Dear reviewer,

First, we want to sincerely apologize for the delayed response to your review. Second, we wish to express our gratitude for your thoughtful and detailed revision. We highly appreciate your thorough and constructive feedback, of which we believe that it will help to significantly improve our manuscript. We revised the manuscript following your valuable suggestions. In almost all cases, we followed your suggestions. All changes have been incorporated directly in the revised version. The changes were made with track changes on. Below we answer your individual comments, which require discussion in detail.

Thank you once again for your great reviewing work.

With kind regards, on behalf of all authors,

Raphael Burchartz

General Comments

The authors have conducted extensive research to understand the role of burial history in self-sealing capacity of several argillaceous host rocks. The authors have focused on hydraulic conductivity as a key characteristic for the sealing integrity of host rocks in the context of high-level nuclear waste disposal in Europe (specifically in Germany, Switzerland, Belgium, France, and Hungary). They have compiled a large dataset including geological, petrophysical, mechanical, and mineralogical data from OPA, COx, To-Do, AMA, BCF, and Boom formations at laboratory and field scales. The manuscript narrates well from Introduction to Methodology, Results, and Discussion; however, I see room for improvement. These mainly involve language enhancement, reference placement, literature improvement, terminology correction, and results-discussion orientation.

Specific Comments

General

- The manuscript uses a mix of American and British English (e.g., "behaviour"/"behavior"). Please choose one and maintain consistency throughout. Additionally, proofread the manuscript thoroughly for grammar, structure, and fluency.

Indeed, thank you very much for pointing that out. We have adapted the whole manuscript to British English.

- The manuscript is filled with abbreviations (e.g., CCC, URL, LSB, OCR). Use abbreviations only if they are well-understood, repeated multiple times, or core to the paper. If abbreviations are used, define each one once in the text and use it thereafter.

We agree and adapted the manuscript accordingly.

- Log returns no unit; Log_{10} K is dimensionless.

Good point, and we fully agree that the logarithm itself is dimensionless. We prefer to retain the notation log10K with the unit of K indicated in parentheses to make the reference unit explicit for the reader, since numerical values depend on whether K is expressed in m/s, m/d, or other equivalent units. This notation is widely used in hydrogeological literature for clarity (see i.e. http://dx.doi.org/10.1002/2017JB014106; https://doi.org/10.1146/annurev-earth-053018-060437) , although we are happy to revise it if the journal prefers strict dimensional formatting.

Abstract

- Line 13: "Hydraulic integrity" is uncommon. "Sealing integrity" is recommended.

We agree and changed accordingly.

Introduction

- Lines 35-40, 42-47: These lines require references.

We agree. Appropriate references were added to the specific sections.

- Lines 41-47: Please provide the equation for hydraulic conductivity.

The equation was added to the revised manuscript. Thank you very much for this suggestion.

- Lines 62-63: Regarding "stress and temperature," specify which stress is being referred to. Pore pressure plays a key role in burial history too as it controls effective stress.

Thank you very much for this important consideration. We completely agree that pore pressure evolution is a critical factor as it directly dictates the effective stress history. By "stress conditions" we meant the comprehensive, time-dependent stress state evolution that the formation experienced along its entire burial history (i.e., the combined effects of total overburden stress, pore fluid pressure, and any lateral tectonic stresses).To clarify this and avoid ambiguity, we reformulated the sentence to: "An important factor shaping the hydraulic characteristics of argillaceous formations is their burial history, including the associated stress states (dictated by the interplay of overburden, pore fluid pressure, tectonic stresses and the resulting effective stress) and temperature conditions."

- Lines 65-75: The critical depth separating mechanical and chemical compaction (e.g., cementation) is typically 2-3 km. Temperatures greater than 80-100 °C are commonly used for chemical compaction (e.g., see https://doi.org/10.1144/GSL.SP.2006.253.01.19). Please revise or provide a reference if using "> 70 °C."

