

NRSA-CA

David Murray-Stoker

18 December 2019

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
484332	2.108420	0.0349947	521862.68	17800.383	CUS
20113	2.969155	0.0029862	25983.07	1977.015	CPL
23614	2.894735	0.0037948	28717.37	1762.985	NAP
16794	3.634621	0.0002784	22933.15	1689.075	NPL
53727	2.042062	0.0411454	58619.43	2395.830	SAP
18485	2.904496	0.0036785	22419.57	1354.648	SPL
23604	2.744539	0.0060596	27248.64	1327.962	TPL
20681	3.112382	0.0018558	24883.14	1350.137	UMW
30335	2.920299	0.0034970	37330.20	2395.370	WMT
18850	2.630525	0.0085253	23611.80	1810.211	XER

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

turnover.Rep	turnover.z	turnover.P.value	turnover.mean	turnover.variance	ecoregion
4949766591	-18.757035	0.0e+00	2708235134	119503505.8	CUS
11163352	-6.579414	0.0e+00	8079993	468637.3	CPL
12340363	-7.685298	0.0e+00	8474040	503080.4	NAP
8052485	-10.451397	0.0e+00	5083595	284066.3	NPL
46311799	-10.927491	0.0e+00	27586229	1713620.3	SAP
5185531	-6.712332	0.0e+00	3625451	232420.0	SPL
8283736	-4.639863	3.5e-06	6337100	419546.0	TPL
8969842	-7.163636	0.0e+00	6323463	369418.5	UMW
23880410	-10.449173	0.0e+00	15036326	846390.8	WMT
7029135	-9.137875	0.0e+00	4544428	271913.0	XER

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

boundary.Index	boundary.P.value	boundary.df	ecoregion
2.550278	0	938	CUS
1.662548	0	93	CPL
1.462868	0	90	NAP
1.675775	0	85	NPL
1.945633	0	171	SAP
1.647613	0	87	SPL
1.785118	0	105	TPL
1.532878	0	83	UMW
1.738667	0	119	WMT
1.458962	0	81	XER

Additive Partitioning

Assess how taxonomic richness scales across spatial scales. Alpha (α) diversity is local richness, while beta (β) diversity is the turnover in diversity from alpha (local, α) to gamma (regional, γ). Additive partitioning scales diversity across samples in a hierarchical sampling design. For this study, α = site, β_1 = site-to-ecoregion, and β_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 α , β , and γ 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	0.0447133	0.1424549	0.000999
alpha.2	0.4088457	0.5240030	0.000999
gamma	1.0000000	1.0000000	1.000000
beta.1	0.3641324	0.3815480	0.000999
beta.2	0.5911543	0.4759970	0.000999

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

	observed	simulated	p.value
alpha.1	16.153289	82.300229	0.000999
alpha.2	94.961067	131.269107	0.000999
gamma	155.645000	155.645000	1.000000
beta.1	9.635499	1.891187	0.000999
beta.2	1.639040	1.185756	0.000999

