We are grateful to the associate editors and the two reviewers for their detailed and constructive feedback. After carefully considering the reviewers comments, we agree that the methodology in presented in our previous draft had substantial flaws that needed addressing. After refinement of our clustering approach as proposed by the reviewers, we find the robustness of our statistical analysis substantially improved. Based on the reviewers' comments, we reconsidered and refined some of our interpretations, which led us to thoroughly restructure the Results and Discussion section. In our opinion, the revised conclusions are better supported by our data and communicated in a clearer way. Please see a detailed point by point response below.

Reviewer 1

1. Statistical support for the two-group GMM classification should be made more transparent.

The two-group division of barite $\delta^{18}O$–$\delta^{34}S$ data is central to the manuscript. However, the Methods state that the Bayesian Information Criterion indicated three components as the optimal GMM solution, whereas the authors selected two components because the three-component model produced negative regression slopes and because two components are consistent with the expected AOM–OSR end-member framework. This choice may be reasonable, but it requires clearer justification. Please report the BIC values for one-, two-, three- and preferably four-component models. The rejected three-component solution should be shown in the Supplement, together with a short explanation of why negative slopes are considered geochemically implausible rather than reflecting mixed growth generations, local fluid mixing, non-coeval sulfate reservoirs, or sampling heterogeneity. I also recommend a sensitivity test showing whether the two groups remain robust when Forsmark and Laxemar are treated separately, or when individual crystals/samples rather than individual SIMS spots are used as the statistical unit. Because multiple SIMS spots may come from the same crystal or fracture, they are not fully independent observations; this should be considered when evaluating the robustness of the clustering and regression results.

- We agree that additional statistical arguments are required to strengthen the reasoning behind our methodology. These are now detailed in our now expanded method section and new Supplementary Figs. 1-2. Upon revisiting our GMM approach based on the Reviewers' comment, we noted it lacked an initial outlier test. This is crucial because GMM is highly sensitive to outliers. After excluding these outliers (new Fig. S1), we calculated the BIC values for one to four component models, showing a clear minimum for two components (new Fig. S2A). This strongly suggests that the previously supported three component model was an artefact created by outliers. As suggested, we additionally filtered the Forsmark data out to test for sensitivity to the locality. The BIC values and the resulting clusters (new Figs. S2C-D) are consistent with the result for the complete dataset. Together, these observations provide strong support for using a two-component model.
- An additional test by running GMM on only the Forsmark samples is, unfortunately, prevented by the low number of datapoints in that group (n =21). We disagree with using whole crystals or samples for the GMM, because crystals can exhibit multiple growth generations and samples may contain crystals of different ages. In our opinion, considering them collectively (e.g., average values) would create artificial signals.

2. Quantify the basis for the AOM- and OSR-dominated interpretations.

The wording that Group 1 was "dominantly associated" with AOM and Group 2 was "dominated" by OSR is reasonable. However, because Group 1 also includes barite associated with less strongly ^{13}C-depleted calcite, and because OSR-related barite appears to occur in both groups, the basis for "dominantly" should be stated more explicitly. I suggest adding a short summary table or a few sentences indicating how many samples or analyses in each group are associated with strongly ^{13}C-depleted calcite, moderately depleted calcite, and non-depleted calcite. It would also be useful to distinguish cases where the calcite is petrographically co-genetic with barite from cases where average $\delta^{13}C_{calcite}$ values from the barite-bearing sample were used. This would strengthen the process-level interpretation without requiring any change to the main conclusion.

- We agree that phrases like "dominantly associated" should have a quantitative basis. We suggest using a threshold of $\delta^{13}C < -40‰$ to define strongly depleted calcite since such values cannot be explained by organic carbon oxidation alone and requires at least an

