Reviewer question: What was the variation in quality (e.g. % undisturbed watershed) in reference sites among ecoregions? Without that qualification, comparison across ecoregions is less informative because quality of reference sites in addition to natural factors remains in important variable. Is there a standardized scale for reference condition across ecoregions?

Response: As noted in the report, the same standard screens for establishing "Reference" were applied across all ecoregions. However, we have produced boxplots to indicate the distribution of % of undisturbed watershed within each of the ecoregions to facilitate a comparison among them. Most of the ecoregions are similar, however the south coast (especially xeric) has a broader distribution than the others—several sites have contributing catchments with a lower percentage of undeveloped land.
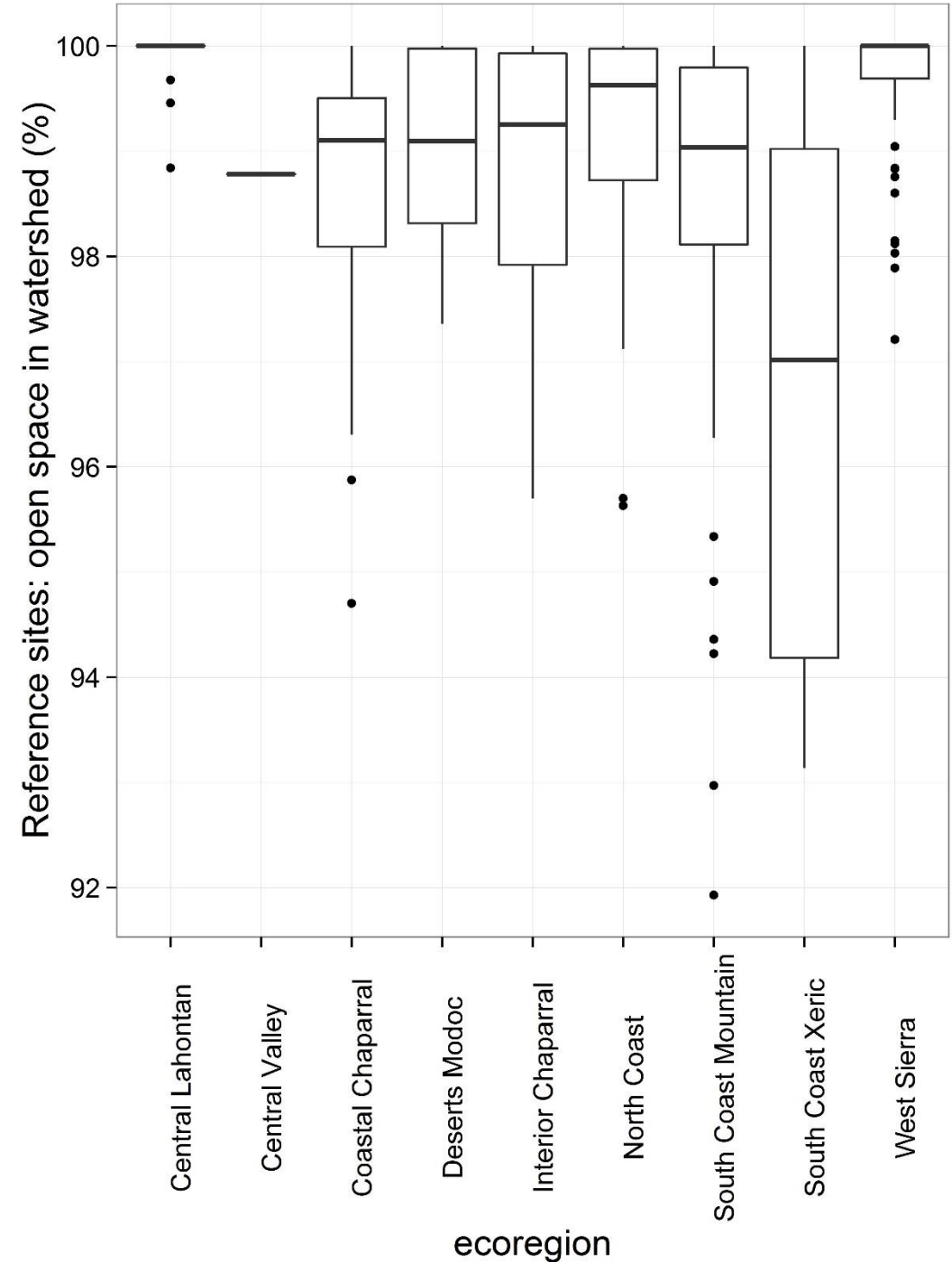
Reviewer question: Were response variables transformed to reduce skewness in the data and overemphasis on high nutrient and biomass levels? Although I have heard it argued that CART and BRT like analyses are robust to assumptions of normality in variables, given the importance of variance analyses, I'm not sure how that could be the case. Some simple runs of analyses using transformed and untransformed variables would provide a valuable evaluation of this potential problem if variables were not transformed.

