order	Section/ Table/ Figure	relevant text	question (comment) posed by reviewer	authors' response
16	Section 4.2.5	We conducted bias analysis using the randomForest package in	Why switch to RF rather than BRT? Reason or just did?	RFs was the first choice for bias analysis. The BRT and Bayesian CART analysis were added in subsequent iterations of the report. RFs straightforward, and robust. Boosting for bias analysis can affected by outliers or noise.
21	Section 4.2.5 Bias Analysis (Objective 2)	After the predictor variables selection, datasets excluding observations with interpolated values of predictor variables were used in the subsequent random forest regression.	You might tell the reader what you did with the datasets that included interpolated values here briefly.	The interpolation was only conducted during the preliminary variable selection process, no dataset in the random forest regression analysis. The text has been modified for clarity.
23	4.3.2 Random Forest Regression for Bias and Variance Analysis (Objective 2)	However, the MSE of water temperature was negative , implying zero influence. Some of the regressions performed poorly for all variables. For example, for the QUAL2K with ACCRUAL model (Figure 4.4c), regression analysis showed no significant relationship between the predicted-minus-observed chlorophyll a and the explanatory variables, and the variable importance ranks also showed insignificant MSE change.	Where is this shown? You can tell which variables are negative in panel B but that is the only one I think.	Though there the x-axis scale doesn't show the negetive numbers, anything below the first number (2 or 5) were either inconsequential or negative. The argument presented in the paragraph is to show that even though the algorithm picks up certain variables, the influence of the variables might be negligible.
22	Section 4.2.5 Bias Analysis (Objective 2)	Because the CGMIidCART program cannot function with missing values, we substituted medians for missing values in the matrices. Based on the protocol suggested by Chipman et al. (2002), we evaluated regression trees using a range of model parameters, chose the "most visited tree" among model iterations	What proportion of the data had missing values. This may skew the results.	Five percent of values were missing in the final data sets used (after variable selection process). This has been added to the text. It is possible that missing values influenced the choice of classification variables.
10	Section 3.3.1	These observations are corroborated by the results of the CART analyses of diatom and BMI NMS axis 1 scores (Table 3.5, Figure 3.5), in which median cut point values for TP and TN were consistently <0.1 and <1 mg/L, respectively	CART often finds the upper change point, rather than the lower change point in curves with assimilative capacity.	Text has been added to note that CART found cutpoints intermediate between perceived resistance and exhaustion thresholds (based on smoothed loess curves through NMS axes); CART cutpoints were slightly higher than perceived resistance thresholds and much lower than exhaustion thresholds.
14	Section 3.4.1	For example, Welch et al. (1988) and Biggs (2000) have suggested that macroalgal percent cover in the range of 20-30% and above is unacceptable from the standpoints of aesthetics and recreation.	Stevenson et al. (2013) relate the 20% macroalgal cover to the thresholds in Suplee et al. (2009) for loss of aesthetic value.	**Suplee et al. (2009) is now cited in the text.**
5	Section 2.4	A plot of stream survey data suggest that South Coast sites may be exhibiting a peak in June-July,	This trend is not well-established in the data. I recommend deleting reference to it in this chapter. Ending a chapter discussion with poor quality analysis and conclusions based on them compromises the integrity of the rest of the chapter.	The graphs were included in response to an earlier reviewer's comment. We have moved the graphs to Appendix 5 (QAQC discussion).
6	Section 3.1	Three basic approaches have been used in establishing levels of stream nutrients and algal abundance that are protective of aquatic life uses (ALUs; U.S. Environmental Protection Agency 2000): 1) reference approach, 2) empirical stress-response approach, and 3) cause-effect approach .	I was confused about the differences between stressor-response and cause-effect approaches until reading further in the paragraph. This "cause-effect" phrase implies c ausal analysis, which is important for stressor-response approaches also. I recommend changing this to process-based modeling. The current wording implies a more formal causal analysis or that stressor-response approaches don't identify cause-effect relationships.	The "cause-effect" phrase has been replaced with "process-based".

8	Section 3.2.3	figure represents the threshold of adverse effects on response	The rationale for using thresholds is they provide a point (threshold) along the stressor gradient above which adverse effects get greater, which drives stakeholder consensus for levels of protection (Stevenson et al. 2008). A figure like the BCG (Fig 1) with some assimilative capacity provides that tipping point that stakeholders do not want to pass. The figures here (Fig. 3.1) and many of the later results do not illustrate that optimal non-linear response for use in management targets. In addition, the phrase "threshold of adverse effects" does not describe the meaning of the thresholds illustrated here, as pointed out in the next sentence. Threshold in adverse effects would indicate effects will get greater rather than lesser above the threshold. In each of these examples, to one extent or the other, the rate of change in attribute loss decreases above the threshold. Again, different threshold responses would illustrate these cases.	The wording has been corrected and text added to emphasize the fact that by the time an exhaustion threshold has been reached, significant ecological damage may already have occurred.
