

# NRSA-CA-QC

## Analyses

### National Rivers and Streams Assessment-Community Assembly

Evaluating metacommunity structure and diversity as well as the drivers of invertebrate community assembly in rivers and streams across the conterminous United States. Metacommunities were defined as the nine ecoregions (coastal plain = CPL, northern Appalachians = NAP, northern plain = NPL, southern Appalachians = SAP, southern plain = SPL, temperate plain = TPL, upper midwest = UMW, western mountains = WMT, and xeric = XER) and the aggregation of all ecoregions as the conterminous United States (CUS) to identify: (1) the differences in community structure and diversity among ecoregions; (2) what drives community assembly at the (eco)regional level; and (3) if the identity and/or magnitude of drivers of community assembly vary among ecoregions.

Data and R code for the manuscript:

Murray-Stoker, D., and K. M. Murray-Stoker. Consistent metacommunity structure despite inconsistent drivers of assembly at the continental scale.

### Load Packages & Data

Data and R packages required for data management and analyses are loaded into the environment.

### Network Centrality

Calculate network centrality for each site. Site coordinates are subset for each ecoregion, and the subset dataframes are combined into a list. The list of dataframes is then supplied to the network centrality function, and all measures of network centrality are appended into a single dataframe. Network centrality is calculated as the mean geodesic distance between a site to all other sites in the ecoregion. Network centrality is a measure of how connected a site is to other sites, or how far organisms would have to disperse between or among sites.

### Data Management

Subset data to create site-by-taxa matrices as well as dataframes with taxa and environmental covariates for each site. The data are subset by ecoregion, and combined into a list. The list is then supplied to the `taxa.matrices()` function, which creates a site-by-taxa matrix for all taxa present in each ecoregion which is created into its own list.

### Site Scores & Reciprocal Averaging

Calculate site scores via reciprocal averaging. Reciprocal averaging (RA) arranges sites with more similar compositions closer together (site scores) and species with more similar distributions closer together (species scores). Site scores derived from RA are calculated for each site within an ecoregion for all nine ecoregions. Site scores will then be used as the response for determining which environmental, geographic, and network variables are most important or have the greatest influence on community assembly and composition.

## Elements of Metacommunity Structure (EMS) Analyses

Calculate the three elements of metacommunity structure (EMS) as defined by Leibold & Mikkelsen (2002): coherence, turnover, and boundary clumping. All three elements are calculated with the `Metacommunity()` function. The EMS for the entire CUS and each ecoregion are calculated (each ecoregion region serves as the regional metacommunity), and then the summary statistics are extracted and appended into a final summary dataframe.

- Definitions:
- Coherence: the number of embedded absences in the ordinated matrix (species absences occurring within surrounding presences).
- Turnover: how often species/taxa replace another between sites.
- Boundary Clumping: a measure of species/taxa dispersions among sites, or how clumped distributions are at edges/boundaries.

Statistical significance for coherence and turnover is calculated in relation to a null distribution on randomized taxa matrices. For the null randomization, use the conservative fixed-fixed null model (method = “swap”) whereby the number of occupied sites and number of species/taxa are held constant. Statistical significance for boundary clumping (Morisita’s index) is calculated relative to a chi-square distribution.

For a full description of the EMS, please see:

Leibold, M.A., and G. Mikkelsen. 2002. Coherence, species turnover, and boundary clumping: elements of meta-community structure. *Oikos* 97: 237–250.

Table 1: Coherence statistics for the CUS and each individual ecoregion.

coherence.embAbs	coherence.z	coherence.P.value	coherence.mean	coherence.variance	ecoregion
558586	-1.864	0.062	594133.36	19073.348	CUS
20113	-2.969	0.003	25983.07	1977.015	CPL
23614	-2.895	0.004	28717.37	1762.985	NAP
30010	-3.010	0.003	37062.42	2342.694	NPL
54976	-2.412	0.016	60597.58	2331.076	SAP
25218	-2.680	0.007	31045.84	2174.368	SPL
24468	-2.750	0.006	28361.35	1415.896	TPL
20681	-3.112	0.002	24883.14	1350.137	UMW
34521	-2.861	0.004	43777.03	3234.985	WMT
27216	-2.784	0.005	34737.53	2702.128	XER

