Dear Reviewer,

Thank you very much for your detailed and constructive review of our manuscript. We sincerely apologize for the delayed response, which was due to other professional obligations and, importantly, our wish to carefully consider and fully address the thoughtful comments and suggestions you provided. We greatly appreciate the time and effort you invested in reviewing our work and believe that your comments have helped us to substantially improve the manuscript.

We have carefully considered all points raised and have made corresponding revisions to the manuscript. For clarity and conciseness, minor corrections, such as typographical errors, spelling, and references, have been implemented directly in the revised manuscript and are therefore not repeated in this response letter. For comments concerning the content of the manuscript, our responses are provided directly below each comment in bold green font. All changes made to the manuscript can be followed in the newly uploaded version, in which track changes have been enabled.

Once again, thank you very much for your great reviewing work, which is highly acknowledged.

With kind regards, on behalf of all authors,

Raphael Burchartz

General Comments

This manuscript presents substantial work in compiling and integrating a comprehensive dataset of hydraulic conductivity (K) measurements derived from both laboratory and in situ tests for six potential host rocks in Europe: COX, OPA, BC, To-Do, AMA, and BDF. The authors aim to identify a systematic links between the maximum burial depth and K by investigating burial history, uplift-related decompaction, cementation, mineral alteration and scale effect on K values of six host rocks.

The text and analysis are generally clear and thorough. The manuscript is supported by an extensive bibliography, comprising 1,250 references, which reflects the considerable effort invested in this work. The study appears to be primarily a synthesis of existing knowledge, and it does not introduce particularly novel findings. Given the distinct characteristics of the six formations considered and their heterogeneity, which make it challenging to derive well-validated general conclusions.

Specific Comments

General:

1. Figure 12 is a key graph used to support the main conclusions, where "maximum burial depth" is used for the y-axis. It assumes that K measured from the same region/borehole share a single value of maximum burial depth. However, maximum burial depth is a parameter with considerable uncertainty and is often difficult to quantify accurately. Table 4 provides approximate values for each region/borehole, some of which are derived from 1D basin modelling.
2. one of the key conclusions of the authors is that shallowly buried (<400 m), poorly indurated formations exhibit limited hydraulic variability and scale independence (L24). However, among the six host rocks investigated, only Boom Clay falls within the <400 m maximum burial depth category. Boom Clay is a relatively homogeneous clay formation, characterized by strong self-sealing capacity, relatively uniform hydraulic conductivity (K), and limited scale dependence, which is one of the reasons it has been selected as a potential host rock for HLW disposal.

   Given this, it is questionable whether the properties of Boom Clay can be generalized to all shallowly buried clays. At a minimum, additional evidence from other shallowly buried formations would be required to support this conclusion.

These are indeed important remarks, thank you very much for pointing this out.

<u>Regarding point 1:</u> We fully agree that maximum burial depth must be treated as a parameter with considerable uncertainty. We have revised section 3.3.2 ("Maximum burial depth estimates") to describe the individual methods underlying each formation's burial depth estimate more explicitly, and have added a column to Table 4 that now provides a quick overview of the method used per formation/borehole (e.g., 1D basin modelling, vitrinite reflectance). We acknowledge the resulting uncertainty, and its implications

for the comparisons drawn in Fig. 12, in a newly added discussion subsection, Sect. 5.4 ("Scope and limitations of the compiled dataset").

<u>Regarding point 2:</u> **We agree that formations with maximum burial depth <400 m are underrepresented in our compilation, currently comprising only the Belgian Boom Clay Formation. We considered other potentially suitable shallowly buried argillaceous formations but found that they either lack sufficient publicly available field-scale (in-situ) hydraulic conductivity data for a robust matrix–field-scale comparison (e.g., the Ypresian Clay, Belgium, largely sampled from the same boreholes as Boom Clay), or are mineralogically distinct in ways that would complicate direct comparison with our clay-mineral-dominated formations (e.g., the opal-A/opal-CT-dominated, clay-poor siliceous mudstones of the Wakkanai and Koetoi formations, Japan; Ishii et al., 2011). We now discuss this limitation explicitly in Sect. 5.4, together with the reasoning above, and note that extending the shallow end-member of this dataset to additional, mineralogically comparable formations with adequate field-scale data remains an important direction for future work. Same applies for formations that faced burial of depth >2,000 m. We have also softened the corresponding statement in the Abstract and Conclusions to reflect that this trend is currently based on a single formation (see also our response to your general comment on the "first-order control" wording below).**

1. data from Lausen is excluded from Figure 12 due to the lack of burial depth information. If the Lausen dataset, characterized by K values spanning approximately 7 orders of magnitude (purple triangles near the surface in figure 4b) , were included in Figure12a, the conclusion that “for depths < 2,000 m, K_lab and K_in situ show a close agreement” would no longer be valid.

Thank you very much for these important remarks.

Indeed, we do not have information about the burial history of the OPA samples from Lausen. However, these data were omitted from Fig. 12 because tests were performed at very shallow present-day depths, within a near-surface decompaction and weathering zone (Sect. 5.3.1) where fracture-related processes, rather than burial-related compaction and cementation, control the hydraulic properties (the same applies to the Swabian Alb OPA and AMA data omitted for the same reason).