admixture of carbon from methane oxidation. We added: *"Out of all classified barite data (i.e., excluding outliers), 71 spots in Group 1 and 79 spots in Group 2 had associated $\delta^{13}C_{calcite}$ data. Strongly ^{13}C-depleted calcite ($\delta^{13}C < -40$ ‰) was almost exclusively associated with Group 1 barite (27 out of 71) (Fig. 4c) (...). In contrast, only 2 out of 79 spots in Group 2 barite were associated with calcite that has a $\delta^{13}C \approx -30$ ‰ or higher (Fig. 4c)."* Rather than Group 1 being dominated by strongly ^{13}C-depleted calcite, we find that ^{13}C-depleted calcite is mostly restricted Group 1. We rephrased all respective statements accordingly.
- Based on the now apparent minority of AOM-associated barite even in Group 1, we modified our interpretation of the data: "*Thus, it seems unlikely that the different MSR pathways are the only control on the different slopes of both groups. Instead, we interpret the different $\delta^{18}O/\delta^{34}S$ trends to result from an interplay of MSR pathways and SRR: Group 1 reflects AOM and OSR at a relatively low SRR, whereas Group 2 dominantly reflects OSR at a relatively high SRR."*

3. Present the preservation-bias interpretation more cautiously or support it with a simple calculation.

The proposed explanation that extremely ^{34}S-enriched barite is absent because barite becomes undersaturated during advanced Rayleigh distillation is interesting. However, the current discussion is mostly qualitative. I encourage the authors either to provide a simple first-order calculation showing how sulfate concentration, Ba^{2+} availability and barite saturation state may evolve during Rayleigh distillation, or to slightly soften the wording of this conclusion. In particular, the maximum observed $\delta^{34}S_{barite}$ value of approximately +52‰ should be presented as the highest value recorded in the dataset rather than as a demonstrated saturation threshold. If no quantitative saturation calculation is added, the manuscript should more clearly acknowledge alternative explanations, such as temporal decoupling between pyrite and barite formation, limited Ba availability during late-stage sulfate reduction, post-formation dissolution, open-system mixing, or sampling bias.

- We agree that +52 ‰ should be presented as the highest measured value, rather than a generalized threshold and adapted our statement accordingly. In our opinion, calculating the evolution of the barite saturation state may add confusion, as it would depend on assumptions of paleo fluid compositions for which we have no constraints. Therefore, we thoroughly revised this discussion, highlighting alternative explanations to our suggested model in section 3.4 (§1).

4. Clarify the inference of sulfate reduction rates and sulfate/electron-donor controls.

The interpretation that differences in $\delta^{18}O/\delta^{34}S$ slopes reflect variations in sulfate reduction rates, fractionation and sulfate/electron-donor ratios is reasonable and consistent with previous conceptual models. However, the manuscript would benefit from a clearer distinction between directly observed isotope patterns and inferred microbial or hydrogeochemical controls. For example, higher sulfate availability does not necessarily imply higher sulfate reduction rates if electron donors are limited. Similarly, apparent $^{34}\varepsilon$, sulfate concentration and cell-specific sulfate reduction rate are related but not interchangeable. I suggest revising this section to state more explicitly that SRR differences are inferred from isotope systematics rather than independently measured. This small clarification would improve the logical flow of the discussion.

- Indeed, the possible variations in SRR between Group 1 and Group 2 are an interpretation, not an independent observation. The respective subsections (now 3.2 and 3.3) were thoroughly revised, making sure to clearly distinguish inferences from observations.

5. The comparison with Antler et al. and marine sedimentary systems should be expanded.

The manuscript notes an important discrepancy: the AOM-influenced Group 1 has a steeper $\delta^{18}O/\delta^{34}S$ trend than the OSR-dominated Group 2, and the Group 2 trend is shallower than OSR-related trends reported from marine sedimentary systems. This difference is important because it supports one of the manuscript's broader implications, namely that isotope interpretation frameworks developed for marine barite cannot be directly transferred to deep subsurface fracture-hosted barite. This point deserves a more mechanistic discussion. The authors should compare the deep fracture system with marine SMTZ settings in terms of sulfate concentration, Ba availability, electron-donor limitation, fluid residence time, openness of the sulfate reservoir, temperature, and barite saturation state. A short comparative table would be useful. Such a comparison would help readers understand whether the

observed $\delta^{18}O$–$\delta^{34}S$ trends are a predictable consequence of low-sulfate deep subsurface environments, or whether they require modification of existing conceptual models for sulfate oxygen and sulfur isotope evolution during microbial sulfate reduction.