Table 2: Turnover statistics for the CUS and each individual ecoregion.

turnover.Rep	turnover.z	turnover.P.value	turnover.mean	turnover.variance	ecoregion
6436119364	18.647	0	3479437990	158561810.4	CUS
11163352	6.664	0	8063703	465116.4	CPL
12340363	7.981	0	8489070	482536.8	NAP
19948363	10.493	0	12088430	749048.8	NPL
49970922	12.251	0	28575001	1746400.6	SAP
10419563	8.330	0	6717606	444408.6	SPL
9027572	4.809	0	6801773	462848.6	TPL
8969842	7.340	0	6330099	359629.5	UMW
32985416	11.293	0	20700971	1087759.2	WMT
15078364	7.364	0	10737531	589477.1	XER

Table 3: Boundary clumping statistics for the CUS and each individual ecoregion.

boundary.Index	boundary.P.value	boundary.df	ecoregion
2.710	0	1060	CUS
1.663	0	93	CPL
1.463	0	90	NAP
1.934	0	129	NPL
2.228	0	175	SAP
1.986	0	121	SPL
1.849	0	107	TPL
1.533	0	83	UMW
1.603	0	130	WMT
1.365	0	109	XER

## Additive & Multiplicative Partitioning of Diversity

Assess how taxonomic richness scales across spatial scales. Alpha ( $\alpha$ ) diversity is local richness, while beta ( $\beta$ ) diversity is the turnover in diversity from alpha (local,  $\alpha$ ) to gamma (regional,  $\gamma$ ). Additive partitioning scales diversity across samples in a hierarchical sampling design. For this study,  $\alpha$  = site,  $\beta_1$  = site-to-ecoregion, and  $\beta_2$  = ecoregion-to-CUS.

For a full description of additive partitioning, please see:

Crist, T.O., J.A. Veech, J.C. Gering, and K.S. Summerville. 2003. Partitioning species diversity across landscapes and regions: a hierarchical analysis of  $\alpha$ ,  $\beta$ , and  $\gamma$  diversity. *The American Naturalist* 162: 734–743.

Evaluating how the observed and expected taxonomic richness varies across hierarchical scales.

Table 4: Results of the additive partitioning of diversity analysis.

	observed	simulated	P.value
alpha.1	0.086	0.244	0.001
alpha.2	0.316	0.707	0.001
gamma	1.000	1.000	1.000
beta.1	0.231	0.463	0.001
beta.2	0.684	0.293	0.001

Table 5: Results of the multiplicative partitioning of diversity analysis.

	observed	simulated	P.value
alpha.1	15.944	68.499	0.001
alpha.2	42.864	101.531	0.001
gamma	106.981	106.981	1.000
beta.1	6.710	1.562	0.001
beta.2	2.496	1.054	0.001

## Taxonomic Richness & Diversity

Compare taxonomic richness and diversity among ecoregions. Taxonomic richness is quantified as the number of taxa present at each site and then compared using a one-way analysis of variance (ANOVA) with Type I sums of squares. Taxonomic diversity is quantified using Shannon's diversity index and then compared using a one-way ANOVA with Type I sums of squares, followed by post-hoc Tukey's HSD. Model assumptions are inspected numerically and graphically.