As discussed in Sect. 5.3.1, this near-surface zone typically extends to depths of approximately 30–100 m, though its extent varies between formations and localities: hydraulic conductivity in the Lausen OPA borehole decreases sharply within the upper 30 m (Vogt et al., 2017; Crisci et al., 2019), whereas elevated values persist to ~40–60 m in the Swabian Alb OPA (Hekel, 1994; Vogt et al., 2017) and to ~95 m in the AMA (borehole BO1). We attribute this variability primarily to differences in uplift/exhumation history and local topographic-geomorphological evolution between sites (Hekel, 1994; Vogt et al., 2017), rather than to a fixed depth threshold.

The pronounced, ~7-order-of-magnitude spread in the Lausen dataset (Fig. 4b) is, in our interpretation, consistent with this near-surface fracture-dominated overprint rather than representing burial-history-controlled matrix behavior, and is comparable in character to the pattern discussed for the AMA (BO4.0) and Swabian Alb OPA data (Sect. 5.3.1). Including the Lausen dataset in Fig. 12 would therefore conflate two distinct controls: near-surface weathering/decompaction and burial-related compaction, that we treat separately in this study.

We agree that this exclusion criterion was not stated explicitly enough in the manuscript. We will revise the Fig. 12 caption and the relevant methods text to state clearly that datasets lacking reliable burial depth information, or known/inferred to be affected by near-surface decompaction and weathering (Lausen, Swabian Alb OPA, AMA), were excluded from Fig. 12 for this reason.

2. In Figure 12a , five boreholes/regions are included for depths greater than 2000 m, namely To-Do, BO3-BO5, and BCF. The To-Do formation includes 8 laboratory and 6 in situ measurements (Table 5). BDF has only 3 laboratory results and no specific in situ data. BO3, BO4, and BO5 include only laboratory tests.

In total, only 17 K measurements (8+6+3) for K_lab and K_in situ coming from To-Do and BCF. Such a limited samples raises the question of whether it is sufficient to support the conclusion that K_lab and K_in situ diverge for max burial depths greater than 2000 m.

1. The points raised above collectively challenge the robustness of the manuscript's final conclusion (Line20) that maximum burial depth and the associated stress and temperature conditions exert a first-order control on matrix-scale hydraulic conductivity (K).

This conclusion appears oversimplified. Hydraulic conductivity (K) is influenced by a range of factors. History of burial, stress and T are only part of the controlling processes. Mode first-hand data and rigorous analysis are needed for characterizing them as "first-order" controls.

We agree that the dataset available for maximum burial depths >2,000 m is limited (17 K measurements from To-Do and BCF combined), and we acknowledge this explicitly in the revised limitations section (Section 5.4). We would like to note, however, that despite the limited sample size, the .divergence between matrix- and field-scale hydraulic conductivity at these depths, though based on a limited number of formations, is in agreement with the two-stage burial evolution independently proposed by Mazurek et al. (2009), based on a broader, separately compiled dataset. We consider this convergence of independent lines of evidence to lend support to the general pattern, even though we agree that the absolute number of measurements at these depths remains small and warrants further first-hand data to more rigorously constrain it.

Regarding the characterization of maximum burial depth as a "first-order control" on matrix-scale hydraulic conductivity: we fully agree with the reviewer that this wording overstates what our dataset can support, given that K is influenced by multiple interacting factors (e.g., original mineralogical composition, spatial mineral and grain-size distribution, bedding anisotropy, fracture network development), of which burial history is only one. We will revise the concluding statement to more accurately reflect the nature of our findings, e.g.: "The compiled and discussed data indicate that maximum burial depth, and the associated stress and temperature history, emerge as a primary and consistent control of hydraulic conductivity across the investigated formations. Together with mineralogical composition, and spatial mineral and pore-size distribution, these burial-related processes govern the spatial and temporal evolution of hydraulic conductivity in argillaceous formations." This reformulation preserves the empirical pattern we observe while avoiding an overstated causal claim that the present dataset cannot fully substantiate.

Abstract

- L14: (K<10-10 m/s): Please provide an appropriate reference to support this value. In the context of safety assessments for HLW, values higher than 10^{-11} m/s are generally considered difficult to justify.

A rock mass hydraulic conductivity of <10^{-10} m/s corresponds to the legally defined minimum requirement for host rock suitability specified in the German site selection act (StandAG). We will add this reference explicitly. We note that the reviewer's suggested threshold of ~10^{-11} m/s may reflect stricter or differently defined criteria used in other national programmes or safety-case contexts; we will clarify in the text that the value cited refers specifically to the German regulatory framework.

- L15: strong aquitard characteristics: Please specify explicitly what characteristics define a strong aquitard in this context. For a rock formation to be considered a promising host for HLW, not only K, many other properties must be taken into account.