- We agree that the comparison between the deep subsurface and marine environments is valuable and should be expanded. Many of the suggested discussion points had been scattered throughout our previous submission. We have now bundled these in the revised section 3.3 (§2) to address this topic more comprehensively, while highlighting a few additional differences between the continental subsurface and marine settings for which we consider their effect on $\delta^{18}O$–$\delta^{34}S$ trends poorly understood.

Minor comments

Please check isotope notation throughout the manuscript. Symbols such as $\delta^{34}S$, $\delta^{18}O$ and $^{34}\varepsilon$ should be formatted consistently in the abstract, main text, tables, figures and captions.

- We have proofread the manuscript for consistency as suggested.

Figure 4 would benefit from uncertainty envelopes around the regression lines. The caption should also clarify whether each plotted point represents one SIMS spot with paired $\delta^{34}S$–$\delta^{18}O$ data, one crystal, or one sample.

- We added uncertainty envelopes around the regression lines and clarified the caption of Fig. 4.

In the caption of Figure 5, two panels appear to be referred to as “(d)”. Please correct the panel labeling.

- We reorganized Figures 5 and 6, paying attention to avoid errors in the labels.

The BSE images in Figures 5 and 6 are useful but would be easier to interpret with clearer arrows or labels marking the key petrographic relationships discussed in the text. For the interpolated $\delta^{34}S$ heatmap in Figure 6, please clarify that the interpolation is illustrative and should not be over-interpreted beyond the density of SIMS analyses.

- Figures 5 and 6 have been thoroughly revised. We have clarified the mentioned petrographic relationships in the BSE images. Moreover, the scatter plots were removed, because they did not add value to the interpretation of the data.

Reference list formatting: The reference list should be revised carefully for consistency with the Biogeosciences style. In particular, please standardize journal-name formatting, as some journals are written in full whereas others are abbreviated. Please also check article-title capitalization, as some titles appear to use inconsistent capitalization styles. In addition, article titles containing isotope notation and chemical formulas should be checked for correct superscript/subscript formatting, for example $\delta^{34}S$, CH_4, and H_2S, rather than plain-text forms such as δ34S, CH4 or H2S.

- The reference list was corrected as suggested.

<u>Reviewer 2</u>

General comments:

1) The introduction accounts for more than a quarter of the total manuscript length and, as a result, feels overly long and at times difficult to follow. While the background material is generally relevant, several sections could be shortened or streamlined, allowing the authors to allocate more space to methods, results, and interpretation.

- We agree that the introduction is relatively long. We noticed that some concepts are introduced again in relevant sections of the discussion. We removed those from the introduction to avoid redundancies.

2) My primary concern with the manuscript relates to the statistical treatment of the isotope data, specifically the use of Gaussian Mixture Models (GMM) to define discrete groups that form the basis for much of the subsequent discussion. Although the Bayesian Information Criterion (BIC) favours a three-component model, this solution is explicitly rejected with limited justification. This decision appears subjective and is not discussed in sufficient depth. From a statistical standpoint, discarding the preferred model without a clear, objective rationale weakens inferential independence and risks introducing confirmation bias. If the data robustly support three components, the implications of this outcome should be explored rather than dismissed. Negative isotope slopes, for example, could plausibly reflect multiple generations of mineral growth, mixing of fluids with distinct histories, or other processes independent of the authors' preferred framework. Presenting and discussing all statistically supported solutions would ultimately strengthen the manuscript.

- Here, we would like to refer to our response to Reviewer 1's comment 1), where similar concerns were raised.

3) In addition, the observation that barite core analyses dominate one group while rim analyses dominate another may suggest that intra-crystalline processes during barite growth contribute to the overall spread observed in the data. This possibility is potentially important, particularly given the distinct zoning patterns shown in some samples (e.g., Fig. 6d), and deserves more explicit consideration. Because these group definitions underpin much of the later discussion, the current statistical treatment leaves the interpretations on uncertain footing. Addressing this would significantly improve the manuscript. I encourage the authors to reduce the length of the introduction and use the freed space to present a clearer and more transparent explanation of their clustering approach and its limitations. They could additionally explore other clustering methods. In any case, I would advise approaching the interpretation of the preferred statistical analysis with an unbiased mind, as potentially "unexpected" trends might be the result of significant mechanisms.

