

## Response to Anonymous Referee #2

We thank the reviewer for their constructive review. Our responses to their comments are in red below.

### Overview

Zander et al. use recently developed hyperspectral imaging methods, including one first introduced in this manuscript, to quantify the abundance of sedimentary photopigment produced by oxygenic phototrophs and anoxygenic phototrophs - specifically purple and green sulfur bacteria (PSB and GSB, respectively). They apply this technique to Lake Cadagno, a lake that has been well studied for both modern and paleolimnology. The authors find that total chloropigments (TChl) and purple sulfur bacteria (PSB) are readily detectable using key absorption troughs in hyperspectral imaging, and calibrate their abundance using concentrations measured on samples using extraction and spectrophotometry and HPLC. They argue that GSB are detectable by a characteristic absorption trough, and around 700 nm, however it is difficult to calibrate as the pigments are not readily measured by other techniques. Instead they use co produced carotenoids to validate the detection of GSB, but do not attempt a calibration, instead they interpret the index as an indicator of relative abundance.

The detection and interpretation of these pigments by hyperspectral imaging has great potential for paleolimnological studies. HSI methods are rapid, high-resolution and non destructive, and analyzing these pigments provides important and direct inference on the past light and oxygenation structure of the lake. GSB is also quite difficult to measure, and it's possible that in situ detection via HSI may actually work better than other approaches, although this is difficult to prove. In the case of Lake Cadagno, the pigment records provide important insights on both long-term lake evolution, and response to external forcings and pressures, but also the response of the lake to flood events, highlighting the value of the high resolution detection.

The manuscript is well written and well illustrated, and effectively articulates the importance and value of the approach. It's an important contribution to the paleolimnologic literature, and I recommend it for publication following the correction of a minor concerns.

### Minor concerns:

#### Non-normal distributions of pigment and RABD data

My only substantial concern with analysis and the results has to do with the treatment of the data in the calibration and reconstruction approaches. Specifically, I worry about non-normality in both the the pigment concentrations measured by spectroscopy and HPLC, and in the RABD indices. Both datasets are left bounded, and tend to be skewed right. Some strongly so. On line 236 the authors note that this may hamper their interpretations. The authors use linear regression to relate RABD indices to TChla and PSB, and to quantify their uncertainties, but this approach assumes normality in both the predictor and predictand. Looking at figure 3, this seems reasonable for TChl-a and RABD670, but the PSB data look right skewed, and potentially bimodal. And looking at figure 5, all the downcore HSI data, and especially the GSB index and PSB reconstruction both look left bounded and right skewed, as expected. I'd

encourage the authors to either demonstrate normality for these data, or consider transforming them before calibration, or use a different approach that does not require normally-distributed data.

I believe this is a common misconception. Normality of X and Y variables is not required for linear regression, the only requirement for normality is in the residuals of the regression models (Poole and O'Farrell, 1971)

Shapiro-Wilk normality tests on the residuals of both regression models used in this study yield p-values > 0.05, suggesting there is not significant evidence to suggest that the model residuals are non-normal. We will include a plot of the residuals and their distribution in the revised supplement.

However, the reviewer is correct in that non-normality may be a problem for the significance test we used for the correlation, i.e. the p-values reported in Fig. 3. We have now calculated the correlations using a non-parametric permutation test, which does not require an assumption about normality of the variables. Furthermore, we adjusted the p-values for the false discovery rate of multiple hypothesis testing using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). The results of the permutation test using 10,000 permutations show that the correlations remain significant for all relationships shown in Fig. 3. We will update the methods section to describe the updated statistical methods.

For the Bphe-a and RABD842, I wonder whether the data really support a continuous calibration, or whether a on/off, presence/absence type calibration would be more appropriate. To my eye (Fig 3A), there are basically two populations of Bphe-a concentrations. Those close to zero, and those around 1000 ug/g. The lower concentration population does suggest a relationship, but the trend is notably lower than that inferred from the dataset as a whole. The two highest Bphe-a concentrations do also stand out, but it's hard to know how much to weight those outliers. I'm open to the argument for a continuous relationship, but I think it needs to be made as it seems plausible that we're mostly seeing presence/absence of PSB in these data.

The reviewer is correct in pointing out that there is some evidence for a discontinuity in the linear relationship between Bphe-a concentrations and RABD842. However, I would not agree that it is a presence/absence type of calibration because there is a considerable range of concentrations in the lower population (0-400 ug/g), which appears to follow an approximately linear relationship based on a limited sample size, although it is true that a linear model fit to only this population would have a lower slope than the general model.

It is possible to consider that a non-linear model could be used to fit this relationship better, and indeed a quadratic polynomial does improve the model fit ( $R^2 = 0.938$  vs  $0.920$ ). Exponential and power models yield a poorer fit. Given that the improvement of the quadratic model is rather small, we chose to use a linear calibration model based on the reasoning that linear calibration models have previously been the dominant choice for calibrating sediment reflectance absorption trough indices and pigment concentrations (Rein and Sirocko, 2002; Wolfe et al., 2006; Butz et al., 2015; Zander et al., 2022).