"Aquitard" is used here in its standard hydrogeological sense (a low-permeability unit limiting groundwater flow relative to adjacent units; Freeze & Cherry, 1979), and in our opinion this does not require further definition. We agree, however, that host rock suitability for HLW disposal depends on a broader set of properties beyond hydraulic behaviour (e.g., sorption/retardation capacity, mechanical stability, geochemical buffering, thermal properties), which are outside the scope of this study. We will revise the sentence to avoid the impression that aquitard characteristics alone determine suitability.

Introduction

- L53-54: the potentially strong spatial anisotropy of hydraulic properties, which can vary across different scales, partly by several orders of magnitude – The discussion here relates to the difficulty of measuring hydraulic conductivity (K) at larger scales due to variability. Anisotropy itself does not necessarily make hydraulic conductivity (K) measurements more challenging. The term heterogeneity would be more appropriate than anisotropy in this context.

We fully agree and adapted accordingly.

- L55: anisotropy arises from …orientation of the bedding… , …fractures, and heterogeneities in sedimentological: Please clarify between anisotropy and heterogeneity. Heterogeneity refers to spatial variability of properties and is a broader concept, whereas anisotropy specifically describes directional dependence of those properties (i.e., variation with the direction of measurement).

We have revised the passage to clearly distinguish anisotropy (directional dependence, primarily from bedding orientation) from heterogeneity (spatial variability from fractures, sedimentological differences, and mineralogical facies), and checked the consistent use of both terms throughout the manuscript.

- Paragraph A (Ln78-89)and paragraph (Ln 90-98): There is some overlap between these two paragraphs. Paragraph A introduces the idea that the rock may undergo either ductile or brittle deformation (L84) and provides explanations for both mechanisms. Paragraph B restates the mechanism for fracture formation in the context of uplift for host rocks. To improve clarity and avoid redundancy, consider consolidating these sections and presenting a unified explanation of fracture formation mechanisms.

You are right, there is overlap and redundancy in these two sections. We have consolidated the two paragraphs by removing the redundant statement and unified in a single explanation for enhanced clarity and flow. Thanks a lot for the suggestion.

Studied Formations and Sites

- L138: the formation is considered homogeneous – Typo error **Done.**
- L145: using pulse or constant-head permeameter tests – For Boom Clay, it is constant-head permeability test that is used to derive K value in the lab. Pulse is for the subsequent diffusion test.

Thank you for this correction. You are right. Upon reviewing the methodological source, we confirm that hydraulic conductivity is indeed derived from a (quasi-)constant-head percolation approach, in which a sustained pressure difference across the sample generates steady-state flow used to calculate K via Darcy's law. The term "pulse" in the original source refers specifically to the instantaneous tracer injection used to determine migration/diffusion parameters within the same experimental set-up. We will revise the text accordingly.

- L157: pronounced overconsolidation – Please add a range of over-consolidation ratio (OCR) for AMA, To-Do, COX, and BCF.

Done. Please note that precise, measured OCR values (e.g., from oedometer consolidation tests) are not available for Cox, AMA and To-Do in the literature compiled for this study. We therefore provide conservative, burial-history-derived estimates, based on the reconstructed maximum burial and present-day depth ranges. We acknowledge this uncertainty in the respective sections.

- L160: variable depths (formation tops between 2 and 74 mbgs)
  Table 1: "mbgs" may not be clear for all readers. Please clarify it. **We clarified the abbreviation.**

Methods

- Table 2 – Aertsens et al., (2004, 2008, 2013); reported Lab results for permeameter tests, not oedometer tests.

Indeed, thank you for pointing that out. We adapted accordingly.

- In permeameter test, it is constant pressure gradient that is used for K measurement.

Yes, but also pressure decay after pulse injection (pressure pulse decay experiments).

Results

- L 398: in situ data shows a pronounced decrease in hydraulic conductivity within the upper 100 m of nine orders of magnitude – The wording is unclear and misleading.

We agree and adapted the wording for better readability and clarity.

- L399 and Figure 4: laboratory data from the AMA and OPA lie in a distinctively narrower range between

- On Figure 4a, data for AMA and Lausen are represented by two dots. Please draw all Klab points available on the figure. Also, please clarify the sample size of data from Lausen borehole.

Done. Please note, that no individual lab data from the borehole Lausen was available and only the data-range could be considered. Hence we kept the data-range here. Further we differentiated the individual sites of each formation for better comprehension and added a porosity versus present day depth plot as requested by reviewer 1.

- L 400: Tests on AMA samples were conducted at effective stresses between 1 and 20 MPa, whereas OPA samples from Lausen were tested under effective stress conditions ranging from 0 to 50 MPa. – Please state explicitly the implications of stress on the differences in measured hydraulic conductivity (K) between in situ and laboratory tests.

We address the effect of stress on K in an individual subchapter (4.3.3). Therefore we deleted the respective sentence to avoid redundancy and confusion.

- L405: The less consolidated Boom Clay – The term "less consolidated" is not appropriate in a geotechnical context. Given that the overconsolidation ratio (OCR) of Boom Clay is 2.4 (Ln141), it should be described as a normally to slightly overconsolidated clay.

You are right and we corrected accordingly.

- L428: OPA data from shallow boreholes in the Swabian Alb (Germany) do not include continuous reports of interval lengths – What does it mean continuous reports? Are the data from boreholes in the Swabian Alb excluded from Figure5?