- We agree that the zoning into a Group 1 core and Group 2 rim points to differences during barite crystallization over time. The sharp increase in $\delta^{18}O$ and $\delta^{34}S$ from core to rim suggests two temporally decoupled precipitation events. The fluids reflected by the core and rim where likely affected by different MSR pathways and/or SRR, as indicated by the shift from Group 1 to Group 2 barite. We further clarified this in section 3.4: *"This sharp increase suggests that barite in the rim precipitated from sulfate that had already undergone ^{34}S-enrichment via sulfate reduction elsewhere in the fracture system (e.g., Drake et al., 2018). Additionally, SIMS spots related to Group 1 predominated in the core, whereas spots related to Group 2 exclusively occurred in the rim (Fig. 7b). This implies that the processing of the respective sulfate reservoirs reflected by core and rim occurred under different SRR and/or MSR pathways. Thus, the distinct textural and isotopic differences demonstrate that the generations reflect two barite precipitation events, rather than continuous barite growth.".* We prefer to avoid drawing implications for the whole dataset from this isolated observation.
- The more detailed explanation of our clustering approach has been added to the Method section and complement by new Supplementary Figures. The main limitation of our clustering approach lies in its application to a merged dataset including barites that formed independent from another, which we address in section 3.2: *" The clusters resulting from GMM are statistical entities and cannot be interpreted as continuous fractionation trends in our study. This is because our dataset contains multiple barite growth generations from different locations and records different fractionation trends, non-coeval sulfate reservoirs between growth zones, or local fluid mixing. Therefore, our statistical analysis can only detect average $\delta^{18}O/\delta^{34}S$ trends of overlapping fractionation or mixing processes involving fracture water sulfate that was processed by OSR and AOM at different points in space and time. These overlapping trends likely explain the low r^2 values."*

Specific comments:

4) A further concern relates to the strong imbalance between datasets from the two study sites. The sulfur isotope dataset for example is dominated by samples from Laxemar (323 analyses) compared to Forsmark (21 analyses). This imbalance likely affects the performance of the GMM and means that the derived clusters largely reflect the structure of the Laxemar dataset, with Forsmark data contributing relatively little statistical weight. As a result, it is difficult to view this as a truly comparative study of two localities. If additional Forsmark data cannot be obtained, I suggest either analysing sites separately or clearly acknowledging the limited interpretive weight of the Forsmark dataset. Alternatively, removing the Forsmark data altogether may lead to a more internally consistent analysis.

- It is true that the derived clusters mostly reflect the Laxemar dataset, which is the main reason why we did not attempt a comparative study between both sites. Unfortunately, additional data from Forsmark cannot be obtained in the framework of this study due to limited and cost intensive instrument time. Instead, we compiled data from two sites with known evidence for OSR and AOM (Drake et al., 2017, 2018b) to achieve a larger overall subsurface barite sample size. Due to the small sample size of the Forsmark dataset we cannot perform GMM for both datasets separately. However, we performed GMM for Laxemar alone to test the influence of Forsmark on the overall result. The results show that excluding Forsmark had no substantial impact on the outcome of the GMM (see Methods and Supplementary Figures). We clarified our reasoning behind this approach and its potential limitations: *"Data from Forsmark and Laxemar were merged for GMM to achieve a larger barite samples size, and a sensitivity test was conducted to exclude an influence of merging localities on GMM performance (see Supplementary Figure S2). Nevertheless, the resulting clusters should not be regarded as globally representative and may differ for barite from other areas.".*

5) The manuscript generally treats barite as a mineral that undergoes minimal isotope fractionation during precipitation and is subsequently resistant to dissolution. However, several observations appear inconsistent with this assumption, particularly the systematic differences between core and rim compositions and their interpretation in terms of barite dissolution. Potential kinetic isotope effects during precipitation are not considered, despite their relevance to zoning patterns. This raises questions about the assumption that barite isotopes always provide a faithful archive of dissolved sulfate compositions.

