Clim. Past Discuss.,
https://doi.org/10.5194/cp-2020-70-SC1, 2020

Climate of the Past Discussions

Interactive comment on "Long term trends in aquatic diversity, productivity and stability: a 15,800 year multidecadal diatom study from Lake Baikal, southern Siberia" *by* Anson W. Mackay et al.

Gavin Simpson

gavin.simpson@uregina.ca

Received and published: 30 June 2020

Mackay and colleagues present data on diatom abundances and associated data from Lake Baikal, with a focus on long term trends in the diversity, productivity, and stability of the diatom record.

I have a couple of technical comments to make regarding the methods used, which speak to the appropriateness of the methods and therefore the evidence they provide in support of the authors interpretations.

My first comment relates to the diversity measures and how they have been handled. Hill's numbers, like other count, or count-based, measures are affected by sampling effort; all else equal, the greater the sampling effort the greater the diversity, the greater the Hill number. In palaeolimnological studies, sampling effort concerns at least two elements of the data collection and analysis process: i) the sample count, here $\sim$300 valves; and ii) the sedimentation processes, accumulation rates, and depth-based sampling that generated the sediment slices that were counted. The authors appreciate these issues and employed methods to handle varying-effort source 1 via a bootstrap approach.

On L234 the authors state the sediment accumulation rates range between 34 and 133 years per cm, which means that each of the 0.5cm slices contains between 17 and $\sim$66 years. As such, we might expect variation in Hill's numbers simply due to systematic changes in accumulation rates over time. The analyses described here do not appear to account for this element of potential bias due to varying sample effort.

It is unclear to what extent variation in accumulation rates might contribute to or obscure the underlying trends in diversity metrics in the core as accumulation rates are not presented alongside the time series of diversity metrics.

It is also unclear exactly how best to accommodate the varying accumulate rate effect in the rarefaction estimate used by the authors. One option might be to simply scale the results of the rarefaction/bootstrap by the number of years that each sediment sample represents. While simple, this solution might not fully account for the problem if the Chao et al (2014) method makes assumptions about equal effort beyond sample count totals.

A more complicated alternative might be to not use any rarefaction at all, and instead model the observed data using a regression approach, with an offset that includes both the count total and the number of years per slice. For the regression model itself, a GAM seems appropriate given the non-linear change in the metrics over time.

For N0, richness, a Poisson or negative binomial model might be a starting point, and then an offset that is 'log(sample_count * n_years)', where 'sample_count' is the number of valves counted in each sample and 'n_years' is the number of years that each sample represents, should result in a model with the correct normalization to expected number of species per valve count per year. The offset used could be varied so that it is terms of expected richness per 100 valves per decade, by suitable scaling of the values inside the 'log()'.

A similar approach for N1 and N2 is little less clear; Tweedie, gamma, or inverse Gaussian distributions for example can all handle non-negative (Tweedie) or strictly positive (gamma, inverse Gaussian) real values which N1 and N2 result in, and each of these models can include an offset in the same way and, assuming a log link function is used, on the same log scale. But I'm not familiar enough with the details of the gamma and inverse Gaussian models as GLMs to comment further on the exact interpretation of offsets in those models.

The second area I wish to comment on is in the interpretation of the moving window coefficient of variation (CV) results, especially Figure 4. What is being shown in Figure 4? CV is a unit-less variable, but each plot has a label on the y-axis. In the upper panel we see variation in the CV but most of these changes are relatively small suggesting CVs of between 9 and <3 % of the mean. Are these values biologically meaningful? Is the fact that we see CV of log(BVAR) declining in the Late Glacial and early Holocene, really an important decline? The N2 signal seems to show more variation; apart from the Late Glacial period, CV values are between 10% and 50% of the mean N2. Those numbers seem, however, at first sight to indicate that the stability of N2 is much less than log(BVAR). My comment really relates to one of whether palaeo productivity stabilised, and if so, how does that mesh with the interpretation that N2 was also stable for much of the Holocene where the variability in CV values for N2 is so much higher than for BVAR?

The Kruskal Wallis analysis is non-parametric, but it is not without assumptions. The

principal assumption that is certainly violated here is an assumption of independent observations, because your data are a time series. I don't know if there is a way to correct the p-values for the loss of degrees of freedom due to the dependence between samples (as is often done for regression, assuming AR noise), but if not, one might have to use a permutation or resampling/bootstrap procedure where the permutation or resampling is done in a manner that preserves for correlation structure (ruling out simple permutation and simple bootstrap resampling).

My final comment relates to the correlation analysis as shown in Figure 5. What type of correlation are you showing here? All the standard correlation coefficients are bounded -1, 1 but the plot shows values >1. I suspect what you're showing is the coefficients from a linear regression through the scatter of productivity and diversity data in the 1000 year moving windows? This needs to be fully explained in the methods. Putting an informative label on the y-axis would also help the reader understand this figure.

Another problem with Figure 5 is the use of the p values encoded as colours on the plot. It's impossible to tell what the p-values are at the low end, where the interest lies, because of the colour scheme used. Also, you (or the reader) are at risk of making a massive multiple comparisons mistake here. If you are going to use the p values then you would need to correct them for multiple comparisons (i.e. for as many tests as the number of data points shown) using the false discovery rate to adjust the p values. Then you could use a binary indicator perhaps to show which values remain significant.

Other questions remain with this correlation analysis;

* is a linear model in a moving window a good fit to the data?

* if relying on the p values, how biased are these because neither the response nor the covariate are independent,

* if this is a linear regression, which variable played the role of a dependent and independent variables?

* if this was a linear regression why assume all the error is in the dependent variable?

I think it would be better if what was plotted was a true correlation value with bounds at -1 and +1 as that is a scale that most readers will be most familiar with. And in that case you probably wouldn't need to use or present the p-values.

Regarding the principal curves, how confident are you that the data are well represented by a single gradient? The eigenvalues of the CA that you used for the starting curve would be a guide as to whether there is remaining structure in the data on axes 2 and higher. PrCs are really useful when there is a strong single gradient in the data, but when there are secondary gradients my experience is that they are much less useful and can get stuck in some weird solutions that don't make ecological sense. You can get an idea of this by looking at the complexity of the smoothing splines fitted to individual taxa; if these are not simple linear, monotonic, or unimodal curves than it is more likely that the PrC is being asked to do too much and is a good sign of problems with the fit. I mention this because of the very rapid changes in the PrC scores, which can happen when the PrC itself is too complex; it would be useful for the reader to see the 2d ordination with the PrC superimposed on it, perhaps in the supplementary materials?

Finally, throughout there are few attempts to quantify the uncertainties in the quantities you estimate, interpret, discuss, and present to the reader, or to compare the observed results with appropriate null models of no change/trend. This makes it difficult to gauge the overall support of your interpretation that comes from the data and the analyses.

Minor Comments

- L189: there's an extra parenthesis before "De'Ath"

- L215: here and throughout, the superscripts in your units appear to have gone missing, perhaps during the conversion to PDF?

- L224: here and throughout, the way you present the Hill's numbers changes through-

out the manuscript. Sometimes the "N" is in italics, sometimes not, and sometimes the number is in italics and sometimes not.

- L243: the R^2 should be upper case R and the 2 is not superscript

- L294: delete "as" between "declined" and "rapidly" or change the sentence to indicate what BVAR changed as rapidly as.

- L501 add "zone" after "photic"

References

Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., Ellison, A.M., 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecol. Monogr. 84, 45–67. https://doi.org/10.1890/13-0133.1

Interactive comment on Clim. Past Discuss., https://doi.org/10.5194/cp-2020-70, 2020.