Thank you for pointing this out. In argillaceous formations, the onset of key mineralogical transformations (such as the smectite-to-illite transition) is widely recognized to begin at temperatures around 70–80°C. We have added the following references to support this in the text:

- **Bjørlykke, K. (1999). Principal aspects of compaction and fluid flow in mudstones. Geological Society, London, Special Publications, 144(1), 73-84.**
- **Bjørlykke, K. (2014). Relationships between depositional environments, burial history and rock properties. Some principal aspects of diagenetic process in sedimentary basins. Sedimentary Geology, 301, 1-14.**

However, we agree that the text can be enhanced to make it clear that this temperature represents a transitional range rather than a fixed threshold. We adapted it accordingly.

- Lines 108-111: The contents within the brackets are inconsistent. Please adhere to one format, either location or sample ID.

Thanks for bringing this to our attention! We adapted accordingly.

Materials and Methods

- Lines 128-130: Boom is missing from the caption.

Thank you very much. Boom was added to the caption.

- Lines 167-168: The character "x" has been used instead of a multiplication symbol in values (×). This also applies to Table 1. Ensure consistency throughout.

Thank you very much for pointing that out. We corrected for all instances.

- Line 288: Please clarify the nature of the permeability (hydraulic conductivity) and porosity data at both laboratory and field scales. Are these matrix permeability/porosity, fracture permeability/porosity, or both?

The compiled porosity data are exclusively laboratory measurements obtained from intact samples, primarily using water-loss methods (105 °C drying),. Macroscopically intact samples may still contain micro fractures, making it difficult to assess their contribution to the overall porosity and permeability measurements. Field-scale values may reflect a response of both the matrix- and fracture-controlled rock-mass hydraulic behaviour. We have clarified this distinction in the revised manuscript and added the underlying methods.

- Figure 2: The legend of Figure 2 lists "ToDo," while the text repeats "To-Do." Ensure consistency throughout.

Indeed, thank you for bringing this to our attention. We have adapted the legend.

Table 2: Permeameters are typically used with samples as received or dried; "saturated" can be misleading. Researchers commonly employ steady-state or unsteady-state methods to determine permeability under single-phase flow conditions. For samples that do not contain swelling clays, saturation with water and subsequent measurement of permeability using that water might be feasible, as might be the case for sandstones. However, for low- to ultra-low permeability samples (e.g., mudrocks), gas is typically used to determine permeability (slip-corrected permeability). Saturating such samples with water is not recommended because it creates two-phase flow, and the resulting measurement would not represent gas permeability but rather diffusivity, or at best, gas effective permeability. Furthermore, the samples used are commonly core plugs, not entire cores.

Thank you for these considerations. We adapted the wording in the table to acknowledge, that commonly core plugs are used. The use of testing fluid (gas or water) depends on the research question, for example shale gas studies require gas perm, rad waste is more concered with hydraulic conductivity. For water-based permeability measurements, the sample must be fully saturated to ensure single-phase flow conditions, which is a prerequisite for obtaining a representative measurement. You are right to mention that if we not use formation water for saturation we will create a disequilibrium that might affect transport properties. But this effect is less strong when compared to full drying and subsequent gas perm measurements. As-received is useless if you do not know the saturation state.

- Lines 331-345:

1. Clarify whether "Burial estimates" refers to maximum burial depth or present-day burial depth. Table 4 suggests it is maximum burial depth; better to list both depths in Table 4.

Thank you very much for pointing this out. Indeed, what we mean with "burial estimates" is the maximum burial depth. We acknowledge that this was not clear and adapted the sub-section title to "3.3.2 Maximum burial depth estimations".

2. The estimation of maximum burial depth is provided for OPA and AMA but not for the other formations. Please provide these estimates for all formations studied.

Please note that maximum burial depth estimates for each formation were already included in the individual formation descriptions in Chapter 2. Additional methodological details were originally provided only for the OPA because the compiled dataset comprises samples from different investigation sites, for which site-specific burial history reconstructions are not consistently available. Instead, maximum burial depths were assigned based on the corresponding sub-regions. In contrast, for the AMA, site-specific burial history reconstructions were available for each investigated location.