Response: We transformed variables as necessary and re-ran 2 of the BRT analyses. Transformation had a negligible effect on results.

variables transformed?	assemblage	ALU type	biomass type(s) included in model	highest ranked predictor (relative influence)	model cv correlation (se)	# trees	# predictors in final model	# obser-vations	relative influence of biomass (rank)	relative influence of TN (rank)	relative influence of NO_x (rank)	relative influence of NH_4 (rank)	relative influence of TP (rank)	relative influence of PO_4 (rank)
no	BMI	CSCI	chlorophyll a	URBAN_2000_5K (16.2)	0.829 (0.012)	5850	31	611	2.35 (12)	**5.38 (6)**	0.43 (31)	NA	2.66 (11)	1.06 (22)
yes	BMI	CSCI	chlorophyll a	URBAN_2000_5K (15.73)	0.834 (0.008)	6150	31	611	2.30 (13)	**5.27 (6)**	0.48 (31)	NA	2.48 (11)	1.04 (23)
no	hybrid	H20	AFDM	URBAN_2000_WS (24.12)	0.847 (0.009)	5950	35	809	2.57 (9)	**18.47 (2)**	1.52 (10)	0.77 (23)	12.37 (3)	2.92 (7)
yes	hybrid	H20	AFDM	URBAN_2000_WS (23.5)	0.847 (0.009)	5900	35	809	2.5 (9)	**19.35 (2)**	1.43 (10)	0.75 (23)	12.45 (3)	2.90 (6)

Reviewer comment: The analyses of results do not distinguish whether thresholds mark the upper bounds or lower bounds of a response range and whether the response is below the threshold is okay or bad. (Also: Although the statistical methods are scientifically valid, they are conceptually flawed for application in policy for two reasons. First, do the thresholds exist? Are the thresholds real or just blips along a stressor gradient?)

