We would like to extend our thanks to the reviewers for two high quality reviews. We found the comments to be thoughtful and in all cases the final product, we feel, is much better that what we originally submitted. Below you will find our itemized response to each comment made by the reviewers. We have also included this document as an upload.

# Reviewer 1

## Major Comments

***1-1.*** I would like more detail on how the averaging was done for each year. For each parameter, an annual average was calculated for each lake. In theory, Secchi depth and temperature were sampled every week and water samples were collected 3 times per season, May through October. If the monitoring design was followed consistently, then a simple average should be sufficient. However, I know from experience with other state databases that there are often gaps in sampling. If some years were represented with only samples late in the season and others only had samples early in the season, then seasonal trends could be misinterpreted as annual trends. How many samples were required per year to calculate an annual average and were there any additional measures taken to ensure that the data collected over the course of the season was weighted similarly between years (e.g., must have at least one sample in May/June, July/Aug, and Sept/Oct)? I think there should be some rules to this effect. If this is not a concern, please state some statistics showing that the majority (x%) of lake-years included in the analysis adhered to the full sampling schedule.

* **Response:** Good catch! Fortunately, it has not greatly impacted the analysis as we have full seasonal coverage for the majority of lake-years. That being said, we still decided to follow the reviewer’s suggestion and filter these out. This filtering step has been added to our data\_clean.R script at lines 157 and 398 and is in the filter\_months() function in functions.R. This removes any lake/year/parameter that does not have a minimum of one sample in May/June, July/August, and September/October for both the Watershed Watch data and the LAGOSNE data. The changes to the analysis are minimal and do not significantly alter any of our conclusions. We added text in the second paragraph of the Methods section describing the selection criteria.

***1-2.*** I also had questions regarding the way that interannual data were represented. At least 10 years of data from 1993 to 2016 had to be available for a lake to be included in the analysis (lines 121-122). This is reasonable. However, this is less than half of the years of the entire period of record. In theory, one lake could be analyzed from 1993 to 2002 and another lake from 2007 to 2016. These are not even overlapping periods of record. I think you need to take it one step further and ensure that there are years represented at the beginning and end of this 24-year period. Oliver et al. 2017 did include rules to this effect. I do not think your analytical method is immune to this possible difference in time frame. Again, please describe the time frames of your data set and adjust your rules if you find that you do have a substantial number of lakes with non-overlapping time frames.

* **Response:** Similar to the approach in Oliver et al. we have added an additional step to ensure that lakes are only included if they have data in the first and second half of the 1993 - 2016 time frame and now only lakes that have at least one measurement in 1993-2004 and 2005-2016, have a minimum of 10 years of data, and have at least 1 measurement in may/june, july/aug, sept/oct are included in the analysis. The filtering step has been added to the data\_clean.R around line 342 and 523 and uses a new function, filter\_early\_late() which is in the functions.R file. The impact on the analysis was minimal. We added text in the second paragraph of the Methods section describing the selection criteria.

***1-3.*** How did Secchi depth change? Why wasn’t this analyzed?

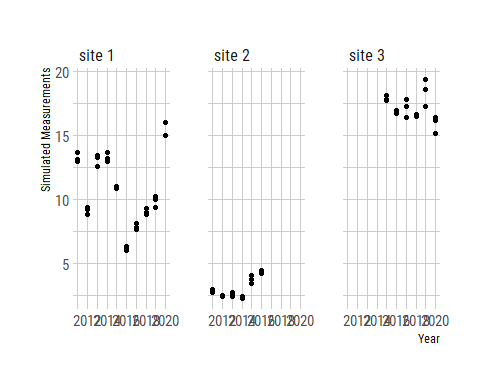
* **Response:** As we were comparing the results from our analysis in Rhode Island to the LAGOSNE data and specifically to the results found in the Oliver et al 2017 paper, we constrained the analysis to a similar set of parameters. The Oliver et al 2017 paper did not include secchi in their analysis and we decided to follow suit.

