For VRS inferences of chlorophyll *a*, we now use a log transformation of the data given in Michelutti et al. (2010).



Figure 1. Log transformed plot of the calibration data reported in Michelutti et al (2010).

The log transformation allows all the data points (n=35) to be retained and satisfies the assumptions of linear regression with respect to the heteroscedasticity (Fig. 2) and normal distribution of residuals (Fig. 3). Using this log transformation, we infer sediment chlorophyll a concentrations with the

equation: Chlorophyll a + derivatives =EXP(0.83784\*LN(peak area650-700 nm)+(-2.48861))



Fig 2. Residuals of regression (Fig. 1) using log transformed data



Fig 3. Histogram of residuals from the log transformed regression in Figure 1. (Shapiro-Wilk normality test W = 0.97649, p-value = 0.6428).

Compared to the equation in Michelutti et al (2010) using untransformed data, there is little-to-no difference in inferred chl-a values (see examples below). Importantly, chl-a profiles generated from the updated equation using log transformed data would not result in any changes in interpretations.



Figure 4. Comparison of spectrally inferred sediment chl-a concentrations between the new log transformed regression equation and that given in Michelutti et al (2010) for Apicocha Pond, Ecuador (Giles et al. 2018; Freshwat Biol).



Figure 5. Comparison of spectrally inferred sediment chl-a concentrations between the new log transformed regression equation and that given in Michelutti et al (2010) for Grape Tree Pond, Jamaica (Heller et al. 2021; The Holocene).



Figure 5. Comparison of spectrally inferred sediment chl-a concentrations between the new log transformed regression equation and that given in Michelutti et al (2010) for Meretta Lake, Nunavut.

I would write something like this for citation:

Sediment chlorophyll a concentrations were inferred using log transformed data from Michelutti et al. (2010) with the equation:
Chlorophyll a + derivatives =

On rare occasions, using the transformed data can result in NA values of inferred chl-a caused by a domain error. This occurs when the peak area (calculated by taking the area under the curve – or integral – of the spectra points between 650 and 700) is negative. Then the formula

breaks down as the negative numbers are not in the domain of the natural logarithm. Simply, this equation becomes NA. Changing the inferred value to zero is not a mathematically sound solution:

If the vrs inferred value for chl-a is 0, this implies that the peak area under the curve is greater than or equal to zero, even though the spectrometer may report peak area as a negative number. Changing the inference to 0 means that peak area is artificially inflated and perhaps even has a sign change in these occurrences. This is a bit of a gray area since “below detection limit” is not exactly a solution either since the spectrometer is reporting values…