

Review of Xu Fei, E. and Harman, C.J. Technical Note: A data-driven method for estimating the composition of end-members from streamwater chemistry observations, HESSD.

Xu Fei and Harman present a method to extend classical end-member mixing analysis by deriving end-members purely based on stream concentrations. By repeatedly delineating the convex-hull around stream concentrations and then classifying results using a k-means clustering approach the method includes an uncertainty assessment of the resulting end-members. The method is successfully applied to the Panola Mountain Research Watershed data.

I consider the outlined method a valuable addition to the end-member mixing literature and recommend publication in HESS after revisions are made that are outlined below. I must admit I am not a math expert and have not checked the given formulae.

Current application of end-member mixing approaches involves the disentangling of stream water concentrations based on pre-conceived 'end-members', or origins of stream water. These end-members are sampled, and hopefully span the spread of stream water concentrations, thus enabling the calculation of flow route proportions. Xu Fei and Harman propose a method that calculates possible end-member concentrations from stream water alone. While this can indeed be very useful in practice, it also defies the purpose of an end-member mixing analysis to some extent: as a hydrologist, we are interested in where the water comes from that makes up the stream, not in its concentrations per se. I am sure the authors agree here, but I miss more discussion in the manuscript on the practical use of the proposed method in a hydrological analysis. Given the end-goal of finding water sources instead of end-member concentrations, where does their method come into play? As a first step, defining possible concentration profiles of end-members, after which you take to the field to 'find' these end-members? Or as a check of a more classical end-member mixing application: are end-members missed? Or vice-versa, as the proposed approach cannot handle end-members that are located outside the convex-hull of stream water concentrations.

Further, I would encourage the authors to provide more discussion of the uncertainty calculation of their method (how certain is the calculation of the convex hull), and how it relates to uncertainty in end-member mixing applications (how clearly defined is a single end-member, how time-variant is its concentration). On l197 you allude to some work you did on this, but this did not make it to the manuscript?

Overall, the manuscript would benefit from a thorough spelling and grammar check.

Minor comments:

l11: streamwater

l30: concentrations; what you mean is the concentration of different solutes are correlated

l33: dividing by the standard deviation. It doesn't require it per se, but yields better results

l34: transforms ~~the~~ from ~~the~~ observation space to ~~the~~ PC space

l35: and each of which accounts. What which? You are referring to the columns of Y? Why not just speak of the principal components?

l45: refer to diagnostic tools of (Hooper, 2003), as they propose a more formal analysis to assess the rank of the data. Hooper also finds evidence of a fourth end-member by the way... see around l173

l46: After thus subjectively determining the number of... Not sure what the purpose of this entire sentence is by the way, seems like it could be skipped altogether.

l50: also: spatial and temporal variability in end-member concentrations

l55: In spite of EMMA's wide applications

l81: to ~~the~~ end-member mixing

l86: Only the top k PCs are retained? So while you criticize this subjectivity in lines 45-47, your method necessitates the same step? Please elaborate on this.

l87: Is there a reason why the analysis must be performed successively in 2D, and not in ND?

l89: "convex-hull": why here in quotes? This term has been used throughout the manuscript. Move its definition to its first use.

l91: algorithm: also define k and d. And: how are they different?

l123: This is quite a central drawback of the method. Can some directions / ideas already be given? Maybe a hybrid between chemma and field sampling for em signatures? Using the time-variance might also provide a way forward, searching for periods where certain end-members dominate and are thus better characterized from stream water concentrations.

l144: Elaborate on 'uncertainty' versus spread of local minima of convex hull. are these really the same thing?