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

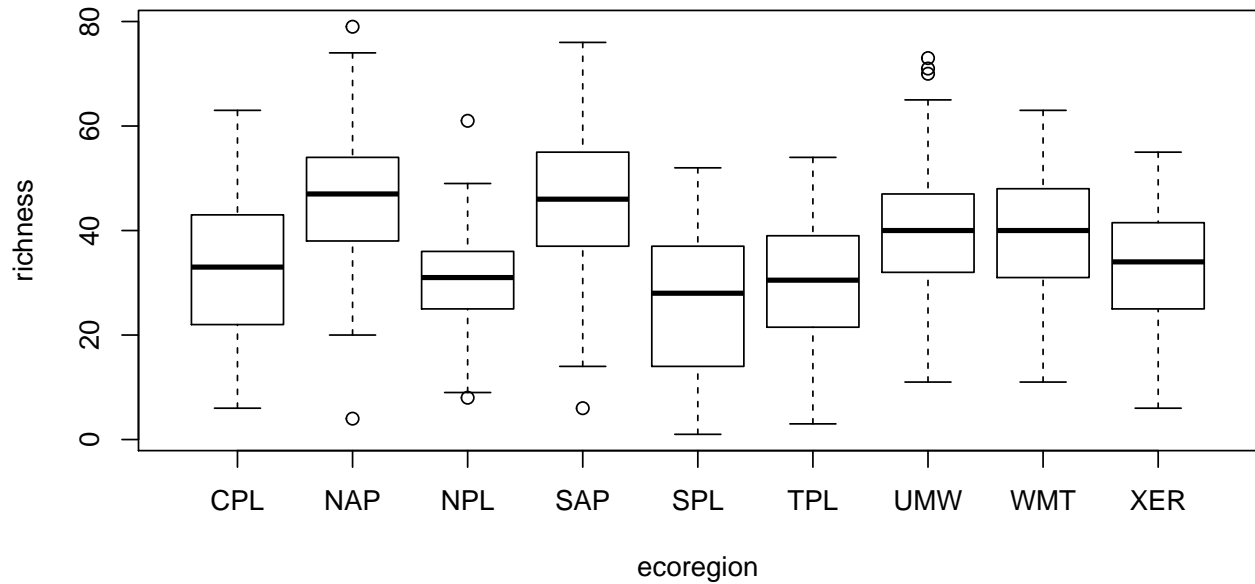


Figure 1: Boxplot of taxonomic richness by ecoregion.

Table 6: Results of the taxonomic richness ANOVA.

term	df	sumsq	meansq	statistic	p.value
ecoregion	8	45641.49	5705.1862	37.46152	0
Residuals	932	141938.55	152.2946	NA	NA