### Taxonomic Richness

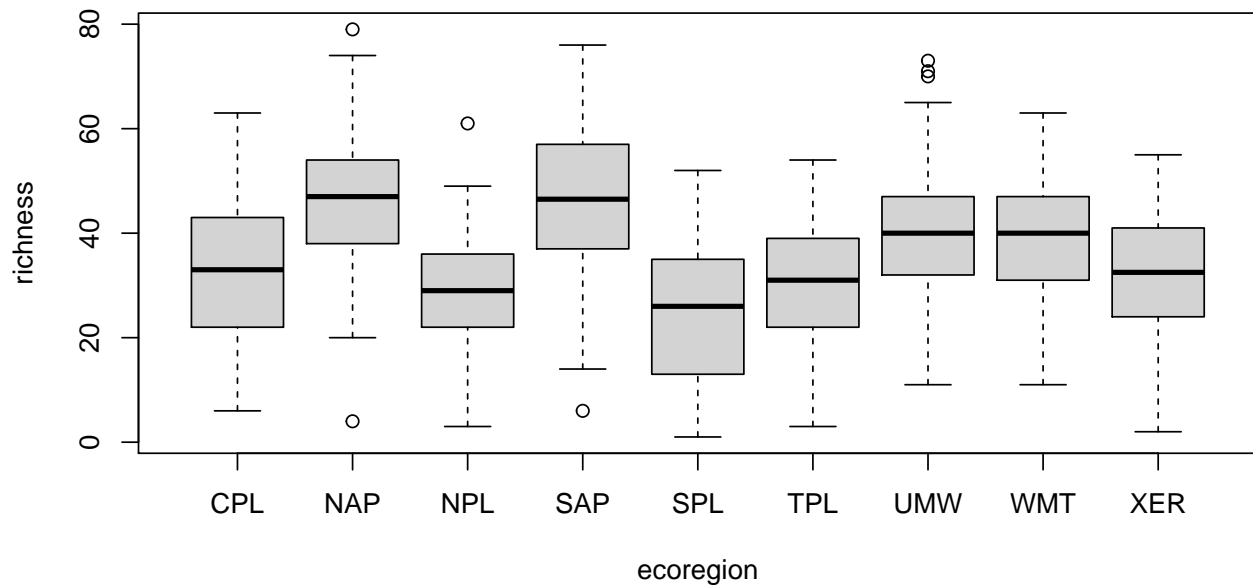


Figure 1: Boxplot of taxonomic richness by ecoregion.

Table 6: Results of the taxonomic richness ANOVA.

term	df	sumsq	meansq	statistic	p.value
ecoregion	8	59276.08	7409.510	48.716	0
Residuals	1055	160461.87	152.097	NA	NA

