Response to Reviewer

A *data-driven method for estimating the composition of end-members from stream water chemistry timeseries* by Esther Xu Fei and Ciaran J. Harman

Jan. 5th, 2021

We would like to thank the reviewer for their thoughtful comments and efforts towards improving the quality of our manuscript. In the following responses, we have updated the manuscript in four main categories:

1. We addressed the concern regarding the end-member distance by adding two paragraphs, one figure, and one table.
2. We clarified the "percent end-member limited" confusion by adding one paragraph and sentences that referring to this concept.
3. We fixed minor problems accordingly.
4. In addition to reviewer's comments, we also made minor revisions to the methodology section to further clarify how CHEMMA works.

Below our responses are in green, and the feedback from reviewers and editors is in black.

I appreciate authors' patience and continuous effort to improve their work and the manuscript. Compared to its last version, this version has been significantly improved. Publication of this work may have a high impact. Thus, the quality of its presentation must be ensured.

My current concerns focus on the following:
(1) Authors misunderstood the end-member distance, which cannot be inferred from Figure 3. It is not the geometrical distance between two end-members in the mixing diagram, but the distance between U- and S-spaces (defined by Christophersen and Hooper, 1992) for individual solutes of each end-member. It is a measure of fitness an end-member to the mixing space. Examining the end-member distances may provide a strong support of the end-member characterization by CHEMMA. Field sampling from a specific location may not perfectly represent the characteristics of an end-member due to spatial (also temporal at some point) variation, but CHEMMA may do better in this case. Any improvement can be reflected by decreases in end-member distances between U- and S-spaces (the shorter the end-member distances, the better it fits to the mixing space; perfect if zeros for all solutes for an end-member).
This calculation is needed because in this study there is a lack of quantitative evaluation of end-member compositions by CHEMMA, other than qualitative comparison with field measurements.

We would like to thank the reviewer for suggesting calculating the distance between U-space and S-space for each end-member. We added a new section in the methodology section to include the paragraph after L185 as "Assessing the goodness of fit". Then we modified the

previous paragraph from L185-188 and added another paragraph to reflect the quantitative assessment of goodness of fit in the Methodology section:

> *"There are several metrics that arise naturally from the CHEMMA framework which could be used to assess the goodness of fit of the inferred mixing subspace. The first and second are the centroid and within-cluster variance of each inferred end-member, which will tend to increase as the number of end-members increases. The third is the orthogonal projection distance from observation space to the mixing subspace, which will be smaller when the end-member lies closer to the linear subspace where the rest of the data live. In this paper, we consider a new cluster to be tenable as a proper end-member if: 1) the spread of previously identified clusters remains similar or decreases, 2) the cluster itself has a reasonable variance, and 3) the orthogonal projection distances of previously identified end-members do not significantly increase after adding a new end-member.*
>
> *We can also assess the degree to which CHEMMA and field-sampled end-members are "similar" to the stream chemical signatures. Field end-member candidate samples typically rely on a few grab samples (for example in \citet{Hooper1990} the groundwater were based on samples from a single well) which may insufficiently sample the overall source variability. CHEMMA end-members may provide a better idea of the time-space averaged chemical signature of a source than the field samples. One way to examine this is to look at the difference between an end-member's composition and its composition when projected into the reduced-rank $k-1$ principal component subspace. This can be done for both field-sampled and CHEMMA end members. A summary measure of that difference is the Euclidean distance of the end-member from the reduced-rank subspace. Where that distance is shorter the end-member has a chemical profile that is aligned with that which is typically found in the stream. This distance can be calculated from the loadings on remaining $n-k+1$ principal components."*

Accordingly, in the result section, we have added one paragraph after L213 to reflect the result from calculating this distance with one figure and one table:

