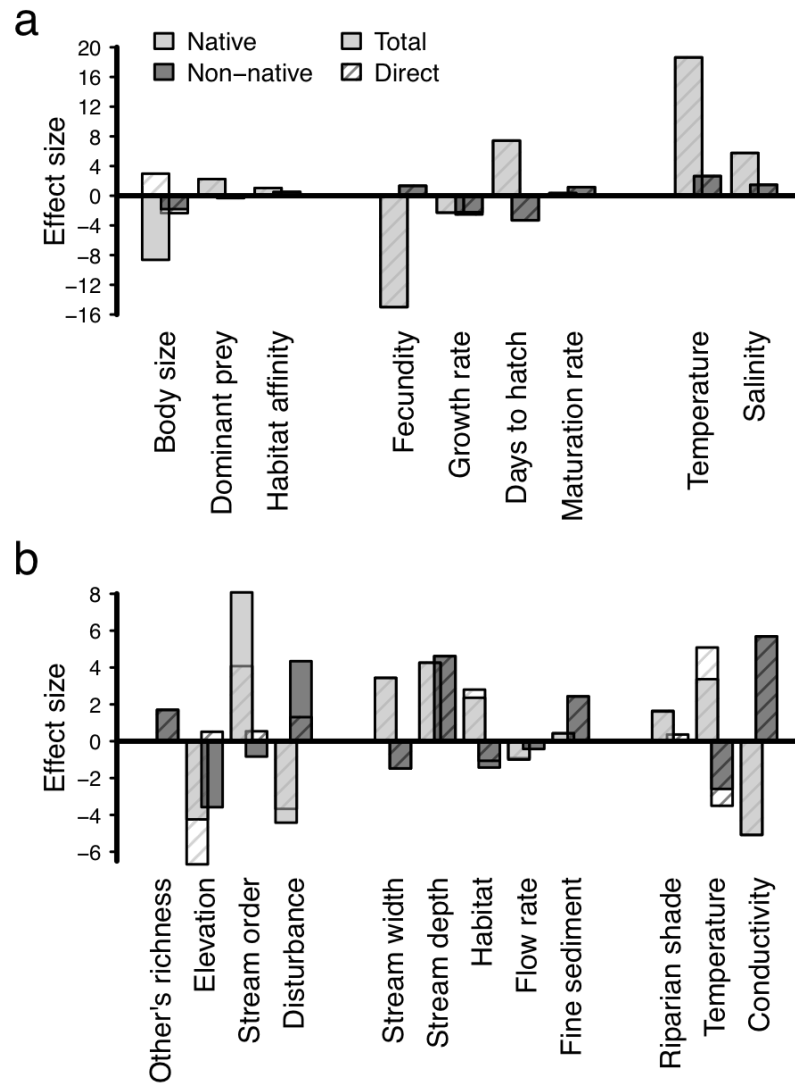
Figure D.2. As in Figure 1 of the main text, but with analysis restricted to surveys conducted in June – September (161 sites and 30 species).

1   **Appendix E.** *Contrast to correlation and partial-correlation approach.*

2   To contrast our results with the inferences of a standard one-dimensional gradient analysis, we also

3 calculated pairwise Spearman's rank order correlations (ρ_{XY}) and partial Spearman's rank order

4 correlations ($\rho_{XY \cdot Z}$) between all 10 species attributes and their site occupancy, and between all 11 site

5 attributes and their species richness.

6   Qualitatively, the permutation-based gradient analyses gave similar results to the method of using

7 rank order and partial rank order correlation coefficients to assess the importance of species and site in

8 explaining nestedness; the two methods agreed 60-80% of the time in comparing of the direction (sign)

9 of their inferred attribute-specific effects. Nevertheless, the quantitative importance of attributes as

10 inferred by our permutation-based gradient analyses corresponded little with their quantitative

11 importance as inferred by the rank order and partial rank order correlation approach; Pearson's

12 correlation coefficients between the two ranged from r = -0.33 to 0.06.

13
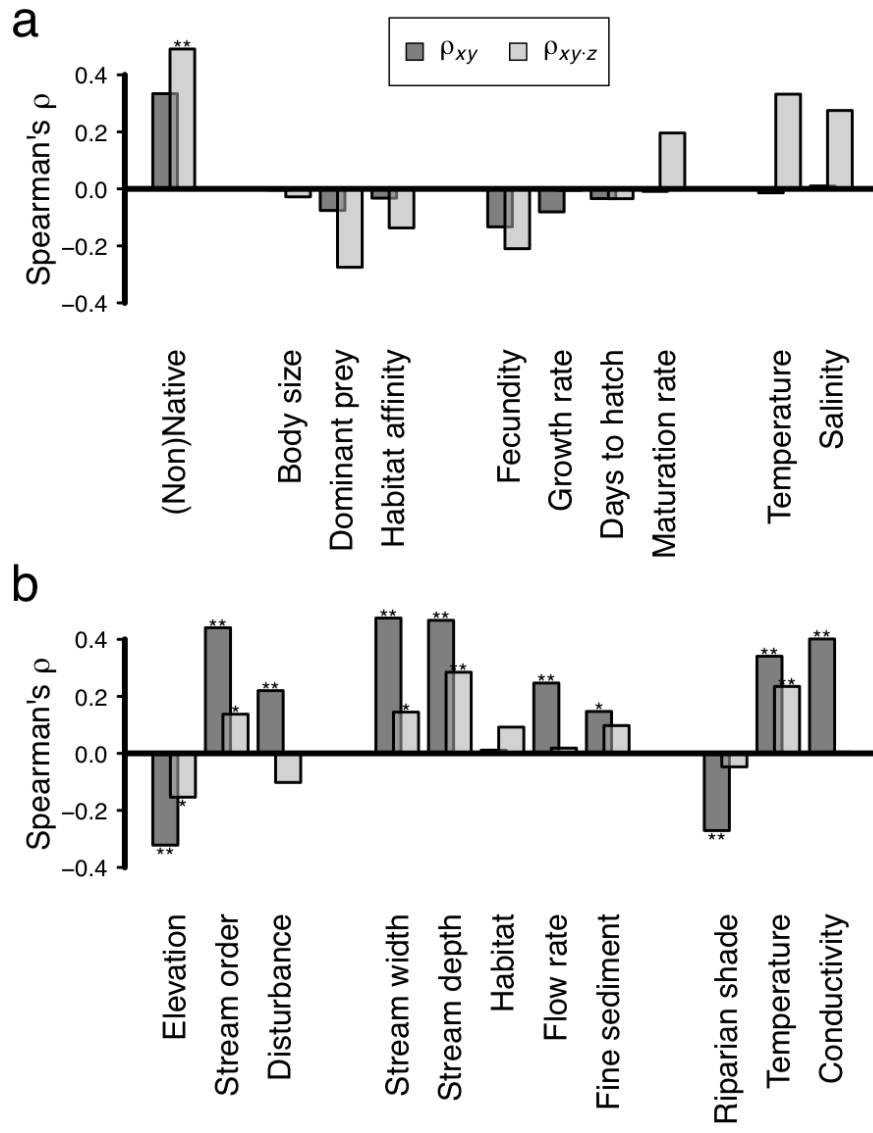
Figure E.1. Spearman's rank order correlations (ρ_{XY}) and partial rank order correlations ($\rho_{XY \cdot Z}$) of (a) species attributes and their site occupancy, and (b) site attributes and their species richness. $*p < 0.05$, $**p < 0.01$.

14

15

16

17

18