However, we fully agree that information on maximum burial depth estimates and the underlying methods should be given for all considered formations in order to keep consistency and flow for the reader. The section has been revised accordingly and methodological approaches of the references were briefly presented in table 4.

3. Several sentences in this section do not end with a period (.). Please add them.

Indeed, thank you very much for making us aware of that. We corrected for all instances.

4. "T_{max}" is commonly used, not "T_{peak}."

Thank you for this important comment. We acknowledge that Tmax is indeed used in parts of the basin modelling and thermal maturity literature. However, Tmax is often the established term for the temperature of maximum hydrocarbon generation determined by Rock-Eval pyrolysis (typically ~400–500 °C).

Since our manuscript discusses maximum temperatures attained during geological burial (typically ~50–200 °C), we prefer the term Tpeak to avoid potential confusion between these two fundamentally different parameters. This distinction is also adopted in parts of the literature. See i.e.:

- **Castro-Vera, L., Amberg, S., Gaus, G., Leu, K., and Littke, R.: 3D basin modeling of the Hils Syncline, Germany: reconstruction of burial and thermal history and implications for petrophysical properties of potential Mesozoic shale host rocks for nuclear waste storage, Int J Earth Sci (Geol Rundsch), https://doi.org/10.1007/s00531-024-02384-z, 2024.**
- **Grobe, A., Littke, R., Sachse, V., and Leythaeuser, D.: Burial history and thermal maturity of Mesozoic rocks of the Dolomites, Northern Italy, Swiss J Geosci, 108, 253–271, https://doi.org/10.1007/s00015-015-0191-2, 2015.**
- **Hardebol, N. J., Callot, J. P., Bertotti, G., and Faure, J. L.: Burial and temperature evolution in thrust belt systems: Sedimentary and thrust sheet loading in the SE Canadian Cordillera, Tectonics, 28, 2008TC002335, https://doi.org/10.1029/2008TC002335, 2009.**
- **Burchartz, R., Seemann, T., Gaus, G., Winhausen, L., Jalali, M., Mbui, B. M., Grohmann, S., Burnaz, L., Colling Cassel, M., Erbacher, J., Littke, R., and Amann, F.: The influence of burial history on physical rock properties: a case study of Lower Jurassic claystones from the Hils and Sack Synclines (Germany), Solid Earth, 17, 485–512, https://doi.org/10.5194/se-17-485-2026, 2026.**

To further improve clarity, we have revised the terminology to "peak burial temperature (Tpeak)" throughout the manuscript.

5. Basin modelling, maximum burial temperature, and vitrinite reflectance are commonly employed to estimate maximum burial depth (e.g., see https://doi.org/10.1016/j.energy.2023.129966). It is suggested to highlight the methodologies used in this study and specify which method was employed to determine the maximum burial depth for each formation.

That is a valuable suggestion. We agree and added the applied methodologies for burial depth estimations for each formation as given in the respective studies. We also updated table 4 accordingly in order to provide a more comprehensive overview.

Results

- Table 5: Please specify what "mbgs" and "Geom." stand for in the caption.

Done.

- Figure 3: The in-situ K data were determined under a specific in-situ effective stress. Were the laboratory K data obtained under the same in-situ effective stress? If not, the observed differences in K could potentially be attributed to variations in effective stress.

Thank you for raising this important point. We addressed this issue in Subsection 4.3.3, "Stress dependency of laboratory data" and 4.3.4 "Summary and baseline definition." Wherever reported in the original studies, we compiled the effective stress applied during laboratory experiments and considered it when interpreting the data. As shown in Figure 8, hydraulic conductivity may decrease with increasing effective stress in some formations and datasets, but the observed variation is generally limited to about two orders of magnitude, with values largely remaining within the 10^{-14} to 10^{-12} m/s range. Thus, while differences in effective stress can contribute to scatter in the laboratory data, they do not explain the full separation between laboratory and in-situ values in Figure 3.