Indeed, the data from the Swabian Alb could not be included in Fig. 5, as interval lengths for the corresponding tests are not reported in the source literature. We have revised the wording in the manuscript to state this explicitly.

- Figure 5: How many samples from Lausen because only two symbols are visible for Lausen. It would be helpful to add the sample size for each borehole.

We revised the figure to make the individual borehole data more visible.

- L 431: two clear interval length-related trends appear in BO2.0 and BO4.0 – For BO2.0, there is only one data point corresponding to a larger packer interval (y=47 m), while five scatter data points are derived for y=10 m. The two data points with hydraulic conductivity values on the order of 10^{-9} and 10^{-6} m/s, are omitted from the trend without justification, which undermines the reliability of the observed trend. Additional data obtained from longer packer intervals are needed to robustly support this interpretation. For BO4.2, getting a trend based on 3 points is not convincing.

We have revised Fig. 5 to use distinct markers per borehole and added the sample size for each location directly in the legend (e.g., "Lausen (n=8)"), which also clarifies that Lausen is represented by more than the two visually overlapping points. We have added a second panel to Fig. 5, showing present-day depth against hydraulic conductivity for the Lausen and AMA boreholes, with each measurement represented as a vertical segment spanning its packer test interval. This allows depth- and interval-length-related effects to be visually distinguished.

We agree that a trend based on two (BO2.0) or three (BO4.0) data points is not robustly supported. We have removed the claim of a systematic trend for BO2.0 and rephrased the BO4.0 statement to note that the observed decrease in K coincides with greater present-day depth of the corresponding intervals. Thank you very much for these important considerations.

- L 432: in the deeper borehole sections between 45 – 95 m – The word deeper is misleading, as this refers to the length of the packer interval rather than depth. Please rephrase to avoid confusion.

We rephrased the sentence to avoid confusion.

- Fig.6 & line 461: Fig. 6 is derived from $\Delta log10(K) = log10Kinsitu - log10Klab$ (line 328)), with data grouped into 10 m depth bins (L326). – It appears that only data points for which both laboratory and in situ measurements are available within the same depth bin have been selected. I am wondering how many samples are included in the two bins for the To-Do formation. Please report or each formation, the sample size in each depth bin. Also please clarify the data selection in the text. This information is important for assessing the robustness of the statistical analysis.

We confirm that $\Delta\log_{10}K$ was calculated only for depth bins in which both laboratory and in-situ measurements were available, since the difference cannot be computed otherwise.

We agree that reporting the underlying sample sizes is important for assessing robustness. At the original 10 m bin width, many bins (particularly for To-Do, Boom, and AMA) contained only 1–2 measurements per bin. We have therefore revised the bin width to 40 m, which increases the number of measurements per bin while still resolving the depth-dependent pattern shown in Fig. 6. The resulting sample sizes per formation and depth bin are provided in a new supplementary table. We note that, even with the wider bins, To-Do remains represented by a small number of measurements per bin, consistent with the overall limited sample size for this formation discussed elsewhere in the manuscript (see our response to your general comment on sample size, and the revised Limitations section, Sect. 5.4). We have added a corresponding note of caution to the Fig. 6 caption and added the individual sample sizes per bin and formation.

- L446: Lausen borehole at depths >30 m – Where can the reader find the depth range for samples taken from Lausen borehole?

As indicated in our text, we can only report on the data-ranges for laboratory K given by Crisci et al. (2019) as no individual data points are given in the reference. However, a depth range can be clearly stated based on Crisci et a. (2019): The depth of samples on which hydraulic conductivity was measured lay between 22.6 and 70.47 m present-day depth. We refer to this depth range in the manuscript: "For laboratory K, derived from oedometric and constant-head permeameter experiments conducted on samples from depths between ~23 and ~70 m depth, hydraulic conductivities consistently fall…". The depth intervals of in situ tests in the Lausen borehole are now additionally depicted in newly added Fig. 5b, together with the in situ test intervals of the AMA and the corresponding lab data. We adapted the text referring to this figure.

- L453: the interquartile ratios – Where is the value for the interquartile ratios?

and

- L419: interquartile range; L435: dashed vertical lines are IQR – Please clarify what is IQR and how to calculate interquartile ratios.

This was a terminological inconsistency on our part. Thank you very much for pointing out this ambiguity. The interquartile range (IQR) is defined as the difference between the 75th and 25th percentiles (in $\log_{10}K$), i.e., IQR = $\log_{10}$Q75) − $\log_{10}$(Q25) and the dashed vertical lines in Fig. 4 and 5a mark these two percentile values. We have removed the inconsistent term "interquartile ratio" used elsewhere in the manuscript and now refer to the IQR consistently throughout. We have also added the corresponding IQR value (in situ = 0.79 $\log_{10}$ units; lab = 0.84 $\log_{10}$ units) to the text and a short method section explaining the IQR.

