

# NRSA-CA

David Murray-Stoker

18 December 2019

## National Rivers and Streams Assessment-Community Assembly

### Model training for boosted regression tree analyses

Boosted regression trees are tuned across a suite of model parameters to optimize model performance and accuracy as well as reducing the potential of overfitting.

### Data Management

Subset data compiled from the community assembly analyses so that only predictors used in each set of boosted regression trees are included in the data.

### Boosted Regression Tree (BRT) Model Training

Analyze boosted regression trees (BRTs) to identify which predictors have the greatest influence on community functional composition. A single BRT is quantified with all data (i.e. all ecoregions included) in order to identify which predictors are most important for the conterminous United States (CUS). In addition to the CUS 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, geographic, and network. Environmental variables are considered “local” while geographic 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, observations per node, and learning rate were set to a series of different values to optimize the BRTs for each ecoregion and the CUS.

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 Model Training Results

Results derived from the BRT model training. These model parameters will be tuned for each model evaluated during the actual BRT analyses to identify the drivers of community assembly.

Table 1: Model training results for the BRTs.

n.trees	interaction.depth	shrinkage	n.minobsinnode	ecoregion
7500	3	0.00100	50	CUS
15000	3	0.00025	10	CPL
15000	2	0.00050	5	NAP
7500	2	0.00050	5	NPL
20000	3	0.00050	20	SAP
7500	3	0.00050	10	SPL
2500	2	0.00100	25	TPL
1000	2	0.01000	5	UMW
10000	2	0.00050	5	WMT
15000	2	0.00025	5	XER