hess-2020-250 Submitted on 24 May 2020

A data-driven method for estimating the composition of end-members from stream water chemistry observations

Esther Xu Fei and Ciaran Joseph Harman

Response to Reviewers

We would like to thank the editor and reviewers for the useful suggestions, which have improved the manuscript. In particular, we have included several major additions:

- *Greatly expanded analysis of the limitations, robustness, and uncertainty of the proposed approach, including:*
  - *Analysis of the uncertainty in end-member identification using bootstrapping*
  - *Analysis of the effect of dataset size by subsampling the data*
  - *Analysis of how data structure controls the robustness of the parameter identification using synthetic datasets*
- *Expanded discussion of the relationship between the proposed method and previous literature*
- *Discussion of the relationship between CHEMMA and DTMM*

These additions have necessitated resubmitting the paper as a regular research paper rather than a Technical Note. We look forward to further feedback from reviewers. Below we provide responses to the second round of reviewer comments on the technical note.

Reviewer #1

The authors proposed an interesting method to decompose stream water into end-members using stream tracer data alone. I like this idea very much.

Thank you for this encouragement, and for the helpful suggestions.

However, the shortages of the method are apparent. For example,

(1) it relies heavily on the used data. When the used stream samples are different, the method likely yields different identifications of end-members. The numbers of stream samples, the extreme points involved in the input data, the tracers measured from the stream samples, as well as the seasons during which the stream water were collected, have significant impacts on the outputs;

We agree that this weakness was not adequately addressed in our previous version of the paper. In the new version we have implemented two additional modeling experiments. First, to examine the influence of sampling variability we bootstrapped the original stream water samples 1000 times to estimate the resulting variability in identified end members. Second, in order to

understand the effect of the numbers of stream samples, we also sub-sampled the original dataset with smaller sample sizes. The results of these analyses can be found in section 3.3 Uncertainty analysis from line 253 to 280, and figures 5 and 6.

(2) The interactions of the tracer concentrations and contributions to stream water of end-members were not treated well. An end-member with extremely high or low tracer concentration may not necessarily result in extreme concentration of stream water when its contribution is rather low. I am wondering the ability of the current method to identify end-members with low contributions to stream water.

(3) Seems extreme points in the data series of stream tracer concentration are required for the implement of the method. If the collected stream water samples do not show any outliers, does the method still work? The mixture of end-members with distinguished tracer concentrations not certainly result in much changes in the stream water tracer concentration, considering their contributions are changing in the time periods.

Thank you for identifying these important points. We have included new analysis and discussion that addresses both. This includes further analysis of the Panola results, and analysis of a set of synthetic datasets.

We should note that the end-members identified using our method do not necessarily have extreme concentrations. Instead they have "unusual" concentration profiles when all variables characterizing them are viewed collectively. That is, they are at the 'fringes' of the data cloud in principle component space. For example, in the revised manuscript, this can be seen in the newly added Figure 5, where the predicted hillslope end-member does not have very extreme SO_4 *concentration, but rather was identified due to its uniqueness in other dimensions compared with the other two end-members. This point is discussed in the manuscript at line 259:*

> "Figure 5 also re-emphasizes that CHEMMA identifies end-members that exhibit collectively unusual combinations of concentrations (i.e., vertex-like structures in the overall data cloud). While many solute concentrations of CHEMMA predicted end-members are located towards extremal values of the observations, they need not be all individually extremes (e.g. the sulfate concentration of end-member 3, corresponding to the hillslope end-member, Figure 5 upper middle plot)."

However the end-member identification has less to do with the extremes than it does the overall boundary of the dataset in n-dimensional space. In other words, it matters more that there are some samples in which an end-member is absent than that there are some where it dominates. You are correct that if a source consistently contributes a minor fraction of the total, it would not be identified as an end-member, since its unique profile would not have an effect on the dataset boundary. However, this dependence on the overall boundary means that relatively few samples may be necessary to robustly identify an end member. We examined this question by generating synthetic datasets with varying degrees of imposed 'end-member' structure. These were generated by creating normally distributed random concentration profiles (essentially spheroids in n-dimensional concentration space), and then censoring samples that fell outside

the mixing space. By varying the variance of the distribution used to generate the samples we can generate datasets influenced to varying degrees by the end-members. The results show that end members can be robustly identified (albeit with some systematic bias, namely an underestimation of their 'extremeness') when only <1% of the synthetic data contained samples that entirely lacked at least one end-member. This point is discussed further in the manuscript in section 3.4 and in figure 8 and 9.

CHEMMA does not require outliers but the structure of the data points. To demonstrate this, we have added a synthetic data section to express when CHEMMA fails and when it starts to work perfectly (Figure 8).

(4) The method may not be able to identify end-members with similar tracer concentrations. This may be not important when we are focusing on the components contributing to the tracer concentration of stream water. But identifying runoff components with similar tracer composition could be very important to understand the changes of hydrological processes in catchments.