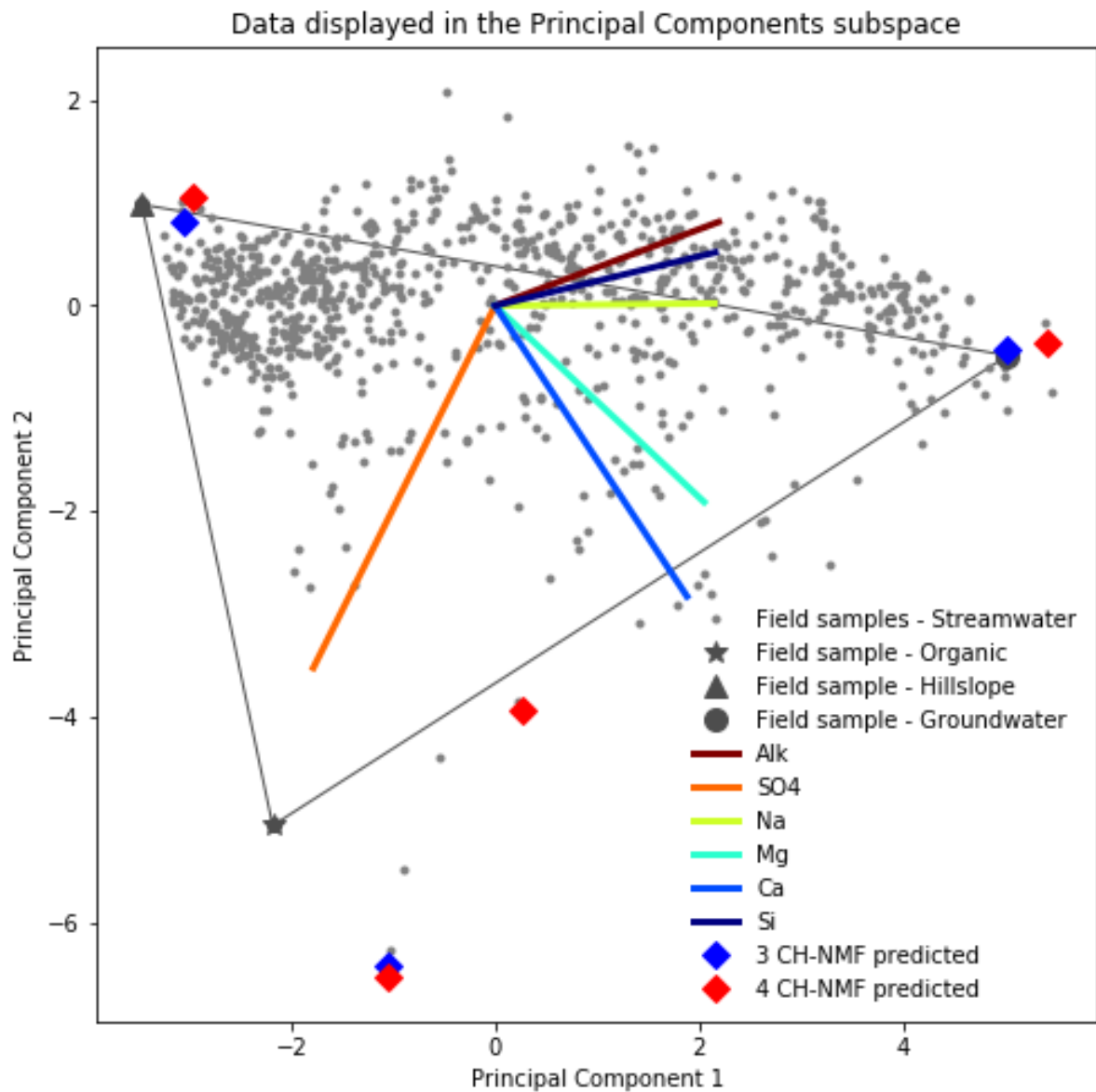
l149: What is a 'reasonable' variance? Is there a suitable metric? Provide guidance.

l197: "Fortunately, CHEMMA itself provides a tool for exploring some of these sources of uncertainty. By partitioning the dataset into time periods (or hydrologic state, etc), the temporal variability of end-members could be explored" What is this statement based on?

On the python code:

Consider making your code available as an importable python module through pip and/or conda.

Why is Figure 2 different on github code? location 4th endmember? accessed 1/7/2020:



References

Hooper, R. P. (2003). Diagnostic tools for mixing models of stream water chemistry. *Water Resources Research*, 39(3), 1–13. <https://doi.org/10.1029/2002WR001528>