Section 4.3.2

- L465: Laboratory measurements for which the flow direction with respect to bedding orientation was reported (n=367) were separated into tests performed parallel (KP; n=135) and perpendicular (KS; n=232) – Please state explicitly the sample size for each clay formation within each category (parallel and perpendicular).
- Each category must have different numbers of samples for each type of clay. According to Table 5, there are 269 laboratory data points for Boom Clay (BC). If BC constitutes the largest dataset in each category, then the calculated geometric mean—where each sample contributes equally—is likely to be dominated by BC.
- Figure 7: 4.3.3 Stress dependency of laboratory data Figure 7 presents the anisotropy ratio for each formation (COX:Kp/Ks= 1/0.8 = 1.25; OPA: 1/0.9 = 1.11; AMA: 1/0.6 = 1.67; BC: 1/0.4 = 2.5), the overall geometric mean anisotropy ratio of approximately 3 seems to be dominated by BC.

Thank you so much for these important observations, which we fully agree with. Boom Clay accounts for 233 of the 363 oriented measurements (~64% of the dataset), and the previously reported geometric means were indeed dominated by this formation, as evidenced by the overall anisotropy ratio (~2.9) closely matching Boom's own ratio (2.5) rather than the substantially lower ratios of OPA (1.11), COx (1.25), and

AMA (1.67). We have revised the text to report formation-specific sample sizes, geometric means, and anisotropy ratios instead of a single pooled value, and now explicitly note that Boom Clay shows the strongest anisotropy among the investigated formations while OPA and COx are close to isotropic. The formation-specific values shown in Fig. 7 (and in newly added table 6) already reflected this pattern; the accompanying text has now been revised to match.

4.3.3 Stress dependency of laboratory data

- The analysis is very vague. OPA: no trend between K and σ' has been identified because only a few data points originate from the same sampling location (lines 481).
- The data supporting the statement that "additional results from the Lausen borehole… show that hydraulic conductivity decreases systematically with increasing effective stress" are not presented in the manuscript.
- AMA: The statement that the decreasing trend "vanishes at confining stresses above ~5 MPa (10–20 MPa range)" (L485) is not clearly supported. From Figure 8a, the decreasing trend appears to remain valid at least within the 0–10 MPa range,

We agree with your comments and have revised the whole section intensively in the newly submitted manuscript. We now place more emphasis on the differing test types and stress conditions applied across formations also acknowledging that terminology was previously used inconsistently. We now support the discussion of the Lausen borehole with the quantitative power-law relationship reported by Crisci et al. (2019), rather than an unsupported qualitative statement. For AMA, we have revised the previous claim that the stress-dependent trend "vanishes above ~5 MPa," acknowledging that it persists up to ~10 MPa. Figure 8 has been revised accordingly.

Section 4.3.4:

- L506: The compiled data shows a median KP/KS ratio of approximately 3, in good agreement with this range, indicating that anisotropy introduces only a minor influence on the overall comparability of hydraulic conductivity values. – This sentence is hard to understand. Anisotropy for the four formations, varying between 1.11-2.5, which is not far from the range of 1.5-5 (line 506). The poenitentia influences of this level of anisotropy on the "overall comparability of hydraulic conductivity values" needs to be clarified.
- Regarding KP/KS ratio of approximately 3, please see comments for line 465.
- L508: Hydraulic conductivity from in situ tests, such as packer experiments, mostly conducted in vertical boreholes, predominantly reflect KP(perpendicular) – In an anisotropic formation, the measured K likely reflects a combination of effect of both Ks and Kp.

We agree and revised the respective section following your comments. In the revised manuscript we now acknowledge for the formation specific bedding anisotropy values and clarified their significance in the broader context.

Section 4.5:

- L552: Across the investigated formations, mean porosity values show a positive correlation with laboratory-measured hydraulic conductivity (Fig. 11a) – The four data points (excluding the Boom Clay point highlighted in yellow) are scattered. As a result, the inferred positive or negative correlations between porosity or bulk density and hydraulic conductivity (K) are not convincingly supported by the data.

We note that a perfect correlation cannot be expected given the multiple additional sources of variability discussed in the preceding subsections (measurement scale, bedding anisotropy, effective stress conditions), in addition to the fact that K is not purely controlled by porosity but also by factors such as pore-size distribution, mineralogical composition, mineral and grain size distribution. We have revised the text to make this connection explicit rather than presenting the porosity-K relationship in isolation.

Section 4.6 and Fig 12:

- Given that Figure 12 is a key figure supporting the analysis and main conclusions, please include a detailed table reporting sample sizes used for each borehole, and state explicitly how the values used for graphing Figure 12 were derived. In the current manuscript, the data are scattered in different places, which makes it difficult for readers to locate them efficiently.

Please note that such a table is already given as Table 5 in the manuscript. However, we agree, that additional information with further differentiation between formation specific sites/borehole locations benefits the comprehension at this point. We added a table (Table 8 in new manuscript version) for a better overview.

- Boom Clay(BC): The maximum burial depth of BC shown in Figure 12 is 200 m, and the porosity is 0.4. However, the K data of Boom Clay, which composed data from 5 boreholes and cover a (present-day) depth range between 59-366 m (line 367 and Table 4). Over-consolidated ratio (OCR) for Boom Clay is 2.4 (Line 241). Based on these information, It is not clear what the procedure used by the authors to process raw data to get a maximum burial depth of 200 m, and a porosity of 0.4 for Boom Clay.