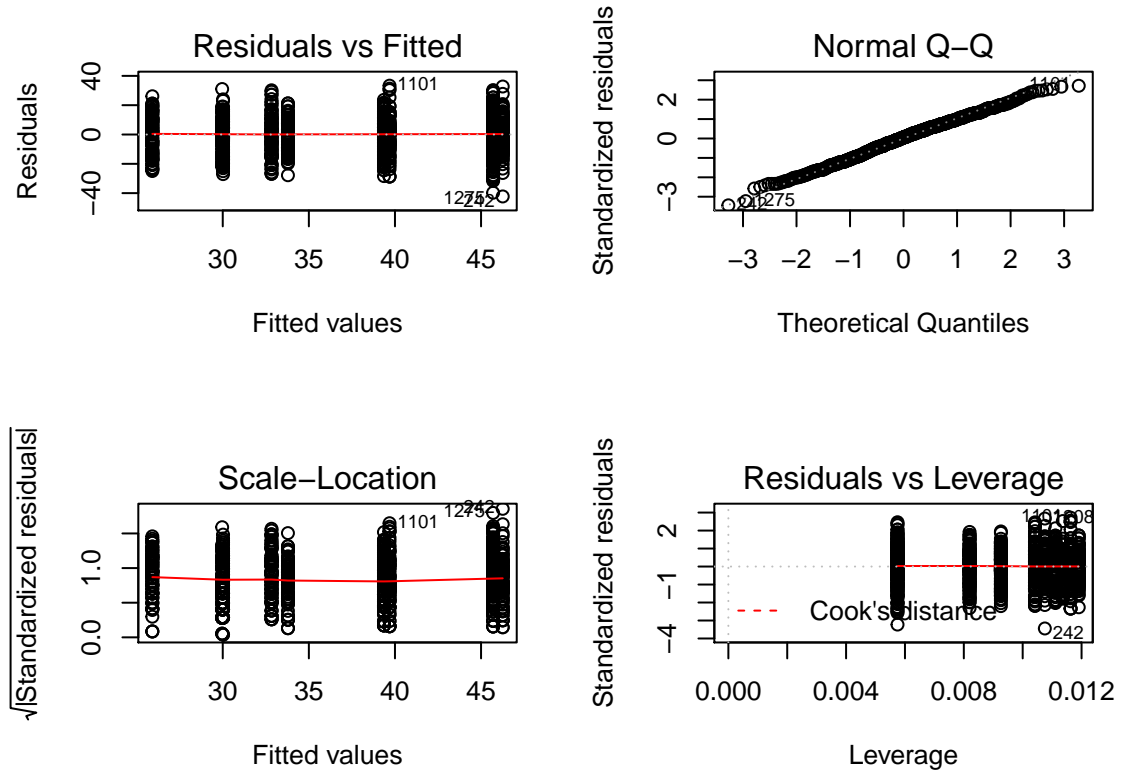


Figure 2: Checking taxonomic richness ANOVA assumptions using diagnostic plots.

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

	richness	groups
NAP	46.25806	a
SAP	45.70115	a
UMW	39.68605	b
WMT	39.36066	b
XER	33.78571	c
CPL	32.83333	c
TPL	30.01852	cd
NPL	29.96591	cd
SPL	25.91111	d

Table 8: Summary statistics of taxonomic richness among ecoregions.

ecoregion	N	mean	sd	se	LL	UL
CPL	96	32.83333	13.681079	1.396319	7.750	60.000
NAP	93	46.25806	12.116198	1.256391	23.900	66.700
NPL	88	29.96591	9.753807	1.039759	10.000	46.650
SAP	174	45.70115	13.729943	1.040864	17.325	70.350
SPL	90	25.91111	13.196564	1.391040	2.225	46.775
TPL	108	30.01852	11.785451	1.134056	7.675	47.650
UMW	86	39.68605	12.868806	1.387679	11.625	69.375
WMT	122	39.36066	11.457478	1.037311	17.000	59.975
XER	84	33.78571	10.664739	1.163618	15.150	53.000

Taxonomic Diversity

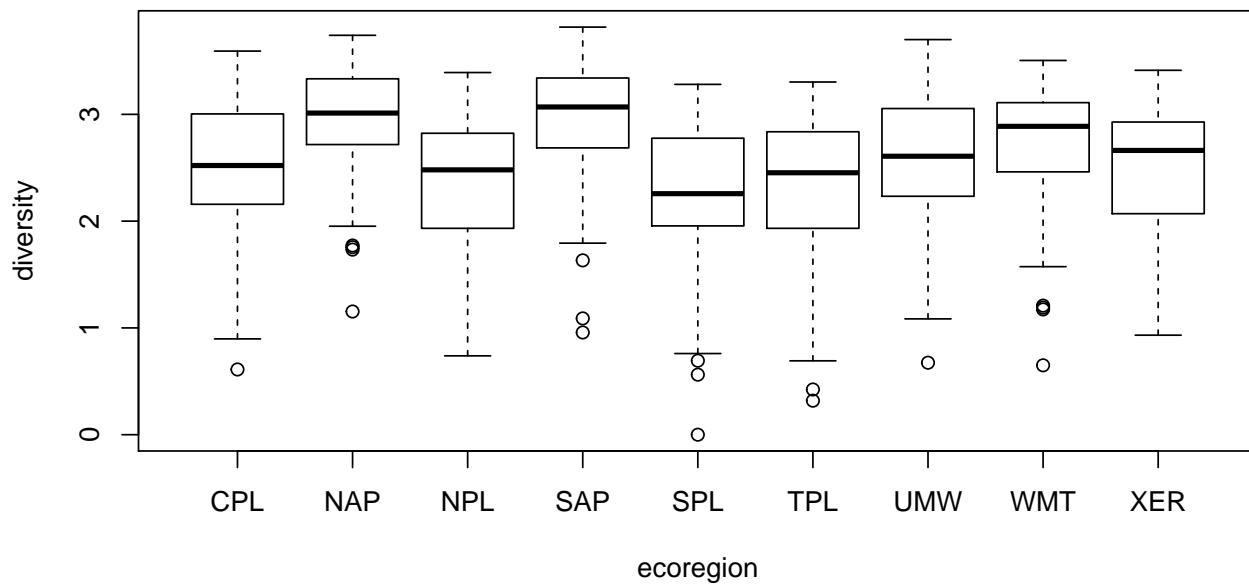


Figure 3: Boxplot of taxonomic richness by ecoregion.