> *"The three CHEMMA end-members are also located closer to the subspace spanned by the $k-1$ PC than the original three field-sampled end-members. The orthogonal projection distances are given in Table \ref{tab: dis}, and show that the CHEMMA end-members are more similar to the stream chemistry than the field samples, particularly for the groundwater end-member (field sample distance: 0.814, CHEMMA sample distance: 0.450). The differences in the chemical signatures of the groundwater end-members and their projections in the data subspace are shown in Figure \ref{fig: distance} (with concentrations given in standardized units, left for field samples and right for CHEMMA predictions). The CHEMMA end member's Alkalinity, SO$_4$, Ca values in*

particular are much closer to that of the data subspace than the field-sampled end-member, which is indicated by the shorter distance from the original 6-D chemical profile in dots (blue for field samples and red for CHEMMA predictions) to the 2-D mixing space profile in flat caps (orange for field samples and green for CHEMMA predictions). Only for Si is the field-sampled value closer. After PCA dimension reduction, both field-sampled profile and CHEMMA-predicted profile are close in the standardized solute space. It is worth noting that CHEMMA does not require dimensional reduction: PCA is only needed to determine the number of end-members."

We also added the following section at the beginning of "3.2 Dimensionality and DTMM" section to present the rest of the results of Table 3:

"For 4 CHEMMA end-member case in Table \ref{tab: dis}, the orthogonal projection distances of organic, hillslope, and groundwater end-members decrease/remain similar with 3 CHEMMA end-member case. Adding a fifth end-member significantly increase the projection distance of identified 4-th end-member."

(2) Thanks for explaining "percent end-member limited"! But I think it is still hard to follow, particularly because it is not explicitly described in the method section. If I understand it right, samples were generated randomly first and then screened by end-member criteria, with those outside the triangle discarded. If so, it should be called "percent of samples limited by end-member criteria", which makes more sense.

More importantly, I do not understand the point you were making using those outside samples. Did you mean the more samples lie outside the triangle the better end- members were characterized, as shown by Figures 7 and 8? If this is indeed what you meant, it does not seem to be correct (well, actually, right results for wrong reasons) and even misleading. It is misleading because readers could consider samples outside the triangle here in your case the same as outliers we have often seen in our real samples. Essentially, cases 4, 5, and 6 did better because they contained more "extreme" samples that are closer to true end-members. I use "extreme" here to indicate samples having extremely high fraction of one end-member but extremely low fraction(s) of the others. This is analogic to using baseflow for groundwater in some real-world cases. If this is what you really meant, I think you have to change your description and make sure readers will not be misled.

Actually, a better, more intuitive way to do this is to generate random samples while applying the constrains at the same time, with varying constrains of end-member fractions, e.g., case 1: 0.4-0.6, case 2: 0.3-0.7, case 3: 0.2-0.8, case 4: 0.1-0.9, and case 5: 0-1.0. This will control the number of samples closer to the true end- members.

We would like to thank the reviewer for pointing out the confusion here. First, we have adjusted several words in sentences from the methodology section to increase the clarity. Second, the 'percent end-member limited' is a measure that is specific to the way we generated

the synthetic data, but serves its purpose in characterizing the spread of the data relative to the mixing space. We have modified the paragraph at line 313 to remove the emphasis on this metric and focus on the insight gained:

> *For the synthetic dataset, the algorithmic uncertainty becomes insignificant when the data cloud just begins to be constrained by the end members. In case 4 in Figure 7) less than 1% of the random samples generated fell outside the mixing space (and were thus discarded). Note that it is the edges, not the verticies, that have affected the shape of the data cloud at this stage. This suggests that the CHEMMA algorithm does not require that there be `extreme' samples containing large contributions from only one end member (i.e. samples close to a vertex in the mixing space). Rather, it can detect mixing structure robustly when the dataset includes samples containing very small contributions of one end member, and intermediate contributions of other (i.e. samples close to an edge/face of the mixing space, but far from a vertex). However, an end-member whose contribution is consistently low may not be effectively detected because it does not affect the shape of the data cloud boundary sufficiently to justify increasing the number of end-members sought (i.e. the number of principal components retained in the analysis plus one).*

(3) Some statements are still pretty speculative (though they may not be incorrect). I suggest to stick with what your data (figures and tables) actually show and avoid stretching too much or being too inclusive. See examples below where this happens.

Thank you for this comment. We have responded to those statements in the section below.

Miscellaneous Comments:
P1/L15-16: This clause is awkward. I think you meant that “a subset of samples with extremely high and low fractions of end-member contributions ... under extreme hydrologic conditions”.
P11/L317-319: Both statements are awkward, but I know what you meant. “...small contribution ...” in the first sentence refers to samples having extremely high and low fractions of end-member contributions under extreme hydrologic conditions and the second refers to some potential end-members exist in the catchment but their impact to streamflow and chemistry is insignificant. If so, change them.
P12/L345: I understand it but it would be hard for others to follow or to get it right away. I think you were talking about extreme samples here as I mentioned earlier.