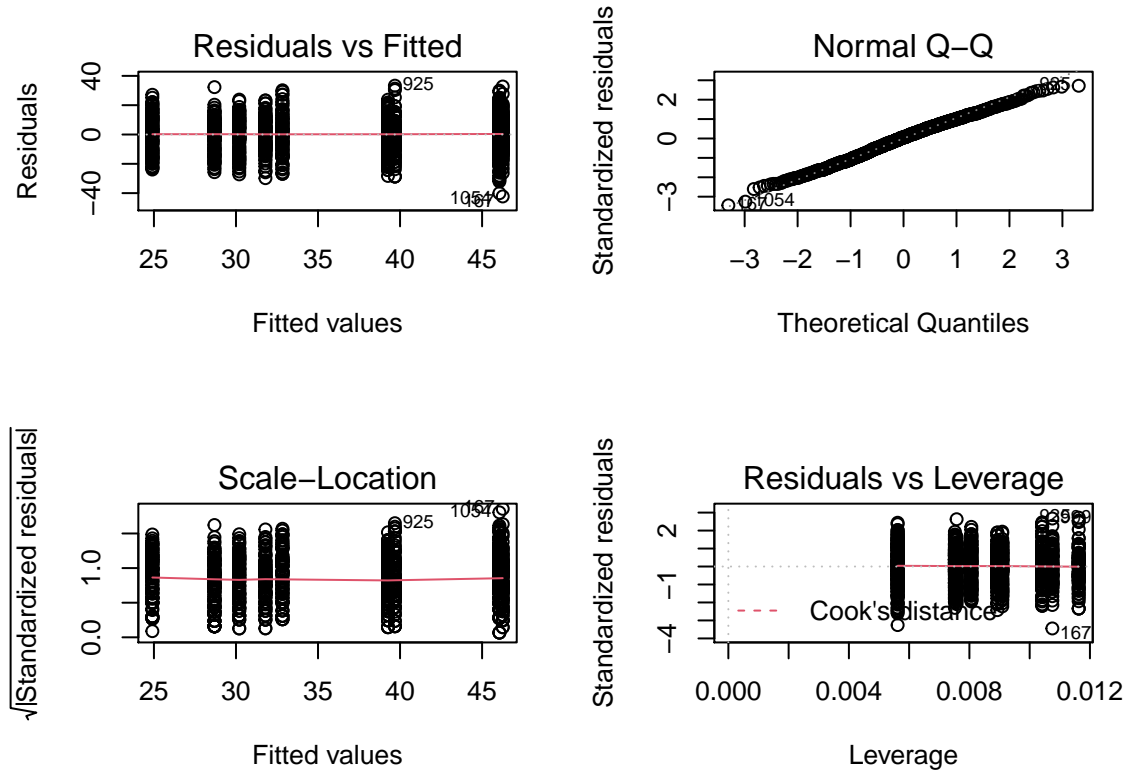


Figure 2: Checking taxonomic richness ANOVA assumptions using diagnostic plots.

Table 7: Groupings assigned to ecoregions differing in taxonomic richness using Tukey's HSD.

	richness	groups
NAP	46.258	a
SAP	46.056	a
UMW	39.686	b
WMT	39.263	b
CPL	32.833	c
XER	31.795	c
TPL	30.200	c
NPL	28.689	cd
SPL	24.903	d

Table 8: Summary statistics of taxonomic richness among ecoregions.

ecoregion	N	mean	sd	se	LL	UL
CPL	96	32.833	13.681	1.396	7.750	60.000
NAP	93	46.258	12.116	1.256	23.900	66.700
NPL	132	28.689	10.596	0.922	8.275	45.000
SAP	178	46.056	13.784	1.033	17.425	69.575
SPL	124	24.903	12.983	1.166	2.075	46.925
TPL	110	30.200	11.759	1.121	7.725	47.550
UMW	86	39.686	12.869	1.388	11.625	69.375
WMT	133	39.263	11.346	0.984	17.000	59.700
XER	112	31.795	11.216	1.060	7.000	53.000

## Taxonomic Diversity

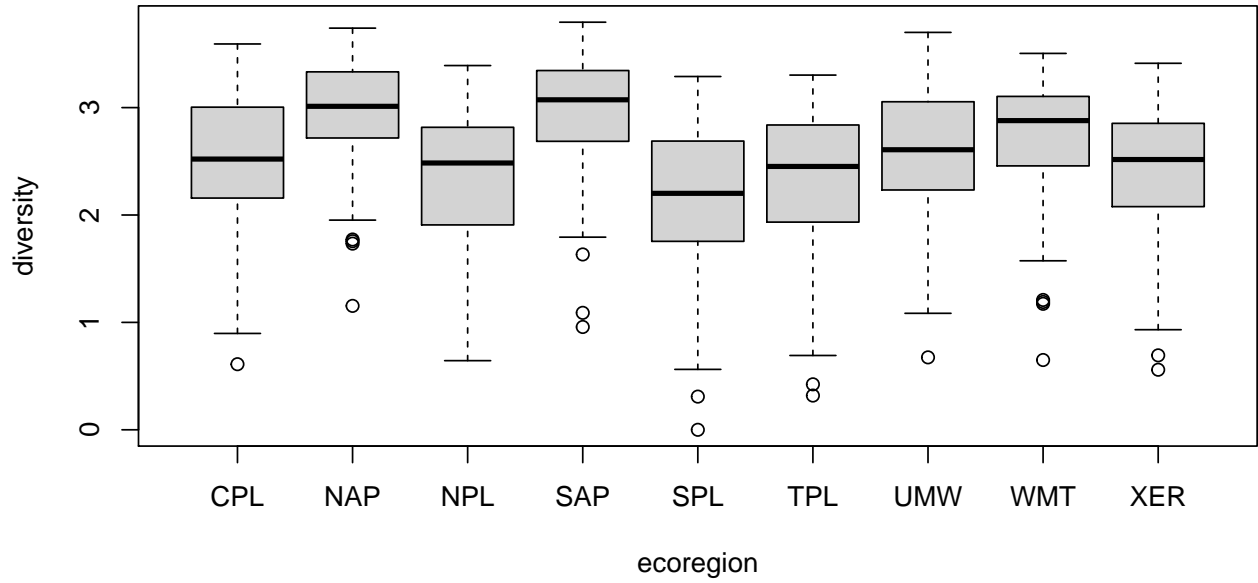


Figure 3: Boxplot of taxonomic richness by ecoregion.

