Response to Reviewer 1.

General comments:
The paper applies NMF (Non-negative Matrix Factorization), which is a machine learning technique, to EMMA (End-Member Mixing Analysis). They use this to calculate CO2 sequestration in three watersheds. The novelty is the application of NMF to EMMA. In general, the paper is well written. I suggest publication if the comments below are addressed.

We thank the reviewer for providing helpful comments towards improving the paper. We incorporated the comments suggested here, which we believe strengthened the manuscript. Below are the specific responses to each comment.

Specific comments

1. Line 19-20, 44-45 and 412. You talk about a "new machine learning technique". Actually, it is not a new technique. What you do is applying an old technique (machine learning or, more specifically, NMF) to EMMA, which is new.

We agree with the reviewer that the technique is not new, but rather the application of the technique. We have updated the manuscript to reflect these changes.

Specific examples of our edits:
- Lines 172-173 we clarify that NMF has been used in other applications. "NMF is an algorithm that has been used for many applications (e.g., spectral analysis, email surveillance, cluster analysis; Berry et al., 2007) but has only recently been applied to stream chemistry (e.g., Xu and Harman, 2020)."

2. Line 132-134. You say that NMF is unique in that it does not rely on assumptions of endmembers a priori. This is repeated throughout the whole paper (figure 2, line 172, 412 and 428). I think this is not entirely true. For instance, Carrera et al. (2004) calculate endmembers without NMF. Carrera, J., E. Vázquez-Suñé, O. Castillo, and X. Sánchez-Vila (2004), A methodology to compute mixing ratios with uncertain endmembers, Water Resour. Res., 40, W12101, doi: 10.1029/2003WR002263.

We thank the reviewer for the reference. In the revised manuscript we emphasize that this technique is different than traditional inverse methods and acknowledge previous work that has improved EMMA through modeling under-constrained endmembers (including Carrera et al., 2004).

Specific example of our edits:
- Lines 79-82 we discuss prior contributions. "Since the inception of EMMA, many researchers have aimed to improve analysis through a more accurate determination of unknown or under-constrained endmember chemistries (Hooper, 2003; Carrera et al., 2004; Valder et al., 2012). But these efforts all use some a priori determination of

3. Line 138: You use SO4 as a reference for solute concentrations. To me it would make more sense to use Cl-, instead, because it is not likely involved in chemical reactions. Is there a particular reason for using SO4?

We thank the reviewer for their question. We specifically normalize to sulfate because it is the target analyte that we wish to separate in the stream. We have updated the text to highlight the rationale behind normalizing to sulfate.

Specific example of our edits:
- Lines 145-151 we discuss the normalization. "Here, cell entries of V are molar solute concentration ratios, $[X]/[Y]$, for stream samples. Indicator n refers to the sampling date, m refers to different solutes X ($= Ca^{2+}$, Mg^{2+}, Na^+, K^+, Cl^-), and brackets refer to concentrations. W is the n x p matrix whose cell entries are proportions, α, for each endmember in each stream sample. Again, n refers to sampling dates, but p is the number of sources of solutes (referred to as endmembers). The proportions refer to the fractions of sulfate in each sample that derive from an individual endmember, where the sum of proportions must equal 1 ± 0.05 for each sample. To derive the mixing proportions of sulfate specifically, we set up the NMF approach by normalizing each analyte concentration by sulfate concentration ($Y = SO_4^{2-}$), the target solute."

4. Line 145: You define end members for shallow, moderately shallow, and deep flowpaths. Of course, they may vary in time as you say in line 149. Could this create some bias? For example, end members of deep flowpaths are generally older with water that fell as rainfall earlier than end members of shallow flowpaths. As acid rain varies with time, differences in chemical signature can be affected by the age of the water.

We thank the reviewer for their thoughtful question. Although different flowpaths do have different transit times, we believe that our model is accurately separating acid rain and pyrite-derived sulfate.

Specific examples of evidence in our text:
- Figure 3 and lines 198-217 describe the separation of reaction fronts in the subsurface at Shale Hills. Lines 230-235 relate the NMF derived endmember chemistries to flowpaths based on the subsurface structure.
- Figure 4A and Lines 244-247 describe how our NMF model results compare to sulfur isotope values, which are not included in the model. The α_{deep} values and low $\delta^{34}S$ values shown in Fig. 4A are consistent with our interpretation that we are separating pyrite-derived sulfate and inconsistent with older acid rain.

5. Line 265: Equation 3 and kstream are not clear to me. Where does the -1 come from? I suggest adding an explanation in the SM like you have done for krock.

We thank the reviewer for this suggestion. We have updated the supplemental to include a derivation of κ_{stream} like we did κ_{rock}.