Table 9: Results of the taxonomic diversity ANOVA.

term	df	sumsq	meansq	statistic	p.value
ecoregion	8	67.30839	8.4135489	24.31641	0
Residuals	932	322.47469	0.3460029	NA	NA

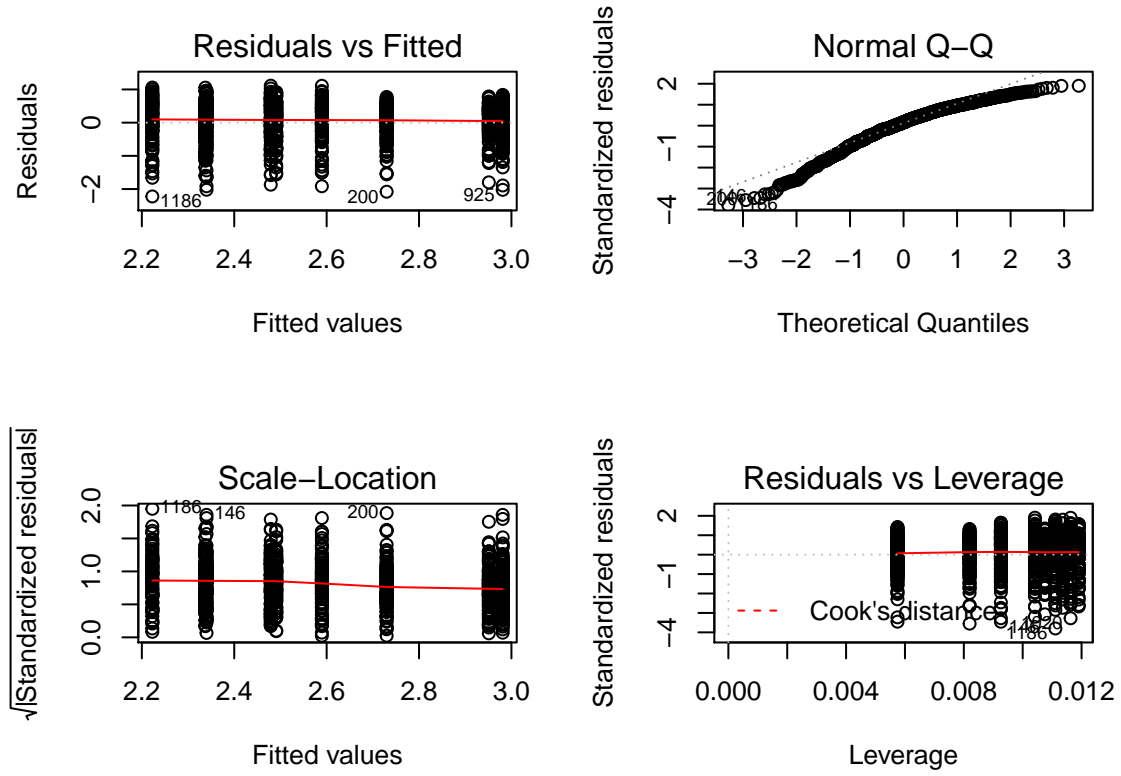


Figure 4: Checking taxonomic diversity ANOVA assumptions using diagnostic plots.