Table 9: Results of the taxonomic diversity ANOVA.

term	df	sumsq	meansq	statistic	p.value
ecoregion	8	85.891	10.736	30.551	0
Residuals	1055	370.756	0.351	NA	NA



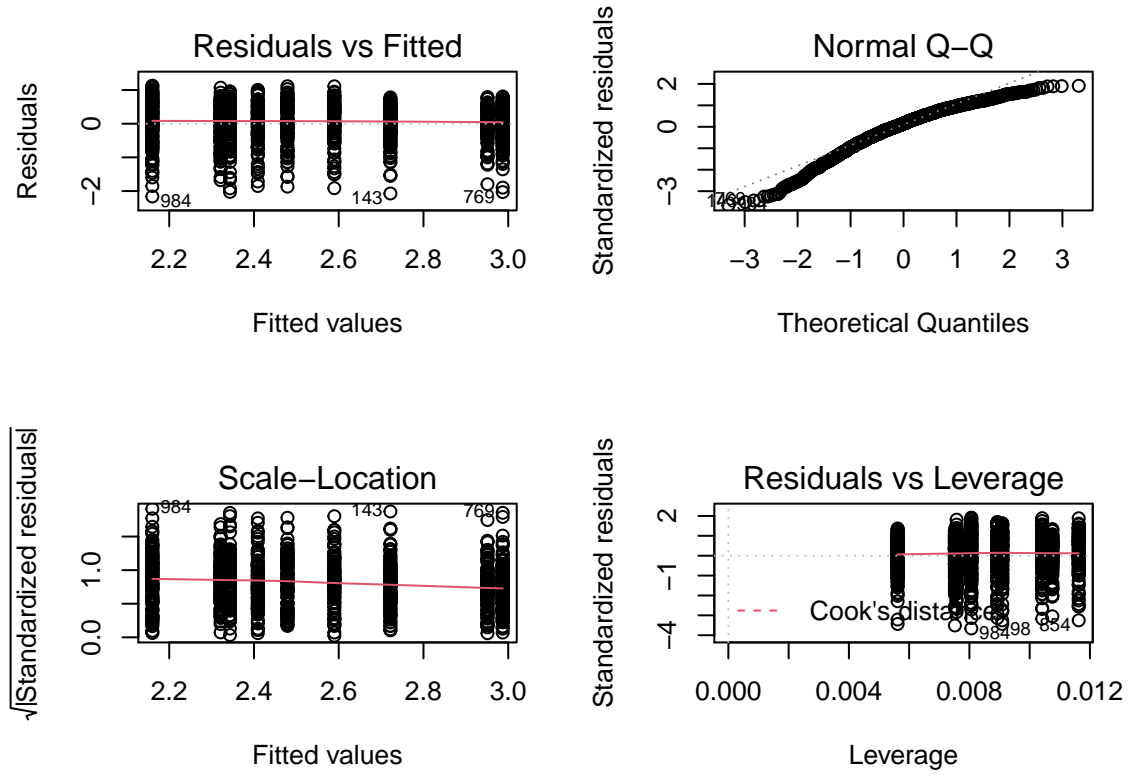


Figure 4: Checking taxonomic diversity ANOVA assumptions using diagnostic plots.

Table 10: Groupings assigned to ecoregions differing in taxonomic diversity using Tukey's HSD.

	diversity	groups
SAP	2.987	a
NAP	2.951	ab
WMT	2.722	bc
UMW	2.590	cd
CPL	2.479	cde
XER	2.409	de
TPL	2.343	def
NPL	2.321	ef
SPL	2.160	f

Table 11: Summary statistics of taxonomic diversity among ecoregions.

ecoregion	N	mean	sd	se	LL	UL
CPL	96	2.479	0.646	0.066	1.034	3.403
NAP	93	2.951	0.513	0.053	1.761	3.658
NPL	132	2.321	0.594	0.052	1.001	3.084
SAP	178	2.987	0.487	0.037	1.864	3.688
SPL	124	2.160	0.683	0.061	0.693	3.166
TPL	110	2.343	0.657	0.063	0.693	3.240
UMW	86	2.590	0.652	0.070	1.097	3.503
WMT	133	2.722	0.531	0.046	1.318	3.422
XER	112	2.409	0.607	0.057	0.967	3.346

## Community Composition

Community composition is compared among ecoregions. A Bray-Curtis dissimilarity index is applied to an untransformed abundance matrix, and then a permutational analysis of variance (PERMANOVA) with 10000 permutations was used to test for differences in community composition. Dimensions were added to the ordination to achieve adequate stress (stress < 0.20).