Specific example of our edits:
- SM Section 2.2, eqs. S11-S14:

"Next, we will derive κ_{stream}, the modern CO_2 sequestration coefficient. In general, both κ_{stream} and κ_{rock} (see SM 2.3) are used as ways to note the extent that weathering in a watershed is sequestering or releasing CO_2. κ_{stream} is the amount of CO_2 emitted or sequestered calculated from $[\Sigma^+]_{total}$ as described above, normalized by $[\Sigma^+]_{total}$(meq/l):

$$\kappa_{stream} = -\frac{\Delta CO_2}{[\Sigma^+]_{total}}, \qquad (S11)$$

The negative sign is used so that a negative κ_{stream} represents sequestration (uptake of CO_2), and a positive κ_{stream} represents release. From eq. S11 it is apparent that the CO_2 emitted or sequestered equals the product, κ_{stream} $[\Sigma^+]_{total}$, with the appropriate sign. Total dissolved base cations in a stream draining a watershed with no carbonate nor pyrite are attributed here entirely as CO_2-weathering: this watershed demonstrates the highest capacity to sequester CO_2 and κ_{stream} equals -0.5. Substituting from eq. S10 into eq. S11 yields:

$$\kappa_{stream} = -\frac{0.5\,[\Sigma^+]_{silicate-CO_2} - 0.25\,[\Sigma^+]_{carbonate-H_2SO_4}}{[\Sigma^+]_{total}}, \qquad (S12)$$

We can further expand eq. S12 by substituting eq. S6 for $[\Sigma^+]_{carbonate-H2SO4}$, eq. S9 for $[\Sigma^+]_{silicate-CO2}$, eq. S4 for $[\Sigma^+]_{silicate}$ and eq. S8 for $[\Sigma^+]_{silicate-H2SO4}$

$$\kappa_{stream} = -\frac{0.5\big([\Sigma^+]_{total} - [\Sigma^+]_{carbonate} - 2\alpha_{shallow}[SO_4^{2-}]_{total}\big) - \alpha_{deep}[SO_4^{2-}]_{total}}{[\Sigma^+]_{total}}, \qquad (S13)$$

This can be rearranged and simplified as:

$$\kappa_{stream} = -\frac{1}{2} + \frac{1}{2}\frac{[\Sigma^+]_{carbonate}}{[\Sigma^+]_{total}} + \frac{[SO_4^{2-}]_{total}}{[\Sigma^+]_{total}}, \qquad (S14)$$

We then define the second term (ratio of carbonate-derived base cations to total base cations in the stream sample) as γ_{stream} and the third term (ratio of the sulfate equivalents (from sulfuric acid) to the equivalents of base cations in the stream) as ζ_{stream}. Note that to obtain the sulfate equivalents, we multiply $[SO_4^{2-}]_{total}$ by 2, resulting in the third term equal to $0.5\zeta_{stream}$. Given these definitions, eq. S14 yields eq. 2 from the main text:

$$\kappa_{stream} = \frac{1}{2}(-1 + \gamma_{stream} + \zeta_{stream})"$$

6. SM, section 2.2. I find this section very hard to follow. Actually, you describe mathematical equations by using text. I think you can make it more readable, if you put the equations as well.

We agree with the reviewer that adding equations to describe the calculation increases the clarity. We have updated SM section 2.2 to reflect these changes:

Specific example of our edits:
- SM Section 2.2, eqs. S3-S10:

"Here we calculate the inferred CO_2 release or sequestration resulting from weathering as recorded in the sum of all base cation concentrations (meq/l) in each stream sample, $[\Sigma^+]_{total}$:

$$[\Sigma^+]_{total} = 2[Ca^{2+}]_{total} + 2[Mg^{2+}]_{total} + [Na^+]_{total} + [K^+]_{total}, \qquad (S3)$$

Here, we use the modeled base cation concentrations from NMF in eq. S3, and we use the uncertainty in the modeled concentrations for the error in $[\Sigma^+]_{total}$. To calculate the inferred CO_2 release or sequestration resulting from weathering, we use the results of NMF, as described below, to identify the extents of 4 weathering reactions recorded in each stream sample: 1) CO_2-driven weathering (CO_2-weathering) of silicates, 2) H_2SO_4-driven weathering (H_2SO_4-weathering) of silicates, 3) CO_2-weathering of carbonates, and 4) H_2SO_4-weathering of carbonates. We note these four quantities respectively as 1) $[\Sigma^+]_{carbonate-CO_2}$; 2) $[\Sigma^+]_{silicate-H_2SO_4}$; 3) $[\Sigma^+]_{silicate-CO_2}$; 4) $[\Sigma^+]_{carbonate-H_2SO_4}$. These are the four unknowns we seek to calculate for SH and ER, as described below.