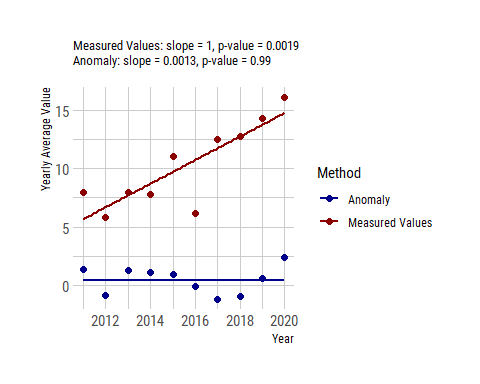
***1-4.*** It is unclear how the trends in temperature, TP, chl-a, TN, and TN:TP compare to water quality condition and any standards that might be set in Rhode Island. Tables 1 and 2 give the range of values across lakes and there is clearly a large gradient in trophic status. Following from the comment made in point 8, it would be helpful to better understand the trends observed here in context of water quality condition. If chl-a is rising, is it still relatively low, or is it reaching “polluted/degraded” levels? Do you observe a shift in trophic status in individual lakes? Somehow, try to put the trends or lack of trends observed in the context of how water quality condition is changing over time.

* **Response:** Another good suggestion. We have addressed this in our response to comment ***1-5b*** and ***1-8***. In particular, we have identified slope values that would indicate a meaningful ecological trend. We use literature values to inform those values. Additionally, brief, preliminary results were added in the discussion looking at coarse trends across trophic state.

***1-4a.*** Line 312: It is still unclear to me how this analysis of site-specific anomalies is robust to variation in timing of inclusion of given sampling locations.

* **Response:** What we mean by this, is if sites with differing base lines (e.g. low nutrient site vs high nutrient site or cool water site vs warm water site) were included at different times, then site-specific anomalies are more appropriate for examining long-term trends because each of the sites will be re-scaled, relative to that sites long term median. For instance, comparing measured values from a cool water site that shows no trend which is measured mostly early in a sampling period to a warm water site without a trend measured late in a sampling period would likely show an increasing trend when in fact none existed. Re-scaling these values relative to the median sets both sites to a common baseline and would then not show a trend. To further illustrate this, look at the two figures below. The first figure shows simulated data from three sites. All sites have no trend with site one being measured across the time period with values mostly in the middle of a range. Site 2 has no trend with low values and is only in the first portion of the time period. Site 3 has no trend with high values and is only in the later portion of the time period. The second figure shows what the average yearly trend would look like if just looking at the values vs looking at the anomalies. The measured values show a trend, when none should exist. The anomalies correctly show no increasing trend. We have added in similar figures and text to the Discussion





***1-5.*** My biggest concern is possibly overstating the lack of trends in TP and TN and the positive trends in temperature and chl-a and then suggesting that temperature, not nutrients, are responsible for increasing chl-a. The authors are somewhat cautious when making this conclusion, but do not fully recognize the limitations of this interpretation. I will try to break my concerns into manageable pieces for the authors to respond to.

* **Response:** We address the reviewers concern regarding the possible association between rising temperature and increasing chlorophyll in the individual pieces below. In short, we concur and have done all we can to avoid the appearance of suggesting causation.

***1-5a.*** I think the title overstates this result. I agree with increasing Chl-a trends, but don’t agree with stable nutrient trends.

* **Response:** Changed title. It no longer focus on the trends. New title is: “Analyzing and comparing long-term water quality of lakes and reservoirs in Rhode Island and the Northeastern United States using an anomaly approach”

***1-5b.*** I have not read Wasserstein et al. 2016, but agree with the general concept that a specific p-value threshold is not the only weight of evidence for determining whether there is a trend. However, I do not think the authors conclusions about trends for each parameter in the Rhode Island vs. LAGOS datasets are consistent with one another need a more detailed framework laid out ahead of time to understand the more holistic way of interpreting whether there are trends. These are the trends that I interpret differently than do the authors:

* **Response:** We like the reviewers concept of a more holistic way of interpreting trends. We attempted to do this in our initial assessment of these trends, but we obviously were not successful in doing so. We now lay out a more explicit approach and describe how we use magnitude of the slope, relative frequency of high/low years in the beginning and and end of the time frame, and the p-value together to assess for trends. This is described in the paper in a new section in the methods, *Assessing regressions for trends*. One of the biggest changes is to suggest values for the slope that would indicate an ecologically meaningful trend. We describe this in the new section, but in general we are using a trophic state change from oligotrophic to eutrophic over one century to indicate ecologically meaningful change. For temperature we look at values associated with ecological change as well as policy values (e.g. the Paris Agreement). While these exact numbers are not universal, we use them as a starting point for assessing the trends. How we interpreted the trends relative to the additional specific issues raised by the reviewer are detailed below and have changed somewhat from our initial submission, although the broad patterns still hold.