For additional reading on PERMANOVA, please see:

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26:32–46.

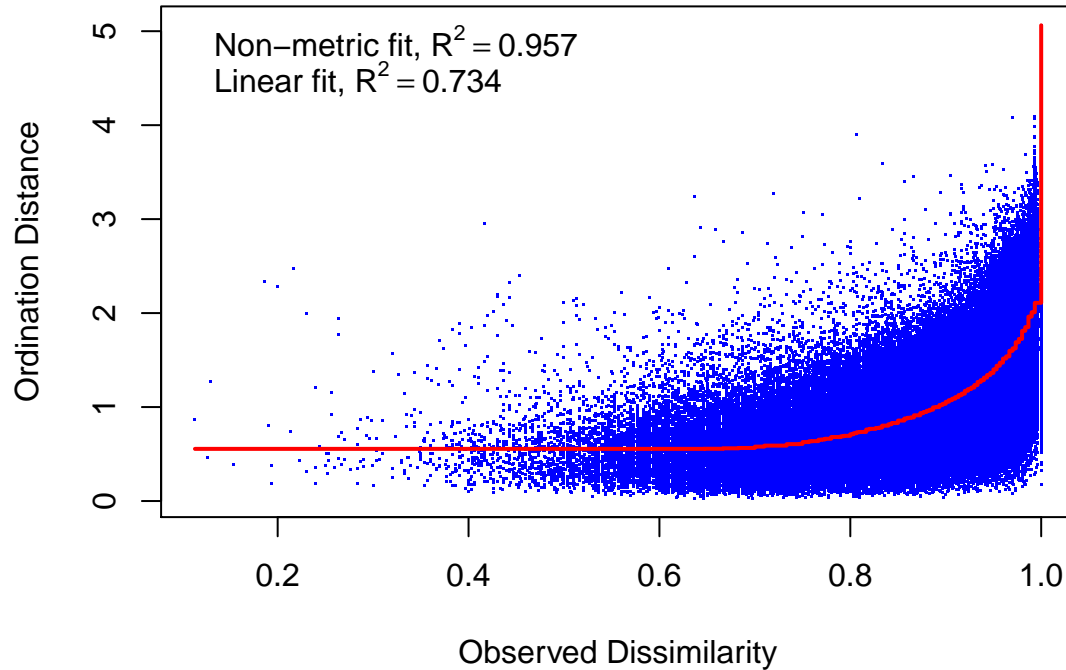


Figure 5: Diagnostic plot of the NMDS.

Table 12: Results of the PERMANOVA.

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
ecoregion	8	33.067	4.133	10.651	0.075	0
Residuals	1055	409.441	0.388	NA	0.925	NA
Total	1063	442.508	NA	NA	1.000	NA

## Boosted Regression Tree Analyses

Analyze boosted regression trees (BRTs) to identify which predictors have the greatest influence on community assembly and composition. A single BRT is quantified with all data (i.e. all ecoregions included) in order to identify which predictors are most important at the continental scale. In addition to the full BRT, separate BRTs are quantified for each ecoregion to examine important predictors that are ecoregion-specific or vary in relative importance among ecoregions.

BRTs have predictor variables grouped into three classes: environmental, landscape, and network. Environmental variables are considered “local” while landscape and network predictors are considered “regional.”

