

Interactive comment on “Interacting effects of climate and agriculture on fluvial DOM in temperate and subtropical catchments” by D. Graeber et al.

Anonymous Referee #1

Received and published: 15 January 2015

Understanding the effects of agriculture, at different levels of intensity and in different biomes, on natural organic matter character and quantity is important for predicting how changing land use and climate regimes are likely to impact biogeochemical cycles. This paper makes a thorough and competent effort to unravel the effects of farming intensity and climate on NOM composition. I especially appreciate the attention to detail throughout the methods and high quality figures.

I have some comments in relation to the statistical treatment of the data following their transformation to ratios. Addressing these comments could change some results in the paper. Especially I expect that the PCA might map differently. I have based these com-

C5

ments on my understanding that the raw data for CDOM and HPSEC were converted to ratios before further statistical analysis was performed.

(1) Indicate sample sizes for statistical tests. No sample sizes are mentioned anywhere in the paper.

(2) It seems to me that after converting some data to ratios, the statistical analysis has proceeded in much the same way as it would have for the raw dataset. Some, but not all, variables were transformed to improve their distributions (relative to the assumption of normality), then ANOVA, PCA were performed. However, methods for statistical analysis of compositional data /ratios are special due to the constraint that the data sum to one (closed data/constant sum constraint). There is a whole field of multivariate statistical analysis devoted to the analysis of compositional data, e.g. in the field of geology. There is an R package called *composition*, and several other packages, specifically directed at analysing compositional data (incl. imputing missing data). The *logratio* transformation is often used prior to linear modelling. See papers by J. Aitchison starting in the 1980s. Also a very readable R tutorial about the problem with ratios at <http://advan.physiology.org/content/37/3/213>.

(3) Currently, some data used in the PCA are bounded by [0 1] and some are not (e.g. fluorescence index), but overall the dataset does not sum to 100% (as it would in a typical compositional dataset). This does not sound like a good situation for starting a PCA. A simple approach would be to autoscale the raw (not compositional) data prior to PCA (transformation of some variables might still be advisable), which takes care of differences in scale between different variables, produces readily interpretable plots, and has other useful properties as described by Bro and Smilde (2014) in their recent PCA tutorial. The autoscaling will allow the PCA to reveal compositional differences between samples, which was the motivation for generating ratio data.

(4) Consider also the underlying assumptions of ANOVA, box and whisker plots and other statistical representations in the analysis of ratio data. When comparing ra-

C6

tio/percentages, it is common to arcsin transform the data first or use a chi-squared test.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 135, 2015.