***1-5b-i.*** TN in Rhode Island (Fig 5A) – The p-value is very small and the slope seems reasonably large. I also see more positive means later in the record. I do not think that 1993 and 1994 have an overly strong effect on the trend. Even when you take them out of the analysis, the slope is still 2.5 and p-value is small. Although the annual increase in TN is not large compared to the range of total TN anomalies, I interpret the stats as an increase in TN. In line 247, I disagree with the statement that removing data from 1993 and 1994 weakens the positive trend “considerably”.

* **Response:** The results for TN in Rhode Island have changed somewhat since we added the additional filtering steps suggested by Reviewer 1 (**1-1** and **1-2**) and made changes to use the median instead of the mean in our response to comment **1-6**. These changes have weakened the positive trend as the slope is smaller and the p-value larger. Additionally three years, 1995, 1996, and 1997 now have insufficient data to be included in the figures and a third, 1998, has only three sites. Using the more explicit definition of a trend we now conclude that TN showed neither a statistically positive trend or an ecologically important trend over the time frame of a century as the slope is less than the amount needed to cause an trophic state change. This is reflected in the manuscript.

***1-5b-ii.*** TN in Lagos data set (Fig 5B) – the text says that “TN showed a slight positive trend”. Here, the slope is smaller and the p-value is large. I would think that the positive slope is driven by the high TN values in 2013. I think there is more evidence for increasing TN in the Rhode Island data set than the LAGOS data set, but the text reads the opposite.

* **Response:** We applied the same within and across year filters to the LAGOSNE, used the median instead of the mean (see **1-6**) and these results have changed as well. Given that and our method for interpreting the trends leads to the following: LAGOSNE is showing almost no statistical support for a change in nitrogen with above and below average years showing little pattern. The slope is much less than the 3.5 we identified. Thus we conclude no trend in TN at the scale of the Region. This is indicated in the manuscript.

***1-5b-iii.*** TP in Rhode Island (Fig 6A) – The authors state that TP does not show a trend over time. I see a positive slope with a marginally significant p-value and more positive means later in the record. I wouldn’t oversell this as a major increasing trend in TP, but I see some evidence for a slight increase in TP.

* **Response:** Again, with the changes to the filtering and the switch to the median the trend lines have slightly changed. TP is showing no trend with no statistical support, a very small slope and above/below average years are relatively evenly distributed. We have indicated as such in the text.

***1-5b-iv.*** TP in LAGOS (Fig 6B) – “TP showed a slight decreasing trend”. I disagree. The slope was negative, but at -0.027, so a much smaller magnitude than the positive slope in the Rhode Island data set. The p-value was 0.32 and the graph itself looks like the slope is a horizontal line centered on 0.

* **Response:** Small changes due to filtering and new criteria for assessing the trend are also seen here. There is statistical support for a trend, that trend is positive, but the magnitude of the slope does not support the idea of an ecologically significant change in TP over a century. This has been added to the manuscript.

***1-5c.*** To put the magnitude of each slope in context, you could report the slope as a percent of the full range of slopes observed in the data set. Even those trends that are “significant” have a relatively small slope compared to the variability between lakes.

* **Response:** This is a good suggestion, but we have chosen not to implement it, primarily because of Reviewer 1’s suggestion to have a more holistic approach to addressing the trends. In developing our approach to assessing the trends we now have ecologically relevant thresholds we can compare to. Both approaches address the same good suggestion of putting the magnitude of the slopes in context.

***1-6.*** This analysis also seems to be saying something about coherence. In some years, there is a wide range of anomalies, meaning that some lakes are much higher than their long-term average at the same time as other lakes are much lower. This seems like an interesting result in and of itself that is not pointed out in the results or discussion. How does the variability in annual anomalies of water quality between lakes compare to that observed in the climate literature, which traditionally uses this approach?

* **Response:** We were a little unclear about this comment. Is it referring to coherence over time within a single variable or coherence between the variables? We agree that for individual variables some years do show greater variability than others. These increases in the yearly standard deviations are largely the result of a few unusual measurements in those years for those variables. Essentially these are outliers, but ones that have no obvious reason to remove from the analysis so they are kept in. Further, there is no pattern to these outliers as they are from different ponds, aren’t seen across variables in the same years, and in the case of the LAGOSNE data, they come from different states. There is only one exception to this and that is the high variability in 2011 for both TN and TP in Rhode Island. That is driven by one sampling event where TN and TP were high for both. Given the lack of a pattern, there is little additional inference we can draw from this, other than the analysis might be susceptible to outliers. Given that our original analysis is based on the mean and std. dev. and that those statistics are sensitive to outliers we have decided, thanks to this comment, to change to median and the inter-quartile range, both of which are considerably more robust statistics in the presence of outliers. With this change the swings in variance are no longer seen as would be expected with statistics that are robust to outliers. Good catch, by the reviewer.