- List of variables and their respective categories:
- Environmental = NO3, NH4, total.P, pH.lab, cond, DOC, LWD.reach, NAT.cover, ALG.cover
- Landscape = pct.for, pct.ag, pct.urb, pct.ISC
- Network = site.lat, site.long, basin.area, mean.annual.flow, site.centralitiy

BRTs parameters were determined with model tuning, whereby the number of trees, interaction depth, observation per node, and learning rate were set to a series of different values to optimize the BRTs. BRTs were optimized for each ecoregion and the CUS, with specific parameters values varying among BRTs.

Table 13: Model parameters for each BRT.

ecoregion	number.of.trees	interaction.depth	learning.rate	minimum.observations
CUS	7500	3	1e-04	25
CPL	1000	3	5e-04	15
NAP	7500	2	1e-04	5
NPL	1000	2	1e-03	15
SAP	7500	3	1e-04	5
SPL	1000	3	1e-04	20
TPL	5000	3	1e-04	20
UMW	1000	3	5e-04	5
WMT	7500	3	1e-04	25
XER	1000	2	1e-03	25

Although certain model parameters varied among BRTs, all BRTs were set with a bagging fraction of 0.5 and 10-fold cross validation.

For further reading on BRTs, please see:

De’Ath, G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* 88:243–251.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77:802–813.

## BRT Numerical Summaries

Exact values of relative influence for predictor variables of community assembly for the CUS and all nine ecoregions.

Table 14: Results of the CUS BRT.

predictors	influence	predictor.type
site.centralcity	15.112	Network
LWD.reach	12.064	Environmental
pct.for	11.296	Landscape
NO3	8.414	Environmental
site.lat	8.277	Network
pH.lab	6.231	Environmental
pct.ISC	5.901	Landscape
mean.annual.flow	5.313	Network
NAT.cover	4.705	Environmental
DOC	4.685	Environmental
site.long	4.236	Network
pct.ag	2.856	Landscape
basin.area	2.219	Network
pct.urb	2.150	Landscape
NH4	2.028	Environmental
cond	1.977	Environmental
total.P	1.504	Environmental
ALG.cover	1.032	Environmental

Table 15: Results of the CPL BRT.

predictors	influence	predictor.type
LWD.reach	20.217	Environmental
NH4	16.839	Environmental
site.centralcity	13.050	Network
total.P	6.315	Environmental
basin.area	5.502	Network
pH.lab	4.856	Environmental
pct.urb	4.776	Landscape
site.lat	4.776	Network
pct.ag	4.260	Landscape
NAT.cover	3.381	Environmental
cond	3.171	Environmental
pct.ISC	2.979	Landscape
DOC	2.379	Environmental
mean.annual.flow	2.247	Network
ALG.cover	1.780	Environmental
site.long	1.487	Network
NO3	1.190	Environmental
pct.for	0.793	Landscape

Table 16: Results of the NAP BRT.

predictors	influence	predictor.type
NAT.cover	29.128	Environmental
LWD.reach	19.466	Environmental
NO3	10.035	Environmental
site.centralcity	5.031	Network
pct.ag	4.965	Landscape
pct.for	4.869	Landscape
ALG.cover	4.584	Environmental
cond	2.723	Environmental
basin.area	2.638	Network
DOC	2.597	Environmental
total.P	2.571	Environmental
mean.annual.flow	2.535	Network
pct.ISC	2.247	Landscape
pH.lab	1.669	Environmental
site.lat	1.460	Network
NH4	1.172	Environmental
site.long	1.156	Network
pct.urb	1.153	Landscape

Table 17: Results of the NPL BRT.

predictors	influence	predictor.type
pct.for	24.362	Landscape
DOC	14.689	Environmental
site.centralcity	10.256	Network
NH4	9.016	Environmental
LWD.reach	7.802	Environmental
mean.annual.flow	7.141	Network
site.lat	3.984	Network
site.long	2.813	Network
cond	2.801	Environmental
pct.ag	2.470	Landscape
basin.area	2.404	Network
NO3	2.373	Environmental
pH.lab	2.211	Environmental
pct.ISC	1.799	Landscape
pct.urb	1.604	Landscape
total.P	1.538	Environmental
NAT.cover	1.424	Environmental
ALG.cover	1.314	Environmental

Table 18: Results of the SAP BRT.