Indeed, the Boom data stems from several boreholes. Here we report on the data-range (whiskers) and the geom.mean (marker) of max. burial depth and K. We clarified this in the figure caption.

- COX: The maximum burial depth is listed in Table 4 as 850 m, with Blaise et al. (2014) cited as the reference. However, the last paragraph on p. 83 of Blaise et al. (2014) states that "the maximum burial depth of the Callovian–Oxfordian claystones in the Andra URL area was estimated to be around 500–700 m ……" Please clarify the origin of the value of 850 m reported in Table 4, or how it was derived.

Thank you for pointing this out. Please note that Blaise et al. (2014) here refer to another, older reference (Landais and Elie, 1997). In their work they use different, independent geothermometers to reconstruct the low temperature burial in the Paris Basin. For COx their results indicate max. burial temperatures of 50 (±5) °C. Following their thermal history diagram (Fig. 10 B) and the CCC (OECD and NEA, 2022) a maximum burial depth at about 850 m is estimated (in dependence on considered thermal flux and conductivity). In addition, Bock et al. (2010) state a max. burial depth of ~820 m. We added the references in our manuscript for clarification.

- The OPA data are derived from multiple regions and boreholes, including: Mont Terri(MT), Rinken (Jira Ost), Weiach (Nördlich Lägern), Benken(Zürich Nordost)
    - Additional boreholes providing in situ data (as shown in Figure 5)
    - Shallow OPA data from Swabian Alb(excluded from Figure 12 due to lack of burial depth information)
    - Shallow OPA data from Lausen (excluded from Figure 12 due to lack of burial depth information)

Please provide a summarized table for each borehole/region, clearly reporting the sample sizes for K_lab and K_in situ for each location.

Done. Please see our answer to your first comment on section 4.6.

Additional comments are provided below for L588.

- L569: depth between 1400 m and 3300 m: The maximum burial depth in Table 3 for AMA is 1300-3300m. Please make the values consistent.

Done. Thanks for making us aware of that.

- L571: The BCF …with burial depths of up to 4,500 m. - The max burial depth of BCF in Figure 12 is 4000m. Please clarify where 4000 comes from.

The CCC gives a range of maximum burial depth for the BCF between 3500 and 4500. In Figure 12 the uncertainty in max. burial depth (if reported) is indicated by vertical whiskers. The mean depth was then chosen as the marker position. We clarified that in the figure caption.

- L588: For formations that experienced maximum burial depths < 2,000 m, laboratory-derived and in situ hydraulic conductivity show a close agreement, indicating consistent matrix and bulk hydraulic properties at depth. In contrast, a divergence between laboratory- and field-scale hydraulic conductivity values emerges in formations with burial depths ≥2,000 m, notably in the To-Do and BCF.

For data with maximum burial depths < 2000 m in Figure 12a, agreement between K_lab and K_in situ is indeed observed for Boom Clay and MT, whereas the remaining regions/boreholes (e.g. JO, NL and COx) exhibit obvious discrepancies between K_lab and K_in situ. Moreover, data from Lausen is excluded from Figure 12 due to the lack of burial depth information. If the Lausen dataset, characterized by K values spanning approximately 7 orders of magnitude (purple triangles near the surface in figure 4b) , were included in Figure12a, the conclusion that "for depths < 2,000 m, K_lab and K_in situ show a close agreement" would no longer be valid. In Figure 12a , five boreholes/regions are included for depths greater than 2000 m, namely To-Do, BO3-BO5, and BCF. The To-Do formation includes 8 laboratory and 6 in situ measurements (Table 5). BDF has only 3 laboratory results and no specific in situ data. BO3, BO4, and BO5 include only laboratory tests.

Thank you for this detailed and important comment. Indeed, MT and Boom show almost no divergence between lab and in situ data. The divergences observed in the other formations <2000 m max burial are all <1 order of magnitude and therefore lie well within the prior defined natural variability range. The divergence might occur as a result of the prior discussed points (anisotropy, test conditions etc.) as well as general dataset heterogeneity. Considering these points, only BCF and To-Do show significant discrepancy between lab and field measurements that we attribute to the points addressed in the discussion. We revised the section accordingly.

- Please provide a summarized table for each borehole, clearly reporting the sample sizes for each borehole regarding K_lab and K_in situ .

Please see comments above.

- In total, only 17 K measurements (8+6+3) for K_lab and K_in situ coming from To-Do and BCF. Such a limited samples raises the question of whether it is sufficient to support the conclusion that K_lab and K_in situ diverge for max burial depths greater than 2000 m.

We agree, that the data in this max. burial depth range is very limited. Please see our comments given further up regarding the newly established discussion chapter to address this and other points.

Discussion

- L608: The compiled porosity data clearly follows the expected burial trend, with progressive pore space reduction accompanying mechanical compaction and cementation. The hydraulic conductivity shows a broadly similar pattern, with a marked decrease between the Boom Clay and COx that faced maximum burial of around 850 m depth. – The conclusion of a "similar pattern" is based on only two data points for BC and COX (see Figure 12a), which is insufficient to support such a claim. Additional data points are needed to substantiate this observation.