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

	diversity	groups
SAP	2.981129	a
NAP	2.951072	ab
WMT	2.729889	bc
UMW	2.589673	cd
XER	2.490968	cde
CPL	2.479081	de
TPL	2.340052	de
NPL	2.337500	de
SPL	2.222758	e

Table 11: Summary statistics of taxonomic diversity among ecoregions.

ecoregion	N	mean	sd	se	LL	UL
CPL	96	2.479081	0.6455157	0.0658827	1.0338034	3.402619
NAP	93	2.951072	0.5126766	0.0531621	1.7605864	3.657523
NPL	88	2.337500	0.5728261	0.0610635	1.2053433	3.071782
SAP	174	2.981129	0.4890645	0.0370759	1.8600463	3.731938
SPL	90	2.222758	0.6815452	0.0718412	0.7081684	3.158389
TPL	108	2.340052	0.6603122	0.0635386	0.6932173	3.240864
UMW	86	2.589673	0.6520973	0.0703175	1.0971820	3.502928
WMT	122	2.729889	0.5357936	0.0485085	1.2179373	3.409602
XER	84	2.490968	0.6002721	0.0654951	1.1458003	3.375190

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.

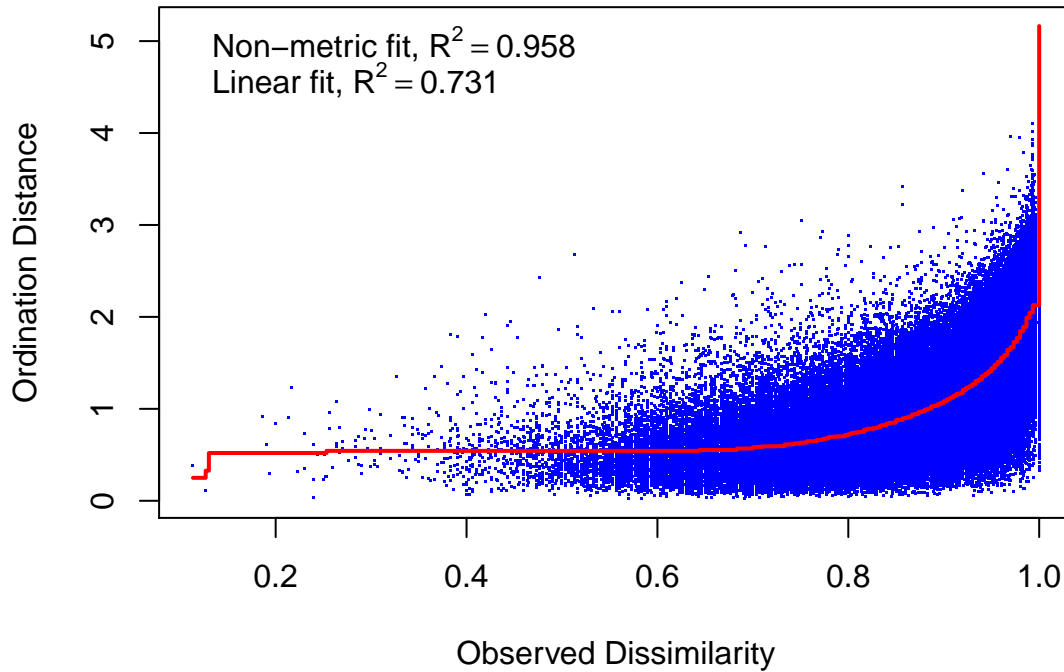


Figure 5: Diagnostic plot of the NMDS.

Table 12: Results of the PERMANOVA.