- Lines 395-410:

1. The Y-axis in Figure 4 denotes "Depth." To avoid confusion between present-day burial depth and maximum burial depth, please be more specific about the Y-axis in Figures 4 and 6, similar to how it is done in Figure 11.

   We fully agree and adapted the figures accordingly.

2. The K medians for lab and field data converge more closely if the filtered depth is increased. Please justify the selection of present-day burial depth > 100 m.

Thank you for this comment. We selected a present-day burial depth of >100 m as a pragmatic threshold to exclude the shallow near-surface interval affected by decompaction, weathering, and other stress-relief effects that strongly influence in situ hydraulic conductivity. Our intent was to compare laboratory and field data under conditions representative of the more intact rock mass, where scale effects are less strongly overprinted by near-surface alteration.

The choice of 100 m is supported by the depth trends shown in Fig. 4, where the strongest decrease in in situ hydraulic conductivity occurs within the upper ~100 m, while data below this depth show substantially reduced scatter and a much clearer convergence between lab and field values. We agree that shifting the cutoff modestly would alter the medians somewhat, but not the overall interpretation. We now clarify in the caption that the 100 m threshold is used as a practical filter to isolate the decompaction-affected zone, and that the detailed implications of this depth dependence are discussed separately in the Discussion section.

3. Significant reductions in conductivity or porosity are typically observed within the first 100 m. Excluding this 100-m interval raises concerns, as the sealing integrity of host rocks needs to be proven for the entire depth, even if they have undergone weathering and decompaction at shallower depths. You may consider highlighting this as a limitation of the study.

We thank the reviewer for this important remark. We agree that the uppermost 100 m represent a critical depth interval, as changes in hydraulic conductivity and porosity within this zone are highly relevant for repository safety assessments. This near-surface zone can extend from a few metres to several tens of metres below ground and is strongly influenced by factors such as mechanical rock properties, exhumation history, local tectonics, and topography. In most cases, potential repository host rocks are overlain by other formations or rock sequences that shield them from significant near-surface processes. Below this zone, laboratory-derived K and in situ K values tend to overlap more closely, allowing a more meaningful comparison of intrinsic properties.

In our analysis, we did not exclude data from this interval but treated the upper 100 m separately from deeper intervals. This distinction was made because the uppermost zone is typically dominated by decompaction, weathering, and stress-relief effects, which can cause extreme scatter in hydraulic conductivity (K) values and overprint the intrinsic rock-matrix properties. Please note that this study is not intended to assess the suitability of specific HLW repository sites. Rather, it aims to improve our understanding of how burial history, including both burial and exhumation, controls the evolution of hydraulic conductivity.

4. Figure 11 displays K and porosity versus maximum burial depth. It is suggested to include a plot of porosity versus present-day burial depth (laboratory and in-situ) in Figure 4, similar to the K versus Depth plot, if such data are available. This would help maintain the link between present-day and maximum burial depths.

Thank you very much for this great suggestion. We agree that adding a porosity versus present day depth plot will benefit linking past- and present day depth. As no in-situ data for porosity measurements is available, a plot only for laboratory derived porosity can be provided.

5. Figure 5: Replace "$\log_{10}$ Hydraulic Conductivity (m/s)" on the X-axis with "$\log_{10}$ K". It is suggested to use symbols rather than full terminology on the axes of figures throughout the manuscript for clarity.

We agree and adapted the figures accordingly.

6. Lines 477-489: Please clarify what is meant by "effective confining stress." Do you intend to refer to the (modified Terzaghi) effective stress, calculated as , where is the (modified Terzagi) effective stress, is the confining stress, is the Biot coefficient, and is the pore fluid pressure? It is assumed that you mean to discuss hydraulic conductivity (K) behaviour as a function of "effective stress." The subsection uses "effective confining stress," "effective stress," "confining stress," and "stress," while Figure 8 labels the X-axis as "effective confinement stress." This lack of consistency is confusing. Please be more specific in your terminology.

