Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-250-RC2, 2020

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Technical Note: A data-driven method for estimating the composition of end-members from streamwater chemistry observations" *by* Esther Xu Fei and Ciaran Joseph Harman

Anonymous Referee #2

Received and published: 13 July 2020

When I read the manuscript on the first time, I was thrilled, as this is something I have been waiting for many years to come out. If we could determine the chemical composition of end-members using streamflow chemistry alone, end-member mixing analysis would be significantly improved and revived. After I read it for a couple of times, I found the fundamental idea is still intriguing, but the assumptions the main method, namely CHEMMA, is based on may be flawed and cause significant uncertainties on the modeling results. A conceptual set-up of why and how this modeling would work could be strengthened. Readability could be improved as well, particularly in regard to some

mathematical details and their connection/implication with/in the hydrologic questions being investigated. Remember that most of readers who are interested in this study are hydrologists not mathematicians.

Major Comments:

The main approach is to use Convex-Hull Non-negative Matrix Factorization (CH-NMF) to infer possible end-member compositions by searching for a simplex that optimally encloses the stream water observations. The assumption for this is, based on authors, that end-members are located near the most extreme points that bound the observations in "mixing" space. From this assumption, it is clear that a simplex is basically determined by the data structure of observations, in other words, the shape of the sample cloud. What if one or more extreme points are missing in our observations? This could happen if samples are collected sparsely or only on certain hydrologic conditions/seasons that do not contain extreme samples (samples with extreme concentrations for at least one solute). The number of samples could also change how samples are distributed. With the same data set, can similar results (with reasonable uncertainties) be obtained from subsets of samples with varying number of samples that are randomly selected?

There is a lack of conceptual set-up where this study came from and where it goes in relation to existing tools in EMMA, particularly the diagnostic tools of mixing models (DTMM; Hooper, WRR, 2003). In one study, Christopherson and Hooper (WRR, 1992) specifically concluded that "Unambiguous identification of the source solution compositions from the mixture alone is impossible; thus, it is necessary that potential source solutions be derived from independent measurements." I do not mean this conclusion cannot be challenged, but the rationale must be stated clearly and explicitly, possibly using a conceptual set-up. Also, what is its relation with DTMM? Will the current study be supplemental or a substitute to DTMM in regard to the number of end-members? Can DTMM actually help to enhance CHEMMA and how?

The study used data collected in late 1980s. That is okay but what I am concerned is about the conservativity of all six solutes. How can we be convinced if all six solutes are conservative? If any of those is not conservative, the results of CHEMMA would be different. In my opinion, this is where DTMM may be able to help. Also, isn't it interesting to compare the number of end-members acquired using CHEMMA to DTMM?

Minor Comments:

L18: Before the first reference, add "e.g.,". Many classical references on EMMA were not actually cited.

L24: This statement should refer to conservative solutes.

L28: The second one is no longer a hypothesis or assumption because of the diagnostic tools of mixing models by Hooper (2003); See Liu et al. (WRR, 2008) for a demonstration and how this was addressed.

L30: "Streamwater concentration are naturally correlated." It is true if you refer to conservative solutes; otherwise it is an ill statement. Use two words "stream water" instead of one word "streamwater". Also, use plural for "concentration".

L31: Need at least one reference (e.g., Christopherson and Hooper, 1992).

L33-35: Multiple issues here. (1) Is Pobs actually eigenvectors? If so, use a parenthesis to annotate so; otherwise explain what it is and how to calculate it. (2) Get rid of the redundant "the". (3) My understanding is that once a standardized data set is used, a correlation matrix is decomposed rather than covariance matrix. Check if this is correct.

L36: If P are indeed eigenvectors, cite Christopherson and Hooper (1992) for the equation.

L41-42: Cite Hooper (WRR, 2003).

L45: True traditionally but not after DTMM is developed. See Liu et al. (WRR, 2008, 2017) as examples.

L51-52: Not true with DTMM.

L52-53: True but DTMM can help identify conservative solutes so that users can use only conservative ones. I mention this because I think your study is also based on mixing of conservative solutes. This should be stated/defined earlier in your text.

L60: Need to specify "extreme points". I think you refer to "extreme points of stream water samples".

L64: I think you mean "end-members' composition".

Result 2: Eigenvectors and PCs are different. PCs are calculated based on eigenvectors and observed concentrations.

Result 3: Is it specified anywhere how to project mathematically?

Result 4: Will the dimension of S differs from one projection plane to another?

Result 5: Is X expression actually [[xem1], [xem2], ..., [xemk]], as each xemi has a dimension of n by 1?

L93: I still think it is correlation matrix not covariance matrix. Also, what you mean here is eigenvectors not PCs.

L94: Spell out PCA as it appears for the first time.

L102: Specify the constraints, each between 0 and 1 with sum of all to be 1.

L125: I think "equifinality" is part of your talking here. Why not citing "equifinality" directly? It is a common term that hydrologists are very familiar with.

L186-206: Need to indicate where this modeling will lead to and how it may work together with DTMM.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-250, 2020.