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
ecoregion	8	32.00908	4.0011349	10.45905	0.0823813	1e-04
Residuals	932	356.53873	0.3825523	NA	0.9176187	NA
Total	940	388.54781	NA	NA	1.0000000	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.centralty

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	0.00100	50
CPL	15000	3	0.00025	10
NAP	15000	2	0.00050	5
NPL	7500	2	0.00050	5
SAP	20000	3	0.00050	20
SPL	7500	3	0.00050	10
TPL	2500	2	0.00100	20
UMW	1000	2	0.01000	5
WMT	10000	2	0.00050	5
XER	15000	2	0.00025	5

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.lat	12.351509	Network
total.P	9.231544	Environmental
pct.ag	8.372920	Landscape
cond	7.410218	Environmental
pct.for	7.120120	Landscape
DOC	6.919976	Environmental
site.long	5.944398	Network
pH.lab	5.263746	Environmental
NH4	5.120600	Environmental
site.centraliity	5.086402	Network
NO3	4.729988	Environmental
basin.area	3.908186	Network
NAT.cover	3.880099	Environmental
mean.annual.flow	3.599771	Network
LWD.reach	3.468289	Environmental
pct.urb	3.388823	Landscape
pct.ISC	3.089224	Landscape
ALG.cover	1.114187	Environmental

Table 15: Results of the CPL BRT.

predictors	influence	predictor.type
DOC	28.7291997	Environmental
pct.for	16.4242187	Landscape
NO3	9.9708354	Environmental
total.P	9.9483353	Environmental
NH4	5.5773930	Environmental
site.centraliity	4.6732420	Network
LWD.reach	3.2736208	Environmental
pct.ag	3.1696737	Landscape
site.lat	3.0352175	Network
cond	2.6262132	Environmental
NAT.cover	2.3847278	Environmental
pH.lab	2.2789846	Environmental
basin.area	1.9665415	Network
site.long	1.7316011	Network
pct.urb	1.5713755	Landscape
mean.annual.flow	1.4997526	Network
pct.ISC	0.9098072	Landscape
ALG.cover	0.2292604	Environmental

Table 16: Results of the NAP BRT.

predictors	influence	predictor.type
NH4	34.478333	Environmental
cond	13.888228	Environmental
pct.for	10.313031	Landscape
site.long	10.287049	Network
total.P	4.008028	Environmental
pct.ag	3.224509	Landscape
site.lat	2.872994	Network
pH.lab	2.791352	Environmental
mean.annual.flow	2.778808	Network
basin.area	2.548968	Network
DOC	2.272747	Environmental
site.centralitiy	2.196276	Network
NAT.cover	1.574417	Environmental
LWD.reach	1.564089	Environmental
NO3	1.548912	Environmental
pct.urb	1.489271	Landscape
ALG.cover	1.130450	Environmental
pct.ISC	1.032538	Landscape

Table 17: Results of the NPL BRT.

predictors	influence	predictor.type
cond	30.5110470	Environmental
site.long	17.9548362	Network
DOC	17.3567343	Environmental
pct.for	12.6557456	Landscape
pct.ISC	6.5272360	Landscape
total.P	4.1733100	Environmental
site.lat	2.2944755	Network
pct.urb	1.5837043	Landscape
NH4	1.3216064	Environmental
NO3	0.8589309	Environmental
ALG.cover	0.8503975	Environmental
NAT.cover	0.8240487	Environmental
mean.annual.flow	0.7842665	Network
pct.ag	0.5662989	Landscape
basin.area	0.5318135	Network
LWD.reach	0.5046404	Environmental
site.centralitiy	0.4125010	Network
pH.lab	0.2884073	Environmental

Table 18: Results of the SAP BRT.

predictors	influence	predictor.type
DOC	20.4409423	Environmental
NH4	12.7859145	Environmental
NO3	9.5527444	Environmental
site.long	8.9558106	Network
pct.for	7.6357228	Landscape
pH.lab	7.2654607	Environmental
basin.area	5.8382434	Network
pct.ag	4.6032581	Landscape
total.P	3.3416431	Environmental
cond	3.3280966	Environmental
NAT.cover	2.9964823	Environmental
site.lat	2.9824967	Network
LWD.reach	2.8341707	Environmental
mean.annual.flow	2.2949118	Network
site.centralitiy	2.1453992	Network
pct.urb	1.2837286	Landscape
pct.ISC	1.2759481	Landscape
ALG.cover	0.4390262	Environmental