It is difficult to explain the apparent jump in pigment concentrations observed in the scatterplot at RABD values around 1.3-1.4, but given measurement uncertainties, limited number of samples, and other sources of error (e.g. discrepancies between pigment abundances at the

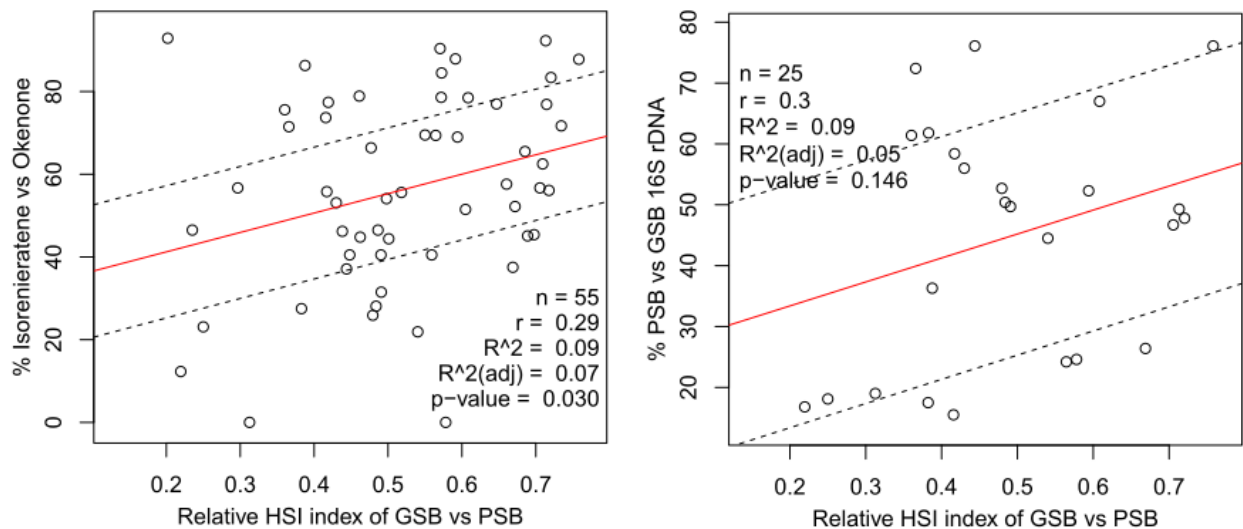
core surface compared to the underlying sampled sediments), I would not overinterpret this feature.

### Other Concerns

Figure 4.

The comparison between the aeDNA data and the PSB and GSB indices is very interesting. I like the presentation in figure 4, but would be very keen to see a direct comparison of the aeDNA sample data and the HSI-based ratio. Could you add a scatter plot where the PSB/GSB indices over the DNA sample depth ranges are averaged and the plotted against a similar metric in the DNA data? There are intervals where it seems to agree quite well, but others where it doesn't, and I'd like to see the comparison explicitly, while also accounting for the different resolution of the datasets.

Based on the reviewer's suggestion, I have made this plot. The relationship between the relative HSI index for BSG vs PSB and the 2013 carotenoid pigment data is statistically significant, but with a low  $R^2$ . Comparing the relative HSI index with the DNA data, the relationship is not statistically significant. This comparison is challenging because the samples for carotenoids and DNA in the 2013 study were taken from different cores, not the same ones as the HSI. The correlation of the sample intervals to the hyperspectral imaging data may include some mismatches, in particular in sections with many floods and slumps, which can have different thicknesses in the two cores. Furthermore, several of the 2013 samples include substantial material from mass movement deposits (slumps and floods). The mass movement deposits tend to have much lower concentrations of pigments, making the data from HSI in these sections noisy and mostly unreliable. Additionally, the 2013 carotenoid data and DNA are reported only as relative percentages of GSB vs PSB material. It is not trivial to relate that type of data to the hyperspectral indices, which are related to concentration, but have different relative scales due to the different absorption of the pigments. We attempted to account for this by standardizing the RABD indices to a 0 to 1 scale, and then taking their ratio. Altogether, these challenges make it not possible to compare the two datasets in an "apples to apples" manner, and therefore, it is difficult to interpret the results of these scatterplots, and we do not plan to include them in the manuscript.



As a small note, in Figure 4 the HSI data shows only data from regular background lacustrine sediment, all mass movement deposits have been removed, whereas in the above scatterplots, the mass movement deposits were not removed if they were also included in the carotenoid/DNA samples from the 2013 paper. We will add some text to explain the possible reasons for discrepancies between our data and the 2013 data.

Figure 5. Add a Tchl label to panel A to match panels B and C.

The label will be added

Figure 6. Continuing the Tchl, PSB and GSB labeling on the curves would benefit figure 6 too.

The labels will be added

## References

Benjamini, Y. and Hochberg, Y.: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289–300, <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>, 1995.

Butz, C., Grosjean, M., Fischer, D., Wunderle, S., Tylmann, W., and Rein, B.: Hyperspectral imaging spectroscopy: a promising method for the biogeochemical analysis of lake sediments, *Journal of Applied Remote Sensing*, 9, 096031, <https://doi.org/10.1117/1.jrs.9.096031>, 2015.

Poole, M. A. and O'Farrell, P. N.: The Assumptions of the Linear Regression Model, *Transactions of the Institute of British Geographers*, 145–158, <https://doi.org/10.2307/621706>, 1971.

Rein, B. and Sirocko, F.: In-situ reflectance spectroscopy - Analysing techniques for high-resolution pigment logging in sediment cores, *International Journal of Earth Sciences*, 91, 950–954, <https://doi.org/10.1007/s00531-002-0264-0>, 2002.

Wolfe, A. P., Vinebrooke, R. D., Michelutti, N., Rivard, B., and Das, B.: Experimental calibration of lake-sediment spectral reflectance to chlorophyll a concentrations: Methodology and paleolimnological validation, *Journal of Paleolimnology*, <https://doi.org/10.1007/s10933-006-0006-6>, 2006.

Zander, P. D., Wienhues, G., and Grosjean, M.: Scanning Hyperspectral Imaging for In Situ Biogeochemical Analysis of Lake Sediment Cores: Review of Recent Developments, *Journal of Imaging*, 8, 58, <https://doi.org/10.3390/jimaging8030058>, 2022.