- Sulfur isotope fractionation between dissolved sulfate and barite during precipitation is negligible (<0.4 per mil; Kusakabe and Robinson, 1977) compared to the observed core-to-rim difference (e.g., ca. 12 per mil in sample KLX01_220; Fig. 7b). Thus, these effects cannot be discussed based on our data. Instead, the observed core-to-rim difference must be related to different processes affecting the sulfate reservoir over time or different mixing ratios of sulfate-bearing fluids between core and rim, or combination thereof. This was clarified as described in our response to this Reviewer 2's comment 3).
- Furthermore, barite is relatively resistant towards dissolution over a broad Eh/pH range, as well as diagenetic overprint, making it a highly useful long-term archive. However, we have not suggested that it could not occur. Indeed, we discuss barite dissolution as a potential reason for the absence of ^{34}S-enriched barite. To avoid speculation on fluid conditions during periods no longer recorded due to barite dissolution, we focus our discussion on the record remaining in our samples.

6) Related to this point, core-to-rim sulfur isotope trends in pyrite are attributed primarily to Rayleigh distillation during mineral growth. While this is certainly plausible, a simple quantitative model using literature fractionation factors would help demonstrate whether this mechanism can reproduce the observed magnitudes. Other potential explanations, including precipitation kinetics or diagenetic modification, should also be considered and discussed.

- Extensive discussion of Rayleigh distillation processes during pyrite formation including supporting evidence and model calculations are presented in the original publication that reported the respective pyrite data (Drake et al., 2018). We would prefer to avoid repeating those here.

7) The treatment of carbon isotopes in calcite is similarly simplified. The use of fixed $\delta^{13}C$ thresholds to infer dominant processes (e.g., -70 to -40‰ for AOM; > -20‰ for OSR) does not fully account for mixed carbon sources, partial methane oxidation, or other processes that can influence calcite $\delta^{13}C$

values. This oversimplification is particularly problematic where averaged $\delta^{13}C$ values are used for barite-bearing samples (Section 3.2, line 211). Given the wide range of observed values, such averaging may obscure meaningful variability and potentially bias subsequent interpretations.

- We agree that the carbon pool from which our calcite precipitated is likely affected by mixing. However, since we can only analyze the product, mixing of different carbon sources cannot be reconstructed in most cases and would be speculative. To avoid this, we have introduced a threshold of below -40‰ for AOM, since only then we can be certain that the observed $\delta^{13}C$ values cannot be explained by OSR (or inorganic carbon sources) alone. We consider this a more robust approach for identifying samples affected by AOM. Additionally, we softened our wording on the association of barite with OSR based on $\delta^{13}C_{calcite} > -30‰$, since we cannot exclude admixture of inorganic C from methane oxidation (e.g.: *"... This calcite likely reflects a more prominent contribution from organic matter rather than methane oxidation, implying that the associated barite was affected by OSR (Drake et al., 2017)".*)

8) The manuscript argues that oligotrophic conditions in the deep subsurface imply low sulfate reduction rates (SRR). However, the interpretations associated with Group 2 require relatively high SRR, which is reconciled by invoking large sulfur isotope fractionation factors (ε^{34}). While this explanation is possible, it remains largely unconstrained and risks becoming flexible enough to accommodate almost any trend. Independent constraints on SRR, or at least a more explicit discussion of uncertainties and alternative parameter combinations, would make this argument more convincing.

- We agree that this explanation is insufficiently explained by independent observations. In fact, the more pronounced ^{34}S-enrichment in Group 2 does not necessarily require a higher $^{34}\varepsilon$ but could also reflect a higher degree of sulfate consumption, possibly due to higher SRR. Therefore, we removed the discussion of variable $^{34}\varepsilon$ values. Instead, we now acknowledge the difference between the OSR-related slope in Group 2 and typically observed ones in marine environments as an unresolved problem. This adds to our conclusion that interpretation of barite stable isotope data cannot directly be transferred from marine environments to the continental subsurface.