Indeed, we revised the wording to acknowledge this fact.

- L611: For mechanically compacted and lithified formations (e.g., COx, OPA), the vast majority of matrix hydraulic conductivity determined from laboratory tests fall within a range of 10-14 to 10-12 m/s, independent of site, formation, or test conditions. This range, previously defined as the natural variability envelope, likewise characterizes the hydraulic behavior at the rock mass scale in the absence of near surface perturbations such as weathering or decompaction. – Among the six types of host rock considered in this manuscript, please specify which formations are classified as "mechanically compacted and lithified formations" and definition of "mechanically compacted and lithified formations".

Thank you very much for pointing this out. The diagenetically lithified formations comprise Cox, OPA, AMA, To-Do, and Boda. However, here we make a distinction with respect to their max. burial history. In the respective section that your comment is addressing we refer to those formations where K lab and K in situ converge (with respect to the natural variability), namely OPA and Cox. We revised the wording accordingly.

- Do the authors intent to classify the data in Figure 12a into 3 categories: 1) BC and Cox for depth <850m, 2) OPA & BO1 for depth between 850-2000m, and 3) AMA, BCF and TO-Do for >2000m? if so, this categorization should be explicitly stated in the text. Please clearly define the depth ranges for each category and specify which boreholes/regions are included in each group.

Yes, a categorization is indeed intended. However, it foresees rather 1) BC as stiff clay rather than claystone with max. burial depth <600 m, 2) Formations indurated and lithified in which self-sealing is effiecient (mainly due to clay mineral swelling), and 3) Formations >2000 m burial depth in which self-sealing efficiency is significantly reduced, mainly due to the gradual loss of swelling potential. This categorization is stated in the abstract as well as in the discussion and conclusion.

- This manuscript distinguish between small scale (matrix) and large scale (rock-mass) (line 32) hydraulic properties. Should the lab tests referred to here represent K at the small scale(matrix)?

Indeed, that is the intention.

- For the category with max burial depths of 850-2000m, please discuss the possible causes of the differences between K_lab and K_in situ for Cox, JO and NL.

Please see our earlier answer to your comment of line 588.

- Line 616-626: The conclusion that K_in situ >> K_lab for the category with maximum burial depths >2000 m is based on a limited sample size from the To-Do and BCF formations, and therefore requires additional data to be adequately supported.

We fully agree with you. Additional data would be important to fully support the conclusions drawn from the compiled data. Same applies to the shallowly buried stiff clay formations such as Boom Clay. Both end-members (shallow and very deep maximum burial depth) play a crucial role in understanding the burial related changes in hydraulic properties. Potential formations that could be considered in this approach are the Mercia Mudstone from the UK that face burial up to 2,000 – 2,500 m or Ypresian Clays (same locations as Boom Clay). However, to our knowledge only little data for hydraulic characteristics, especially in situ data, is currently available in the literature that would allow for scale comparison of the hydraulic properties. We now clearly address these limitations in subchapter 5.4.

5.1 self-sealing characteristics of investigated formations

- L661: In the moderately indurated formations (Clay, COx, OPA), Please clarify what is meant by “Clay” , Boom Clay?

This is a typo. “Clay” was deleted.

- L664: indicating efficient sealing of natural or excavation-induced discontinuities below approximately 250 m depth.. – Please clarify what is meant by “below approximately 250 m depth” ,

We rephrased the section to clarify this.

- L673 Toarcian-Domerian argillites exhibit generally overlapping in situ and laboratory hydraulic conductivity – For the To-do data in Figure 12 a : K_lab and K_in situ differ by at least one order of magnitude. Please clarify the definition of “overlapping” in this context. It should also be noted that the To-Do formation includes only 8 laboratory and 6 in situ measurements (Table 5), which may limit the robustness of any inferred overlap between the two datasets.

What was meant here, is that the in situ and lab K range overlaps. However, we agree, that this statement is unclear and we revised the sentence accordingly.

- L684: In the case of the BCF, the highly indurated nature and general depletion of swellable clay minerals do not favour rapid self-sealing. – Please clarify the fraction of swellable clay minerals for each formation considered in this manuscript, since it is the controlling factor of self-sealing property.

and

- Line 723 suggests that smectite may be the primary swellable clay mineral. In this context, it would be more informative to report the smectite fraction for each formation.

We added information about the smectite fraction to the respective sections and for the individual formations. Please note, that for AMA no discrete clay-mineralogical information could be found. However, currently a publication by Seeman et al. is in preparation that addresses the specific mineralogical composition of the AMA in detail as part of the MATURITY research project (see Burchartz et al., 2026).

5.2 The role of maximum burial depths

- L700: An initial porosity loss exceeding 20 % is observed between the Boom Clay and the COx, corresponding to a differential burial of approximately 600 m. – For the category with max burial depth <600m, the porosity-maxi burial depth relationship is derived from BC and OPx,. Additional data points within this depth range are required to robustly support the proposed relationship.