20	Section 4.4.2	Other explanatory variables, such as "Code 21" land use and W1-Hall, indicators of development and riparian disturbance	And the importance of these variables has to be indirect, but indicating their values as proxies for the factors that have the direct effects on benthic biomass, light, temperatures, substratu, current velocity, and time since last disturbance.	Agreed, but these "indirect" variables could represent important combinations of suites of factors that together influence the response to nutrients. This has been noted in the text.
1	Section 2.2.3	A CDF depicts the estimated probability distribution of values of a given indicator relative to the cumulative proportion of the geographic unit of interest, i.e., percent of stream length in the state.	So this implies that site-weights are used to account for the stratified random probabilisitic sampling design, right?	If the question is whether conducting a probability survey requires the calculation of weights, the answer is "yes", *if* strata (or otherwise unequal probability) are specified as part of the survey design, and densities of sites relative to (in this case) stream length totals are variable across strata (and/or multidensity categories). If conducting an equal probability survey that lacks stratification, then no weights need be calculated, as they are all equal and therefore do not affect estimates.
2	Section 2.2.3	and by ecoregion relative to the 75th percentile of reference sites	Were weights for sites recalculated for each ecoregion?	No. There was no need to recalculate them. Weights do not change just because estimates are being generated for different subsets of the target population.
3	Section 2.3.2	The proportion of sites for which samples were collected represented about 10% of the state total stream kilometers.	I find it hard to believe that you sampled and assessed 10% of the stream kilometers in CA. Were these samples from stream segments (confluence to confluence) and those segments represented 10% of stream kilometers in the state?	It doesn't mean we physically sampled 10% of the stream kilometers in the state. It means that what we sampled is taken to represent (statistically) 10% of the state's stream kilometers in the condition estimates on which we report. The point of conducting a probability survey is to collect a random subsample's worth of data that can then be scaled up to represent a stated portion of the population within the sampling frame.
4	Section 2.3.2	CDFs of site disturbance classes show a good amount of separation of reference, intermediate and stressed sites for chlorophyll a and AFDM, but not for macroalgal % cover (Figure 2.2).	Figure 2.2 is misnumbered. I had trouble determining what good amount of separation meant. So, I plotted box plots in my head with this information and figured that good separation does not mean that biomass would be significantly different among sites. However it does mean that the tails of the distributions do vary. I.e. about 15-20% of disturbed sites had chl above 50 mg/m2, but only almost none of the reference sites were that high. These results and the relative significance of these results should be explained. Differences in biomass and nutrients among disturbance	We did actually include such box plots in the report (in Appendix A). However, through an oversight, we failed to reference the appendix appropriately within the text. That has been rectified.
7	Table 3.2	5-km radius from sampling site	Future work should distinguish upstream and downstream land use, which I assume the radius approach does not. A circle arounds a sampling site probably is composed of about 1/3rd of land upstream of the site and 2/3rds downstream.	We did in fact clip the circle to the watershed (as delineated in GIS), such that the area in question only reflected land uses draining to the site. We have added a footnote to Table 2.2 that drives this point home.
11	Section 3.3.1		The report of the results emphasized the high end of the nutrient and algal biomass ranges for thresholds. That means that all changes have already occurred. You should report the nutrient and biomass concentrations above which thresholds occurred, i.e. the low end of the stressor gradient above which bad effects occur. That should be the focus of the management targets.	We are aware that resistance thresholds would be of more value than exhaustion thresholds for setting management targets. However, despite our best efforts to identify resistance thresholds, most analyses did not yield any. The reviewer had also commented that we failed to mention some resistance thresholds derived from the NMS plots. This was an oversight on our part, and we have now pointed them out. However, other than the few partial dependence plots from the BRT analyses in which resistance thresholds were apparent, none of the other analyses revealed them. This was true even when we bored down in the piecewise regression analyses, focusing on the low ends of the various gradients (**see PowerPoint attachment to these responses**). In summary, the general lack of resistance thresholds reported is not because we did not look for them, but because they appear to be largely undetectable using our methods and the available data. We have carefully qualified this fact in the report.