We would like to collectively clarify the meaning of these two sentences. As discussed in our response to (2) above, we mean that to successfully identify end-members CHEMMA requires a subset of samples that are lacking in one particular end-member but are enriched in others. We have modified the last sentence in the abstract to “*The results suggest that the mixing space can be identified robustly when the dataset includes samples that contain extremely small contributions of one end-member -- samples containing extremely large contributions from one end-member are not necessary, but do reduce uncertainty about the end-member composition*.”

For L317-319, see our response to major comment (2) above. For L345, we slightly changed the sentence to "*In other words, the algorithmic uncertainty was essentially eliminated if at least a few samples contained nearly zero contribution from at least one end member.*"

P4-6/L116-146: I think "k" has a different meaning before and after Line 127. I think it means the number of PCs or the number of dimensions before Line 127, but the number of end-members or vertices after Line 127. Because the number of end-member is one more than the number of mixing dimensions, "k" can not be used the same in the two sections.

Thank you for your suggestion. We have corrected the dimension from k to k-1.

P5/L132: Cite "Step 3, Figure 1a" here.

Thank you for your suggestion. We have changed it to "Step 3, from Figure 1a to Figure 1b, notice the changing distribution of the blue points" to reflect the process of projecting samples from the observational space to the principal component space.

P8/L234-235: Good to have such a statement, but it is too casual and lack of specifics.
Thank you for pointing out the need of clarification. We have added a following sentence to clarify it as "CHEMMA identifies sources that can be found through their control on the boundary of the sample space."

P9/L244-245: Why not cite the results of DTMM to support your statement here and also in the conclusion (P12/L359-360)?
P12/L359-360: DTMM results should be cited in the results section.

We have cited the DTMM for L244-245 and add additional sentence to clarify the conflict between DTMM and the rule of one: "An additional dimensionality (additional eigenvector to be retained) can be added until residual structure is unseen or is not improved."

For L359-360, we also added another sentence citing DTMM result to support our conclusion. "DTMM (Hooper 2003) was used to conclude that 1) the dimensionality of Panola dataset is at least 3 (i.e., at least 4 end-members are required), 2) the possible fourth source (end-member) may be weathering products containing calcium and magnesium. CHEMMA was able to identify a fourth end-member with such characteristic without run through DTMM analysis."

P9/L255-256: An example of a speculative statement without a demonstration or support (though not necessarily incorrect).

We have added a citation to Inamdar et al 2013, which corroborates our point.

P11/L330-339: Discussion mode.

We feel that this section works better in the Conclusion and would prefer to keep it here.

P11/L336-338: Awkward and hard to follow.

We have clarified this sentence by changing ‘biased’ to ‘uncertain’

P12/L363-364: Clueless to readers with a sudden introduction of these terms. Citing a perception is better.

We have added a citation where readers can get more information.

P12/L366 (beginning): Awkward and hard to follow.

Changed ‘observations’ to ‘samples’.

P2/L30-33: Cite literature where the statements came from. I think some languages/phrases are not accurate, e.g., “approximately” and “additional end-members” used in the statements.
P2/L46: Try not to use ambiguous phrases. Change to "a similar approach".
P3/L85-86: Not completely true. People used baseflow for groundwater in many cases. Change the statement to “there is not a method ..., other than using baseflow to characterize groundwater ...”. Add a reference (e.g., Liu et al., Ecohydrology, 2008).
P8/L214: You mean “four magenta diamonds”?
P8/L229: Not assumed. Using "suggested" may be a better word choice.
P10/L280-286: Parenthesize these sequential numbers to avoid unnecessary confusion with counting numbers.
P11/L342: Fix the grammar here.
P12/L351-354: Use consistent numbering. Used numbers before but letters here.
P12/L361: Change “might” to “should”.
P12/L362-369: Use consistent numbering, including parentheses (half or complete) throughout the manuscript.

We would like to thank the reviewer for pointing out these minor problems. We have corrected them accordingly.