As this statement only refers to the two mentioned formations, we do not see the need for further data points. However, we fully agree, that it would be beneficial to close this gap of max. burial depth with additional data. Unfortunately, to our knowledge there is currently no data publically available that could serve as suitable input. We discuss the general data limitations in the newly established chapter 5.4.

- L706: Neuzil, (1994) demonstrated the existence of a log-linear relationship between porosity and hydraulic conductivity across a broad range of argillaceous formations, a trend also reflected in the data compiled for this study. – Please add a porosity-K graph based on figure 12. Please include also the log-linear relationship proposed by Neuzil, (1994) in the figure for comparison.

Our wording was misleading. Neuzil (19914) states: "Figure 1 suggests that a log-linear relation between permeability and porosity exists over an exceptionally wide range of consolidation states; materials ranging from recently deposited marine clays to mildly metamorphosed argillite are represented. Excluding regions 11 and 12, the data fall in a band approximately three orders of magnitude wide that spans porosities exceeding 0.8 to less than 0.1. Permeability decreases by approximately an order of magnitude with each 0.13 decrease in porosity and ranges over eight orders of magnitude." We have adapted our wording accordingly.

- L708: hydraulic conductivity would be expected to decrease further with continued pore-space reduction beyond approximately 850 m burial. This is generally consistent across the investigated formations and observational scales (laboratory versus in situ) up to maximum burial depths of around 2,000 m – An explicit porosity – K curve based on figure 12 is necessary to prove this.

We included a new figure (Figure 13) in the respective section to show the correlation between porosity and hydraulic conductivity. Within those formations where different sites could be considered, namely OPA and AMA the correlation is mostly visible. For OPA a regression analysis gave $R2=0.807$. In AMA the correlation is less pronounced. We revised the text section accordingly.

- Figure 11 present a porosity-K graph for BC, Cox, and OPA, all with maximum burial depths < 2,000 m. The OPA data appears to deviate from the overall trend of decreasing K with decreasing porosity . This variability should be discussed in more detail, as it may indicate that the formation-specific effects play a more significant role than a consistent global trend.

Yes, we fully agree with your observation. However, to us COx is the "deviating" formation with comparably low K for the sampled porosity-range. This might be a result of early diagenetic precipitated carbonates that potentially results in the clogging of pore throats, reducing water accessible flow paths and associated K: see i.e.

- **Esteban, L., Géraud, Y., and Bouchez, J. L.: Pore network connectivity anisotropy in Jurassic argillite specimens from eastern Paris Basin (France), Physics and Chemistry of the Earth, Parts A/B/C, 32, 161–169, https://doi.org/10.1016/j.pce.2005.11.001, 2007.**
- **Menaceur, H., Delage, P., Tang, A.-M., and Conil, N.: The thermo-mechanical behaviour of the Callovo-Oxfordian claystone, International Journal of Rock Mechanics and Mining Sciences, 78, 290–303, https://doi.org/10.1016/j.ijrmms.2015.07.002, 2015.**
- **Robinet, J. C., Sardini, P., Siitari-Kauppi, M., Prêt, D., and Yven, B.: Upscaling the porosity of the Callovo-Oxfordian mudstone from the pore scale to the formation scale; insights from the 3H-PMMA autoradiography technique and SEM BSE imaging, Sedimentary Geology, 321, 1–10, https://doi.org/10.1016/j.sedgeo.2015.02.007, 2015.**

We acknowledge the hydraulic variability as a result of factors such as mineralogical composition, grain- and pore size-distribution across different formations and discussed this in more detail in the revised section.

- L715: At maximum burial depths exceeding 2,000 m, a divergence emerges between laboratory-derived matrix hydraulic conductivity ….This paragraph summarizes plausible explanations for the scale effect on hydraulic conductivity (K) in formations with maximum burial depths exceeding 2000 m. However, the discussion is primarily based on 17 hydraulic conductivity measurements from To-Do and BCF formations. Additional datasets are needed to robustly support this observation.

We fully agree with you. Complementing data would be an important step in robustly supporting these observations. Unfortunately, the current data-base does not provide such data. See our earlier comment on lines 616 – 626.

5.2 The role of present-day depth

- L752: Mazurek et al. (2023) demonstrated that, aside from hydraulically active fractures, diffusion remains the dominant transport mechanism even in the weathered Opalinus Clay encountered in the Lausen borehole. – This sentence is oversimplified and it needs some context. It is suggested to improve it as “Aside from hydraulically active fractures, diffusive transport still remains the dominant transport mechanism governing solute exchange between fractures and matrix, even in the weathered Opalinus Clay encountered in the Lausen borehole.”

We agree. Thank you very much for this much improved formulation.

- L758: However, in situ data from AMA borehole BO4 reveal a depth-dependent trend in hydraulic conductivity comparable to that observed in the OPA. - Where can the reader find the depth dependence of the in situ data for BO4? Please indicate clearly in the manuscript or figures where this relationship is presented.

Done. The K-Depth relationship observed on in situ data in BO4.0 can now be seen clearer in revised Fig. 4 and additionally in Fig. 6.