12	3.3.2	There was generally a high degree of correspondence between the piecewise regression output and the SiZer map for the various ALU/gradient combinations.	Some SiZer maps showed lower thresholds that could be associated with assimilative capacity and low nutrient levels. The skewness in the the independent variables prevents observation of that assimilative capacity. However, even with the lack of transformation in independent variables the SiZer figures for all independent variables except TP (e.g. figs 3.14-3.18 show a downward trend in the red zone, indicating the areas (bands) of negative slope increase, and then decrease. That zone in which the areas of negative slope increase (red increase and purple and/or gray decrease) indicate zones where negative trens are not as pronounced at the lowest "stressor" levels.	It appears the reviewer is referring to the fact that some of the maps show "doublet peaks" at the low end of the stressor gradient. Both peaks, in such cases, are indicating the same phenomenon: that of a strongly negative slope along the stressor gradient. One can understand what is happening by visualizing the scatterplot above each map. These are not indicative of resistance thresholds, but rather something more along the lines of a disjunct exhaustion threshold. We now point out the doublets, where they exist, to minimize potential confusion for the reader.

13	Section 3.4.1	thus lending additional support for the numbers we derived	The ranges in the literature represent almost the full range of possible nutrient concentrations. The fact that your thresholds like within these simply shows you were sampling someplace on earth.	It is true that some of the literature values are high, but it is nonetheless accurate to state that our values fell largely within the ranges (albeit wide) that have been reported. To address where it appears that the reviewer was going with his comment, though, it should be noted that the full range of nutrient concentration values occuring across streams throughout the state shows that nutrient concentrations can, and often do, reach values that exceed (in some cases, by orders of magnitude; see Table 2.6) the thresholds we report based on our own analyses. Therefore, within this context, the thresholds are not inherently meaningless for nutrient-management purposes (as we believe was being implied).
15	Section 3.4.1	Thus, based on the output of widely different analytical techniques for multiple biotic assemblages, narrow ranges of thresholds with high confidence were realized for both TN and TP .	Are you sure that both N and P are causally limiting nutrients?	We cannot prove (nor do we claim) causation, based on available data, but we have support for thresholds for both nutrients based on the reported relationships.
17	Section 4.3.3	Percent canopy cover was the most important predictor of percent cover of macroalgae (PCT_MAP; Table 4.8), accounting for over 13% relative influence.	Did you also account for reach direction (N-S-E-W) and effect of canopy on light availability (reaching the stream)?	No. For these ambient surveys, nothing so fancy was undertaken. Field crews simply used spherical densiometers to assess coverage.
18	Figure 4.6	Interaction plot of NOx and mean monthly maximum ambient air temperature (the mean of the month the sample was collected and the two months prior) from BRT model for chlorophyll a.	Be careful not to imply and interaction if there is not one, given the name of the plot. Maybe rename the plot. There is no evidence of interaction in this figure in the sense that differences in temperature affect the magnitude of effects of nitrate, and vice versa. For example, the nitrate effect is little affected by temperature, as far as I can see in this figure.	There were actually several significant interactions between predictors in the various BRT models (e.g., the one shown in Fig. 4.8). However, it is also true that the specific pair of predictors shown in Fig. 4.6 did not exhibit a significant interaction. We had called the graph in that figure (and its two analogs) an "interaction plot" because that's what those plots are called in the *dismo* package (whether there's a statistical interaction or not). However, we agree that this naming convention can be misleading and therefore have changed their name to "three-dimensional plots" (and have modified the text slightly) to ensure that there is no further confusion about this.
19	Section 4.4.2	In preliminary BRT models, both antecedent ambient air temperature and stream temperature on the day of sampling exhibited substantial influence on chlorophyll a and was significantly related to this biomass type, the levels of which increased dramatically in over the course of a relatively narrow range of antecedent ambient air temperature (approximately 26-28˚C).	Does air temperature also correlate with areas of the landscape that have been altered for agriculture and/or urban development? Plus these temperatures seem higher than what I would expect the optimal temperatures to be for nuisance filamentous algae or diatoms that would produce the highest benthic biomasses. Distinguish causal and correlational relationships.	There may be a correlation, but that does not obviate the possibility of a direct, positive effect of temperature on biomass accrual. BRT allows us to see how well predictors account for responses within the context of other (possibly confounded) predictors. Because the strong relationships between temperature and biomass were realized within the context of the types of landscape variables the reviewer mentions, there is little support for the possbility that land use is the ultimate driver of biomass accrual (with temperature subordinate), which is what it appears the reviewer is suggesting. Regarding distuinguishing causal from correlative, we have modified the language slightly to avoid unjustifiably implying causation.