We acknowledge this is a problem – however, it is not a problem that is unique to CHEMMA. Nevertheless, as demonstrated in previous section two source waters with similar tracer concentrations may be distinguishable by CHEMMA if enough samples are available in which one is absent and the other dominant. This is discussed on Line 259 (given above) and Line 315:

> "The CHEMMA algorithm appears to detects structure more robustly when the dataset includes samples containing very small contributions at some time. However, a consistently very low contributed end-member will not be effectively detected because it does not affect the shape of the data cloud boundary."

Reviewer #2

After I examined the revision, my concerns about the validity of this pure mathematical tool grow instead of diminish. This mathematical tool is surely beautiful, but whether or not it yields hydrochemically meaningful results has not been well demonstrated. No matter how beautifully can chemical concentrations in end-members be inversely derived mathematically from streamflow chemistry alone, it has to be proved to be accurate with foreseeable and acceptable uncertainties. The uncertainty I am talking about here is not the uncertainty arising from chemical analysis as we always know but one caused by this tool per se. There are two questions that are related to this type of uncertainty, which were inquired earlier but not actually addressed in the revision. If the number of samples changes significantly, can chemical concentrations in end-members be still determined within an acceptable range? If non-conservative solutes are included in the analysis, are the results of chemical concentrations in end-members and the number of end-members determined by this tool consistent and valid? These questions have to be answered with actual data analyses before it becomes a convincing tool. As it currently stands, I do not feel it is ready for others to use this tool with high confidence.

Chemical concentrations in end-members were determined by data structure by CHEMMA. No doubt that a significant change in the number of samples will change chemical concentrations in end-members. But as long as the determined concentrations are within certain range, we are fine with it. Simply saying CHEMMA applies to large data set without a demonstration is not an appropriate answer. Also, how large can we consider it to be "large"?

We appreciate these suggestions on how to improve this manuscript, and have made extensive additions to the manuscript to address them. We have added sections to examine how the robustness of the identification depends on number of samples available, the uncertainties associated with the sampling variability, and the robustness of the method to different data cloud structures. We may not be able to develop a universal standard to quantify how many samples is 'enough', this really depends on the structure of each given study dataset. However, as we saw from the results of exploring the Panola dataset (Figure 6), CHEMMA converged on some components of some end members with as few as 45 samples, while for others as many as 500 samples were needed before the estimated component concentrations converged to the values estimated using the full dataset of 905 samples.

In my opinion, non-conservative solutes should not be included in the analysis, nor can chemical concentrations of non-conservative solutes in end-members be derived from CHEMMA. Otherwise, CHEMMA should be a totally different set-up and have to deal with chemical equilibrium. With non-conservative solutes included in CHEMMA, however, solutions can be achieved mathematically by increasing the number of end-members. Rick Hooper noticed this problem in EMMA in his later work, which eventually led the publication of diagnostic tools of mixing models in 2003. Including non-conservative solutes in EMMA will cause polygon to bend outward. Yes, this problem can be mathematically resolved by adding additional "end-members" to obtain a more complex polygon, but they are not truly end-members because in such cases the assumptions of mixing models are violated. Analogically, the same issue applies to CHEMMA. Why not testing using DTMM if all solutes included in CHEMMA are conservative and examining whether or not non-conservative solutes caused the fourth end-member? Also, why not determining the number of end-members using DTMM and comparing with CHEMMA? Simply citing the published results of Hooper et al. (1990) cannot guarantee those solutes are conservative, as conservative behavior of those solutes, along with the number of end-members, were determined subjectively at the time.

We acknowledge your thoughtful comment, and agree that non-conservative solutes should not be included. CHEMMA is based on the assumption of conservative mixing, as is EMMA. We agree that DTMM should be used conjunctions with CHEMMA to test for whether solutes are conservative, and to select the number of end members. CHEMMA is compatible with DTMM in this regard. We have introduced a new section discussing the use of DTMM for these purposes. In particular:

> "To carry the idea of DTMM rank determination further, we performed a five-fold cross validation analysis on PCA fit residuals on the Panola data with varying dimensionality (Figure 4). The mean square errors of residuals (Figure 4a) exhibited the greatest decrease when the dimension was increased from one to

two, which suggests three end-members might constitute a parsimonious set. However, the small normality test p-value in Figure 4 b) shows that residuals of sulfate, magnesium, and calcium solutes still maintain some structures in a two dimensional mixing space. Residual structures persist until the dimension goes beyond five(Figure 4 b)). Thus even with DTMM, the 'true' rank of the dataset remains uncertain. However, DTMM analysis at least provides an established method to identify conserved solutes and to justify the choice of rank. The robustness of CHEMMA end-members could also serve as a check for DTMM-determined rank of mixture."