9) Finally, the absence of isotopically heavy barite is not uniquely constrained by the current discussion. Several alternative explanations could be considered, including sulfate depletion prior to barite saturation, temporal decoupling between pyrite and barite formation, or preferential sampling of late-stage, sulfate-poor conditions. Incorporating barite saturation index calculations or explicitly evaluating these scenarios would strengthen this part of the manuscript.

- Indeed, our model involving barite undersaturation prior to strong ^{34}S-enrichment is not the only possible explanation. We have addressed this issue in our response to Reviewer 1's comment 3, where we also highlight the difficulties of a more quantitative treatment.

Minor comments:

10) Several parts of the manuscript, particularly in the introduction, rely on rather vague statements that would benefit from greater precision. For example, at line 45 the authors refer to a "range" of isotope fractionations without clearly defining what that range is. Providing representative values or at least approximate bounds would greatly improve clarity. Similarly, at line 116, Laxemar is described as being "a couple of kilometers" from the Baltic Sea coastline, while Äspö is located on the coast. Reporting the actual distances between these localities would be more informative and avoid ambiguity.

- We specified the distance between Laxemar and Äspö. However, we could not find an unspecified fractionation range in line 45 or elsewhere in our manuscript.

11) The manuscript would also benefit from careful attention to language and terminology. There are several grammatical and spelling issues (e.g., line 54: "the slopes (…) is") that should be corrected. In addition, some statements rely heavily on scientific shorthand or jargon, such as the phrasing at lines 73–74 where AOM and OSR are described as "producing trends." Rephrasing these statements in clearer, more explicit terms would make the manuscript more accessible to a broader readership. I

have not undertaken a comprehensive review of grammar and spelling, as the structure and wording of the manuscript will likely change substantially in response to the major revisions outlined above. Nonetheless, I strongly recommend that the final revised version be carefully edited for language, clarity, and consistency prior to resubmission.

- We checked the manuscript for spelling mistakes, grammar, and unnecessary jargon before resubmission.

12) Finally, regarding QA/QC, I was unable to identify data from standards or reference materials that would substantiate the uncertainty estimates and error propagation described in the methods section. Including these data, or explicitly stating where and how they are reported (e.g., in supplementary materials), would strengthen confidence in the analytical results.

- We have now summarized the independently determined isotope values of reference materials this information in a separate data sheet (S4) within the Supplementary Data Table. Measurements of barite reference materials during analytical sessions, are listed in the respective data sheet of each individual session (S5-S10). Measurements of pyrite and calcite reference materials are published in the cited original literature (Drake et al. 2017, 2018).

References

Drake, H., Heim, C., Roberts, N. M. W., Zack, T., Tillberg, M., Broman, C., Ivarsson, M., Whitehouse, M. J., and Åström, M. E.: Isotopic evidence for microbial production and consumption of methane in the upper continental crust throughout the Phanerozoic eon, Earth Planet. Sci. Lett., 470, 108–118, https://doi.org/10.1016/j.epsl.2017.04.034, 2017.

Drake, H., Whitehouse, M. J., Heim, C., Reiners, P. W., Tillberg, M., Hogmalm, K. J., Dopson, M., Broman, C., and Åström, M. E.: Unprecedented ^{34}S-enrichment of pyrite formed following microbial sulfate reduction in fractured crystalline rocks, Geobiology, 16, 556–574, https://doi.org/10.1111/gbi.12297, 2018a.

Drake, H., Whitehouse, M. J., Heim, C., Reiners, P. W., Tillberg, M., Hogmalm, K. J., Dopson, M., Broman, C., and Åström, M. E.: Unprecedented ^{34}S-enrichment of pyrite formed following microbial sulfate reduction in fractured crystalline rocks, Geobiology, 16, 556–574, https://doi.org/10.1111/gbi.12297, 2018b.

Kusakabe, M. and Robinson, B. W.: Oxygen and sulfur isotope equilibria in the BaSO4□HSO4−□H2O system from 110 to 350°C and applications, Geochim. Cosmochim. Acta, 41, 1033–1040, https://doi.org/10.1016/0016-7037(77)90098-9, 1977.