Response: In general, additional analyses further support our assertion that the thresholds we identified are real and nearly all of them are exhaustion. There is, in general, greater variance and higher ALI values (among the ALIs that decrease with stress) among the sites that had gradient values below the threshold identified, than among those that had higher. The same 5 ALI-gradient combinations that were illustrated in the report in Figs. 3.14 – 3.18 are used to illustrate this:
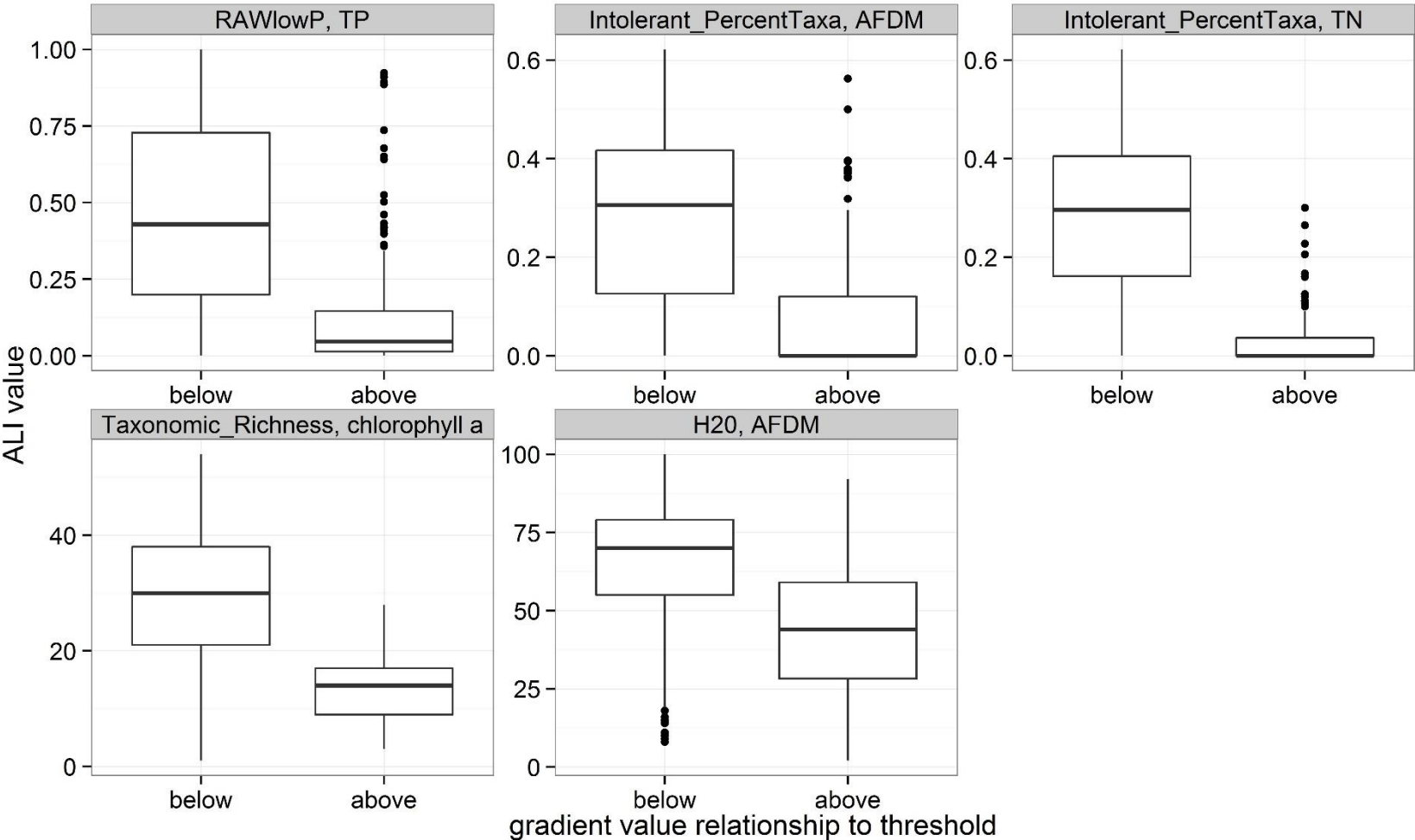
Distribution of aquatic life indicator (ALI) values among sites with stressor gradient (i.e., biomass or nutrient concentration) values below vs. above the threshold that had been determined in the report based on piecewise regression. The strip above each panel in the plot indicates the type of ALI followed by the type of gradient in question. The ALI values are indicated by the y-axis.

Reviewer comment: The piecewise linear regression results show the problem with the highly skewed results, and potential problems with loosing pattern recognition of assimilative capacity at really low nutrient concentrations and levels of productivity.