Based on the high proton and low metal concentrations of the measured rain chemistry, the rain contributes negligibly to the base cation concentrations of the study streams; therefore, we apportioned all the base cations to weathering reactions. First, we note that the meq/l of cations derived from carbonate minerals, $[\Sigma^+]_{carbonate}$, equal $[\Sigma^+]_{carbonate-CO_2} + [\Sigma^+]_{carbonate-H_2SO_4}$. Likewise, the meq/l of cations derived from silicate minerals, $[\Sigma^+]_{silicates}$ equal $[\Sigma^+]_{silicate-H_2SO_4} + [\Sigma^+]_{silicate-CO_2}$. The summation of silicate-cations ($[\Sigma^+]_{silicate}$) is the difference between the summation of total cations ($[\Sigma^+]_{total}$) and that of carbonate-derived cations ($[\Sigma^+]_{carbonate}$):

$$[\Sigma^+]_{silicate} = [\Sigma^+]_{total} - [\Sigma^+]_{carbonate}, \qquad (S4)$$

We use a few field observations to complete the calculations for SH and ER, as explained in the main text. First, carbonate minerals only dissolve in water flowing along the deep path because carbonates have been depleted from shallow depths. Second, although some chlorite dissolves into water flowing along the deep path, the release of Mg at depth is insignificant compared to Mg released from carbonate. So we ignore shallow dissolution of carbonates and deep dissolution of silicates in both SH and ER.

With these observations, we can write:

$$[\Sigma^+]_{carbonate} = 2\alpha_{deep}[SO_4^{2-}]_{total}\left(\left(\frac{[Ca^{2+}]}{[SO_4^{2-}]}\right)_{deep} + \left(\frac{[Mg^{2+}]}{[SO_4^{2-}]}\right)_{deep}\right), \tag{S5}$$

Here, α_{deep} is the proportion determined through NMF of sulfate in a given water sample that was derived from reactions along the deep flowpath, $[SO_4^{2-}]_{total}$ is the total concentration of sulfate in the stream water sample under consideration, $([Ca^{2+}]/[SO_4^{2-}])_{deep}$ and $([Mg^{2+}]/[SO_4^{2-}])_{deep}$ are the model-derived ratios of $[Ca^{2+}]$ and $[Mg^{2+}]$ to $[SO_4^{2-}]$, respectively, that characterize the deep flowpath endmember for that sample.

Remembering that Mg release from chlorite dissolution at depth is insignificant compared to Mg from carbonates, all of the generated sulfate in the deep weathering endmember is balanced by cations from dissolved carbonate minerals:

$$[\Sigma^+]_{carbonate-H_2SO_4} = 4\alpha_{deep}[SO_4^{2-}]_{total}, \tag{S6}$$

(We multiply the concentration of deep sulfate by 4 because 4 eq of cations are released per mol of sulfate, noting that $[\Sigma^+]$ is in eq/L and $[SO_4^{2-}]$ is in mol/L). Any carbonate-derived base cations that are in excess of what could have been produced by pyrite-derived sulfuric acid are attributed to CO_2-weathering of carbonates:

$$[\Sigma^+]_{carbonate-CO_2} = [\Sigma^+]_{carbonate} - [\Sigma^+]_{carbonate-H_2SO_4}, \tag{S7}$$

Remembering that no carbonates dissolve into water flowing along the shallow path, then similar arguments for the shallow flowpath yield:

$$[\Sigma^+]_{silicate-H_2SO_4} = 2\alpha_{shallow}[SO_4^{2-}]_{total}, \tag{S8}$$
$$[\Sigma^+]_{silicate-CO_2} = [\Sigma^+]_{silicate} - [\Sigma^+]_{silicate-H_2SO_4}, \tag{S9}$$

From these equations, values for the four unknowns can be calculated for SH and ER. A similar approach was taken for HB except that no carbonate minerals were present, and only two unknowns were determined ($[\Sigma^+]_{silicate-H_2SO_4}$, $[\Sigma^+]_{silicate-CO_2}$).