***1-7.*** In the discussion (lines 236-237) and in the title, the authors seem to want to conclude that because there were positive trends in temperature and chl-a but not in TP or TN (which I don’t agree as strongly with), that increasing chl-a is likely caused by increasing temperatures and not nutrients. This is certainly interesting if it were the case, but it is well beyond the scope of what can be concluded here.

* **Response:** We have softened our language in the discussion on original lines 236-237 considerably. They now read, “While our analysis is not capable of detecting causation, that both chlorophyll *a* and temperature is increasing with less obvious trends in nutrients is interesting and warrants further exploration to see if increasing chlorophyll *a* can be described by temperature.”

***1-7a.*** This study simply lists the direction of trends of 5 parameters, but does not analyze correlations between them.

* **Response:** Agreed. Language softened and title changed to hopefully steer readers away from causation.

***1-7b.*** Even if it had tested correlations between parameters, it is still difficult to reach for causation.

* **Response:** Agreed. Language softened and title changed to hopefully steer readers away from causation.

***1-7c.*** The analysis of mean anomalies over time is interesting, but that simply means that the entire set of lakes in the state shows trends over time. What I’ve learned from reading some of the papers cited by this study and from my own investigations is that lakes near one another can have opposite trends. Processes at the scale of the individual lake and its watershed are important and can have opposite trends in neighboring watersheds. I think this study would benefit by adding an analysis at the individual lake scale to compliment the state scale. I would feel more comfortable with the statement that increasing chl-a may be a result of warming waters, not nutrients, if this was also tested at the lake scale. One could build a mixed effects model:

* Chl-a anomaly ~ temperature anomaly + TP anomaly + TN anomaly + lake

where lake is a random effect that has a unique id for each lake in the study. Years would be replicates. I think it’s possible that you could see a nutrient effect if you examined the data this way. There is a lot of variability between lakes within a year, so while I find the statistical method used in this study interesting, I don’t think it tests of drivers of change.

* **Response:** We agree with the reviewer that lake scale trends and the drivers behind those trends are both very interesting questions and something our approach could contribute to; however, we feel that it is beyond the scope of this particular paper. Given the changes we have made to the title and the stated goals of the paper we feel that adding this type of analysis would greatly expand the paper and take away from our primary purpose of looking at long term trends in Rhode Island and the Northeast/Midwest. We do feel this is a very good suggestion for next steps and my co-authors and I are in active discussion on how to move forward with such an analysis. In the Discussion at the the end of *Trends* we briefly mention this application and that the anomaly approach could also be used at the individual lake level to assess for trends.

***1-8.*** Line 274-275: the authors pose the interesting question about whether lakes of similar trophic state are showing similar trends. I disagree that this test is beyond the scope of the study. I think it is a timely question that would be good to test following the Stoddard et al. paper. It should be relatively easy to test – dividing the lakes into groups of hypereutrophic, eutrophic, mesotrophic, and oligotrophic categories based on their long-term mean TP or chl-a. If there are not enough lakes in each category, with would be interesting to compare a few.

* **Response:** We agree that this is indeed a timely question and fairly easy to test. Initially we were going to include this analysis in our revisions and did some preliminary analysis. There were differences in trends across trophic state with many of the water quality variables and to fully address this would require significant additional analysis and discussion. After looking at these results and with quite a bit of discussion among the co-authors we do still feel that this type of analysis is 1) beyond the scope of this paper and 2) worthy of its own paper. We have added to the Discussion under *Trends* about this and briefly address preliminary results for differences in chlorophyll trends across just oligotrophic and hypereutrophic lakes. Our internal conversations are ongoing about completing this analysis for a subsequent paper.

## Editorial Comments

***1-9.*** Line 64: Better define the spatial scales of the studies referenced in this paragraph. Macro-scale, regional, subregional could be interpreted in different ways. Please be explicit about the geographic region that each study includes.

* **Response:** Added information on extent, plus entire paragraph was re-written

***1-10.*** Line 71: I’m not sure why it was surprising that little change in nutrients and chl-a was reported in Oliver et al. 2017? How is this study different from Lottig and Collins? All three use the LAGOS data set. I think it would be helpful to be more specific about both the spatial extent and the results of these studies – particularly when you talk about complexity in discussing the Lottig and Collins papers. Complex just doesn’t explain the patterns very well – please be more specific in describing what complex means in this case.