Table 19: Results of the SPL BRT.

predictors	influence	predictor.type
site.lat	16.286141	Network
site.long	13.713756	Network
NO3	13.494424	Environmental
DOC	11.023835	Environmental
pct.ag	7.371879	Landscape
site.centralitiy	5.727863	Network
ALG.cover	5.053961	Environmental
LWD.reach	4.179257	Environmental
pH.lab	3.917487	Environmental
basin.area	3.837246	Network
mean.annual.flow	2.884376	Network
cond	2.586143	Environmental
pct.urb	2.561533	Landscape
NAT.cover	1.687666	Environmental
total.P	1.678962	Environmental
pct.for	1.409987	Landscape
pct.ISC	1.387481	Landscape
NH4	1.198003	Environmental

Table 20: Results of the TPL BRT.

predictors	influence	predictor.type
basin.area	35.1997549	Network
NO3	18.4203581	Environmental
pH.lab	9.4762760	Environmental
LWD.reach	7.9008157	Environmental
NH4	7.6744213	Environmental
DOC	5.6998098	Environmental
site.long	4.0899052	Network
site.centralitiy	3.6876847	Network
site.lat	2.4403579	Network
pct.ag	1.1637302	Landscape
cond	1.0366559	Environmental
mean.annual.flow	0.9707870	Network
pct.urb	0.5083977	Landscape
ALG.cover	0.4337641	Environmental
pct.for	0.4304421	Landscape
total.P	0.3620201	Environmental
NAT.cover	0.2777532	Environmental
pct.ISC	0.2270662	Landscape

Table 21: Results of the UMW BRT.

predictors	influence	predictor.type
DOC	25.165404	Environmental
NO3	11.206464	Environmental
NH4	11.113927	Environmental
pH.lab	7.403112	Environmental
site.centralitiy	6.784653	Network
total.P	6.162017	Environmental
NAT.cover	5.071966	Environmental
pct.for	3.821120	Landscape
site.long	2.912266	Network
LWD.reach	2.749846	Environmental
basin.area	2.695048	Network
pct.ISC	2.590454	Landscape
site.lat	2.535323	Network
mean.annual.flow	2.396979	Network
ALG.cover	2.313240	Environmental
pct.ag	2.199434	Landscape
cond	1.517581	Environmental
pct.urb	1.361168	Landscape

Table 22: Results of the WMT BRT.

predictors	influence	predictor.type
site.lat	20.1008528	Network
cond	10.3737148	Environmental
basin.area	10.1512262	Network
DOC	8.9008676	Environmental
ALG.cover	8.8863927	Environmental
NAT.cover	7.9984966	Environmental
LWD.reach	7.0579996	Environmental
NH4	5.8492324	Environmental
pH.lab	4.1541184	Environmental
total.P	3.8826824	Environmental
pct.urb	2.7681576	Landscape
NO3	2.7403230	Environmental
pct.for	1.8725301	Landscape
pct.ag	1.7125326	Landscape
site.long	1.3793442	Network
mean.annual.flow	1.0062318	Network
site.centraliity	0.8555877	Network
pct.ISC	0.3097094	Landscape

Table 23: Results of the XER BRT.

predictors	influence	predictor.type
basin.area	26.4512880	Network
pct.ISC	11.1235318	Landscape
ALG.cover	8.8352475	Environmental
pct.ag	8.0219606	Landscape
cond	7.8148714	Environmental
site.lat	6.2596968	Network
pct.urb	5.5019569	Landscape
LWD.reach	4.8977165	Environmental
NH4	4.6424673	Environmental
DOC	3.4157143	Environmental
NAT.cover	3.2068245	Environmental
pH.lab	2.8935636	Environmental
site.centraliity	1.9104060	Network
NO3	1.4585930	Environmental
total.P	1.1890544	Environmental
site.long	0.9097713	Network
pct.for	0.7816872	Landscape
mean.annual.flow	0.6856489	Network