This is an important point. Thank you very much for bringing it to our attention. We agree that the terminology in this subsection was not sufficiently consistent. We revised the whole section to clearly state the type of applied stress during different laboratory experiments (vertical effective stress in case of constant-head tests during oedometer experiments and effective confining stress during pressure-pulse decay tests). The labels x-axis labels of figure 8 were adapted accordingly.

- Line 502: It is assumed that "stress conditions" should be "effective stress conditions." The same applies to line 517.

Indeed, we have adapted accordingly.

- Lines 554-557: "However, if porosity alone would determine hydraulic conductivity, COx would be expected to show higher K values than the OPA and AMA, both of which have porosity values approximately 3–4 % lower than COx. Instead, COx generally exhibits lower hydraulic conductivity." This observation is natural, as porosity alone does not solely control permeability or hydraulic conductivity (k or K). Porosity is a single metric that cannot fully represent the entire pore structure; pore size distribution (PSD) is a more informative parameter. Moreover, fractures, while not significantly contributing to total porosity, largely control permeability. Figure 11 supports this observation. Although you focus on fractures, consider discussing the role of PSD in controlling K in the Discussion section.

Thank you very much for pointing this out. We fully agree that porosity alone does not solely control K. What was meant here is, that at similar porosity ranges in claystones, second-order controls (mineralogy, connectivity, tortuosity, PSD) become important in controlling the hydraulic conductivity. We have adapted the respective section to fully acknowledge this fact.

Discussion

The section discusses the self-sealing capacity of the host rocks, which is related to their hydraulic conductivity. It indicates that this capacity is controlled by swelling clay minerals and fractures. However, it is unclear what type of hydraulic conductivity data (matrix, fracture, or both) was used in the Materials and Methods section. This ambiguity has not been resolved in the Results section either. Furthermore, the swelling clay mineralogy and fracture properties of the formations (laboratory or in-situ) have not been specified in the Materials and Methods or described in the Results. The discussion appears to be based on metrics that were not previously measured or described. Please clarify.

Thank you for this important clarification request. We agree that the manuscript did not sufficiently distinguish between matrix-scale and rock-mass-scale hydraulic data. The compiled laboratory data predominantly represent matrix hydraulic conductivity measured on intact core plugs under controlled conditions, whereas the in situ data represent rock-mass hydraulic conductivity and may include the effects of fractures, bedding-plane discontinuities, and near-field disturbance where present. We have revised the Materials and Methods, Results, and Discussion to make this distinction explicit.

We also clarify that swelling clay mineralogy and fracture properties were not newly measured in this study for all formations, but were compiled from the original source studies where available and used as qualitative context for interpreting self-sealing behaviour. The discussion of self-sealing is therefore based on the hydraulic dataset and the associated literature on argillaceous rock behaviour, not on a direct new measurement of clay mineralogy or fracture geometry in each formation.

Please note that we have added a new subsection entitled "5.4 Scope and limitations of the compiled dataset" in response to this and several related comments. The new section explicitly discusses the strengths and limitations of the compiled dataset, clarifies which conclusions are directly supported by the data and which represent literature-based geological interpretations, and outlines the principal uncertainties associated with the compilation.

Conclusions

- Line 818: Pore size distribution is also a key characteristic that should be mentioned.

Thank you very much for pointing this out. We fully agree that pore size distribution is an important key characteristic in controlling fluid flow in argillaceous media. We considered this to be included under the umbrella term "microstructure". However, we see the need to designate this specifically and adapted the sentence accordingly.

Technical Corrections

- Line 12: "to a large extend" is incorrect. Use "to a large extent." **Done**

- Line 366: "ssamples" is a typo; it should be "samples." **Done**
- Line 837: “depth” should be replaced with “depths.” **Done**