With respect to the atmosphere considered over the long-term (10^5-10^6 yr), H_2SO_4-weathering of silicates and CO_2-weathering of carbonates are CO_2 neutral, while CO_2-weathering of silicates sequesters CO_2 and H_2SO_4-weathering of carbonates releases CO_2 (Fig. 1). As seen in Figure 1, per mole of $CaSiO_3$ or $CaCO_3$ weathered, CO_2-weathering of silicates sequesters 1 mol of CO_2 and H_2SO_4-weathering of carbonates releases 0.5 moles of CO_2. In terms of $[\Sigma^+]_{total}$, CO_2-weathering of silicates sequesters 0.5 moles of CO_2 per base cation equivalent released into solution and H_2SO_4-weathering of carbonates releases 0.25 moles of CO_2 per base cation equivalent released into solution (Fig. 1; Reactions 2, 3, 6, and 7). For a given water sample, the cation concentrations record the extent of dissolution of carbonate and silicates, as long as the contribution of these base cations from acid rain is minimal. (For simplicity, we do not correct $[\Sigma^+]$ for rain chemistry but see SM Section 4). Therefore, the uptake or release of CO_2, ΔCO_2, can be calculated for any given stream water sample:

$$\Delta CO_2 = 0.5 \, [\Sigma^+]_{silicate-CO_2} - 0.25 \, [\Sigma^+]_{carbonate-H_2SO_4}. \tag{S10}"$$

7. SM, line 55-57. If I understand correctly, here you attribute all Ca and Mg to carbonate dissolution. However, it can also come from silicates. In fact, in figure 1 you represent silicates by CaSiO2. Do you simply neglect Ca from silicate weathering?

We do not attribute all Ca and Mg to carbonate dissolution, but rather all Ca and Mg in the deep weathering flowpath to carbonate dissolution. Ca and Mg in the shallow-weathering flowpath are thus attributed to silicate dissolution.

Specific example of our edits:
- Eq. S5 shows that carbonate cations are only derived from Ca and Mg in the deep weathering flowpath: $[\Sigma^+]_{carbonate} = 2\alpha_{deep}[SO_4^{2-}]_{total}\left(\left(\frac{[Ca^{2+}]}{[SO_4^{2-}]}\right)_{deep} + \left(\frac{[Mg^{2+}]}{[SO_4^{2-}]}\right)_{deep}\right)$

Technical corrections

1. SM, line 28. Change "in in" to "in".
   - SM line 31 was revised to correct this mistake.
2. SM, line 133-134. I think this equation is equation 3 from the main text, not 2.
   - SM line 233 was revised to correct this mistake.
3. SM, line 148. You refer to Fig. 3C. However, this figure contains nothing related to lag times. I think you mean Fig. 4C.
   - SM line 249 was revised to correct this mistake.

Response to Reviewer 2.

This study focuses on applying machine learning to endmember mixing analysis of weathering chemistry in subsurface groundwater flow paths. They apply an NMF scheme and train it on syntactic data generated using a multivariate normal distribution of log-transformed stream water chemistries. The NMF is then applied to 3 measured stream water samples to delineate mixing proportions. The study is well presented and written, the SM is seminal to the understanding of the study and holds the key details for the optimization of the NMF. The main finding is within the sensitivity of the reaction to the groundwater flow paths which are unknown, yet they control the concentration relation between the components and therefore they are controlling the overall chemistry. In a way, these flow paths are a spatial localization of the reaction in space and time, due to the seasonal effects, as shown here. I found the paper very interesting, well written, and clear and supports the publication of the study, yet the main missing part that is not discussed here and must be added is a discussion on the how.

We thank the reviewer for their thoughtful summary, reviewing our work, and for their support in publishing our work. Below are reposes pertaining to your questions of "how".

How does the NMF manage to capture the effect of the subsurface groundwater flow paths? What is the additional mechanism that is deciphered by the NMF? The spatial and temporal

effect of the subsurface groundwater flow paths must be captured in a meanfield way by the MNF, and this is not clear how it managed to do so and what was the missing mechanism.

NMF determines patterns in datasets. As domain scientists, we then interpret those patterns in the context of our system. We describe in detail in section 3.2 how we interpret the endmembers and relate them to flowpath and sub-surface structure. In a practical sense, water that flows along a flowpath dissolved minerals that are present along that flowpath. Because there is a separation on reaction fronts at our sites, each flowpath interacts with different minerals and, therefore, has a unique chemical signature that can be detected and separated with NMF. We have updated the text to clarify this idea.

Specific example of our edits:
- Lines 259-262 describe the importance of reaction front separation. "The dissolution of different minerals along these flowpaths lead to patterns in stream chemistry that our NMF model discerns and separates. If mineral reaction fronts are not separated in the subsurface, different flowpaths might not be separated by NMF; however, Brantley et al. (2017) and Gu et al. (2020a) have argued that separation of reaction fronts is common."

I agree with referee 1 remarks 5 and 6, do clarify the mathematical components with a mathematical expression.

Specific example of our edits:
- See SM section 2.2 eqs. S3-S14