Response: We re-ran a subset of the piecewise regressions (the same 5 ALI-gradient combinations that were illustrated in the report in Figs. 3.14 – 3.18) using a truncated data set. Specifically, for each analysis, we used only data from sites in which the stressor gradient value was less than the threshold that we had determined for that ALI-gradient combination in previous analysis (as presented in the report). The goal was to see whether this would allow any potential resistance threshold to be detected. In 4 out of the 5 cases, the signature of the regression was not indicative of a resistance threshold (in fact, it looked more like the signature of an exhaustion threshold, again). In the one case in which a resistance threshold was possibly detected, it occurred at the extreme low end of the overall gradient (practically zero) calling into question how meaningful it was.
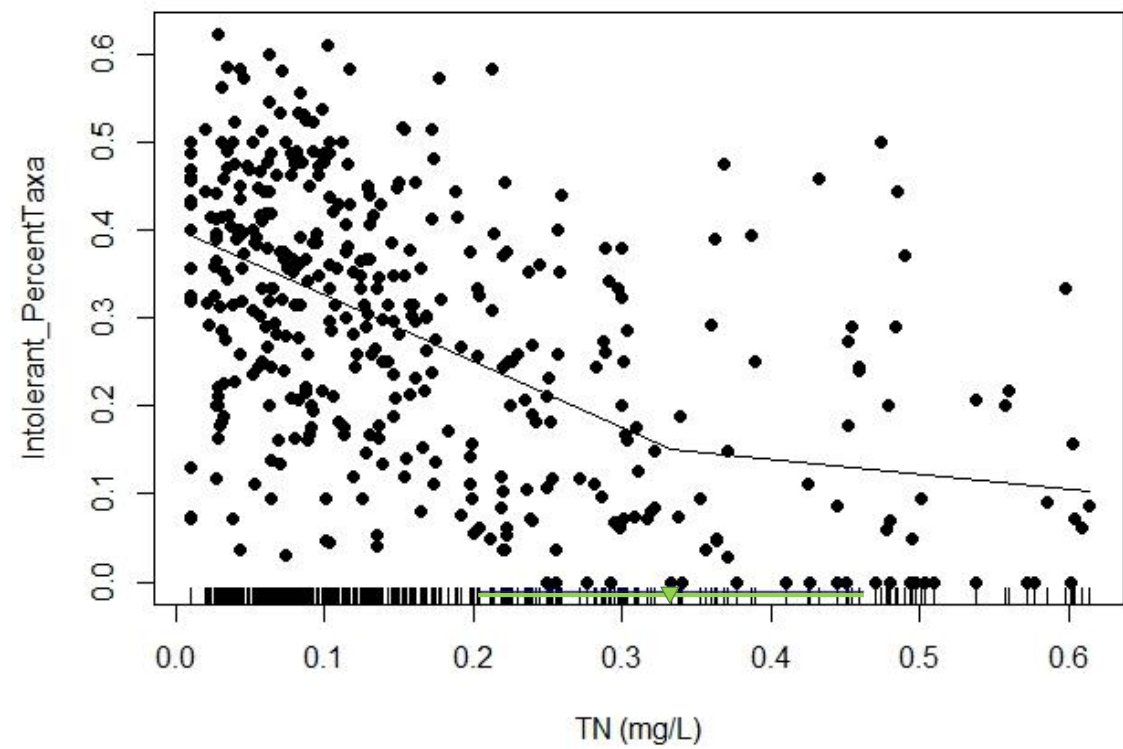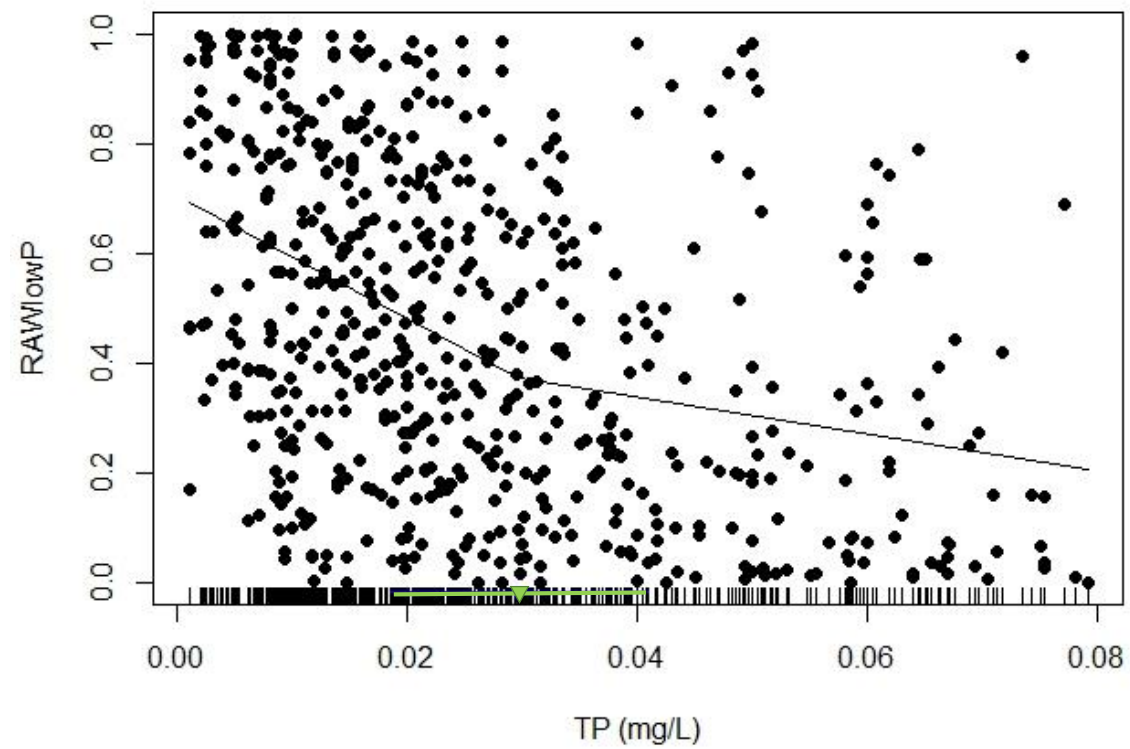
= confidence interval with breakpoint