* **Response:** This paragraph was re-written and better describes the papers.

***1-11.*** Lines 99 – 102: please delete. Unnecessary.

* **Response:** Fixed

***1-12.*** Line 113: capitalize Secchi

* **Response:** Fixed

***1-13.*** Line 123: delete approximately. It seems that you are giving specific numbers here.

* **Response:** Fixed

***1-14.*** Lines 202-207: The text simply repeats the numbers reported in tables 1 and 2. It would be more interesting to compare the populations at the two scales – in terms of the means and the ranges and the variety of trophic states represented and to also summarize the comparison of trends over time at the two scales.

* **Response:** Added an additional paragraph to the results that compares the two scales and includes a new table summarize the percentage of lakes at both scales in chlorophyll based trophic states. Another table was added to the trends sections of the results summarizing how we interpreted the trends at each scale.

***1-15.*** Line 285: “receive feedback” is a vague phrase here. Perhaps you mean that is important for managers and stakeholders to know what the long-term trends are? I think you could also add more references and specific language about why it is important to share results with volunteers.

* **Response:** We changed the wording on this to reflect a focus on the need to know about long term trends. Also we added three citations that specifically address that sharing results back with volunteers meets their expectations and can motivate continued involvement.

***1-16.*** Line 292: I was a little bit confused by this paragraph. Why discuss ways of reducing nutrients when you have not identified increased nutrients as having an increasing trend? There are a lot of leaps made to get to this paragraph – temperature affects chl-a more than nutrients, resultant chl-a and nutrient values are too high, lower nutrients can offset warming temperatures, ways to lower nutrients. Some of the analyses I suggest adding above may better motivate this paragraph being included in the discussion section.

* **Response:** This section was re-written. We still mention nutrient reductions, but frame it as long-term trends, like those seen in RI, can help develop new hypotheses and inform existing ones.

# Reviewer 2

## Major Comments

***2-1.*** The paper needs some contextual description of the RI lakes that could be achieved from GIS analysis. For example, what is the distribution of size and depth of the lakes in the data set? What are the land-use area percentages (e.g. agriculture, residential, and forested areas) within the catchment areas of each lake in the data set? A contextual description will be of value to other researchers who may wish to compare these results to other areas. Smaller lakes potentially could show a different warming trend than larger lakes, for example, and therefore knowing a lake size distribution could help qualify the results. Some context on the land-use patterns is also important for understanding the nutrient-level data, for comparing results between study areas, and perhaps shedding some light on why the Rhode Island mean water quality values are less than mean values for the larger, LAGOSNE dataset (p. 11, lines 205-207). At a minimum, an overview description of the lakes that make up the dataset is needed.

* **Response:** We added a summary table that summarizes average lake area, maximum depth, and average percent forest, ag, and developed for the LAGOSNE and URIWW. Some additional text is in the methods as well as in the results. Thanks for the suggestion. There are a few differences between the regions and this is a useful addition to the paper.

***2-2.*** There should be no reason to rely on “anecdotal evidence” for explaining warming air temperatures over the timeframe of the investigation (p. 14, lines 269-272). There is no source/reference for this claim. Weather station data for RI should be available to assess whether regional climate is warming and to see if it is consistent with the observed increases in lake temperature.

* **Response:** Thanks again. We have cleaned up this sentence to better reflect a focus on lengthening growing season and also refer to a source of data (NOAA Climate Time Series) that make this point as well.

***2-3.*** Some clarification on the methods is also needed. The authors say that they used the same analytical methods for the LAGOSNE data set as they did for the Rhode Island lakes. Please verify whether the LAGOSNE data represent the same seasonal timeframe (May – October) and lake sample depth (< or = 2m) as the Rhode Island data (as they should for a comparative analysis). Additionally, please state why the LAGOSNE data extend only until 2014 vs. the year 2016 for the RI dataset.

* **Response:** In the paragraph about LAGOSNE, we provided some additional clarification. We use the same selection criteria on both datasets to make sure the time frame is comparable. The data for LAGOSNE are reported to all be from the surface or epilimnion. While not a direct match to the depth filter we used for URIWW, they are roughly comparable. Data for LAGOSNE do not extend beyond 2013, hence the truncation. This is all reflected in the manuscript.

## Editorial Comments

***2-4.*** p. 6, line 90: Capitalize “States.”

* **Response:** Fixed.

***2-5.*** p. 10, line 188: Check spelling of “parameter.”

* **Response:** Fixed.