Reviewer #3

The authors present a promising method that allows to identify and characterize end-members using stream water tracer concentrations only. While I believe their work is a valuable contribution to existing literature, one major shortcoming is the lacking literature review and thus relating their work to the existing literature (beyond EMMA). I am aware that the field of end-member mixing modeling is wide, however, the authors miss to acknowledge key papers of the field of hydrology such as Carrera et al., 2003, (https://doi.org/10.1029/2003WR002263) or Genereux, 1998 (https://doi.org/10.1029/98WR00010). Likewise, they neglect recently published papers that provide methodological advances in mixing analyses within hydrology, e.g., Beria et al., 2020 (https://doi.org/10.5194/gmd-13-2433-2020), Popp et al., 2019, (https://doi.org/10.1029/2019WR025677), or Barbeta and Peñuelas, 2017, (https://doi.org/10.1038/s41598-017-09643-x). The authors claim that no method exists to characterize missing or unmeasured end-members, however, Popp et al., 2019, have provided an approach that allows to identify unmeasured end-members.

Thank you for identifying these key papers. The cited papers focus on similar topics but take different approaches (using Bayesian framework to reduce and quantify uncertainties) than we are presenting. We have added a paragraph in the introduction to acknowledge their work.

"It is worth distinguishing CHEMMA from previous applications of statistical learning methods (such as maximum likelihood estimation, Bayesian inference, and Markov Chain Monte Carlo, MCMC) to end-member mixing analysis. Genereux (1998) presented a linear estimator for uncertainties in end-member concentration and mixing ratios. Carrera et al. (2004) achieved something similar using a maximum likelihood method. By combining likelihood methods, Bayesian inferences, and probabilistic linear models with a MCMC algorithm, various authos including Barbeta and Peñuelas (2017); Beria et al. (2020); Delsman et al. (2013); Popp et al. (2019) have been able to obtain time-evolving uncertainty estimates. These contributions all focus on quantifying uncertainty resulting from the use of field-sampled candidate end-members. In contrast, CHEMMA aims to infer the end-members themselves."

Popp et al., 2019 does provide an approach that allows identification of unmeasured end-members, but we would also like to clarify the difference between our approach and theirs. Popp et al., 2019 introduce a single residual end-member that represents the collective effect of all unobserved end-members in addition to the observed ones, which are used to initiate the Bayesian mixing model. That is, their approach is best suited to where information on some observed end-members is available. However, CHEMMA do not require any observed end-members. It relies solely on the mixture data. The contrast with the Popp model is discussed on line 79:

> "Popp et al. (2019) comes close, introducing a residual end-member that represents collective behavior of all unobserved end-members. Their method still requires some a-priori knowledge of "observed" end-members to initialize a Bayesian mixing model. In contrast, CHEMMA allows for inference of the entire suite of end-member compositions, and their associated uncertainties."

Another shortcoming is that the method is only applied to one data set. I strongly suggest to apply the method to other data sets (using different tracers and in a different geographic setting) to prove the robustness of the method. It's been shown that the tracer set size has a major influence on end-member mixing modeling (Barthold et al.). Testing the method on other datasets should be feasible since many datasets are readily available nowadays.

It is certainly true that the validity of the approach will be strengthened by applying it to multiple datasets. We have chosen not to here, and hope that the material we have included in the new manuscript is sufficient to motivate further investigation. The reviewer's point is well taken though, and we have partially addressed it with the addition of the synthetic datasets in section 3.4, which we use to analyze to examine the robustness of CHEMMA. We have also added a section where we subsample and bootstrap the Panola dataset in order to explore sources of uncertainty and the robustness of the results. We look forward to applying CHEMMA on other datasets in the near future.

I really appreciate the detailed description of the methods and the valuable reflections (section 4) added to the latest version (v3). It is also great that the code and data are available in a Jupyter notebook.

Thank you for recognizing our work in this area. We will continue refining the code and may make it to be a Python package in the future.

The manuscript is well written and clearly structured, however, readability should be improved by correcting for a couple of grammatical flaws (e.g., articles are often missing) that I believe to have detected.

- L. 4 and throughout the text: consider replacing "candidate" with "potential"
- L. 24: replace "should" with "can" be explained
- L. 28: the hypotheses

• L. 29: I would rephrase it saying that the "1) stream water consists of the identified end-members and 2) all end-members were identified correctly".

• L. 63: please add a reference to the 4th statement

• L. 67-68: this statement is not true. See comment above about method provided by Popp et al., 2019, (https://doi.org/10.1029/2019WR025677)

• L. 74-76: I would really appreciate to see this method applied to other datasets

• L 92: "end-member mixing" approach/method?

• L. 96: subspace

• L. 152: analysis many times (instead of "a large number of")

• L. 163: can you please specify "reasonable"?

• L. 181 following: please statistically quantify the similarity between field measurements and your values. E.g., not only alkalinity (hillslope) seems to differ considerably. Also, consider removing decimal

points in Table 1 given the high st.dev.

• L. 204: please use a statistical test to quantitatively describe how well you can reproduce end-members of Hooper et al.

• L. 225: that is a great suggestion! You could also indicate that a time lag representing a delay caused by tracer transport from the source to the output (see Beria et al.) adds uncertainty.

• L. 232: I would remove this statement.

• increase font size in Fig. 3

• Figs. 2 and 3 and table 2: specify what is meant by "organic"

Thank you for helping in improving the readability of our manuscript. We have adapted your comments in our manuscript.