predictors	influence	predictor.type
site.lat	18.093	Network
site.centraliity	17.196	Network
pct.ag	8.168	Landscape
total.P	6.735	Environmental
pct.for	6.401	Landscape
pH.lab	5.034	Environmental
NO3	4.729	Environmental
pct.ISC	4.384	Landscape
DOC	4.310	Environmental
NH4	4.282	Environmental
LWD.reach	4.230	Environmental
site.long	4.030	Network
cond	2.676	Environmental
pct.urb	2.635	Landscape
NAT.cover	2.497	Environmental
basin.area	2.114	Network
ALG.cover	1.436	Environmental
mean.annual.flow	1.049	Network

Table 19: Results of the SPL BRT.

predictors	influence	predictor.type
pH.lab	23.045	Environmental
pct.ISC	13.698	Landscape
mean.annual.flow	12.198	Network
LWD.reach	8.680	Environmental
NAT.cover	8.652	Environmental
ALG.cover	5.174	Environmental
NH4	4.462	Environmental
basin.area	3.718	Network
site.centraliity	3.049	Network
pct.urb	2.682	Landscape
DOC	2.417	Environmental
pct.ag	2.203	Landscape
site.long	2.142	Network
site.lat	2.104	Network
pct.for	2.047	Landscape
NO3	1.725	Environmental
cond	1.241	Environmental
total.P	0.762	Environmental

Table 20: Results of the TPL BRT.

predictors	influence	predictor.type
cond	27.516	Environmental
site.centrality	20.119	Network
LWD.reach	15.876	Environmental
site.long	9.832	Network
pct.ISC	4.580	Landscape
pct.for	4.011	Landscape
NO3	3.322	Environmental
pH.lab	2.562	Environmental
DOC	2.362	Environmental
basin.area	1.559	Network
pct.urb	1.549	Landscape
site.lat	1.474	Network
NAT.cover	1.366	Environmental
NH4	1.296	Environmental
total.P	1.205	Environmental
pct.ag	0.766	Landscape
mean.annual.flow	0.401	Network
ALG.cover	0.204	Environmental

Table 21: Results of the UMW BRT.

predictors	influence	predictor.type
pct.for	19.753	Landscape
basin.area	10.013	Network
pH.lab	8.227	Environmental
pct.urb	7.633	Landscape
site.long	6.921	Network
LWD.reach	6.570	Environmental
NO3	5.887	Environmental
pct.ag	4.450	Landscape
mean.annual.flow	4.121	Network
NAT.cover	4.117	Environmental
total.P	4.071	Environmental
site.centrality	4.055	Network
DOC	3.571	Environmental
NH4	2.926	Environmental
pct.ISC	2.774	Landscape
cond	2.696	Environmental
site.lat	1.858	Network
ALG.cover	0.356	Environmental



Table 22: Results of the WMT BRT.

predictors	influence	predictor.type
pct.urb	27.016	Landscape
DOC	15.670	Environmental
NO3	12.772	Environmental
pct.ISC	9.844	Landscape
pH.lab	6.671	Environmental
site.centrality	6.472	Network
LWD.reach	5.605	Environmental
total.P	3.424	Environmental
site.lat	1.979	Network
mean.annual.flow	1.767	Network
NAT.cover	1.599	Environmental
pct.ag	1.596	Landscape
NH4	1.506	Environmental
pct.for	1.078	Landscape
ALG.cover	1.021	Environmental
site.long	1.010	Network
basin.area	0.720	Network
cond	0.251	Environmental

Table 23: Results of the XER BRT.

predictors	influence	predictor.type
site.centrality	36.197	Network
pct.for	20.055	Landscape
total.P	10.084	Environmental
site.lat	9.939	Network
DOC	9.865	Environmental
cond	2.274	Environmental
NO3	1.828	Environmental
LWD.reach	1.762	Environmental
NH4	1.290	Environmental
basin.area	1.090	Network
pct.urb	1.081	Landscape
pH.lab	0.891	Environmental
NAT.cover	0.864	Environmental
site.long	0.806	Network
pct.ISC	0.779	Landscape
mean.annual.flow	0.748	Network
ALG.cover	0.319	Environmental
pct.ag	0.128	Landscape