Referee # 6 Comments (Report 1):

General Comment

Comment:

Technical review of the manuscript titled "Novel insights into deep groundwater exploration by geophysical estimation of hard rock permeability"

The paper presents a significant advancement in hydro-geophysics by demonstrating the use of controlled-source audio-frequency magnetotellurics (CSAMT) to estimate two- and three-dimensional permeability (K) distributions at depths exceeding 1 km. While previous geophysical approaches like vertical electrical sounding (VES) and electrical resistivity tomography (ERT) are generally restricted to shallow depths less than 500m, this study successfully bridges the gap between regional mapping and site-specific, high-resolution modeling for deep aquifers in crystalline and sedimentary terrains. So, the current study is scientifically rigorous, utilizing 116 core samples across three lithologies (granite, hornstone, and sandstone) to establish a strong empirical correlation between resistivity and permeability $\{R2 = 0.96\}$. Whole manuscript was written in a highly technical and professional academic style. Furthermore, the authors successfully argue that CSAMT bridges the resolution depth gap that MT and TDEM cannot fill in resistive hard rock settings. However, they rightly emphasize that the derived empirical constants are site-specific and must be recalibrated for new geological environments.

Despite these strengths mentioned above, I have identified a few areas, in my point of view, where the manuscript would benefit from clarification and refinement.

Response:

Dear Anonymous Referee,

We sincerely thank you for your comprehensive and technically rigorous review of our manuscript. We greatly appreciate your recognition of the scientific contribution of this study, particularly the demonstration of controlled-source audio-frequency magnetotellurics (CSAMT) for estimating two- and three-dimensional permeability distributions at depths exceeding 1 km.

Your positive evaluation of the empirical calibration based on 116 core samples across granite, hornstone, and sandstone lithologies, and the resulting strong resistivity–permeability correlation ($R^2 = 0.96$), is highly encouraging. We are also grateful for your acknowledgment that the study effectively bridges the resolution–depth gap between shallow methods (VES, ERT) and deeper regional approaches (MT, TDEM) in resistive hard rock settings, while appropriately emphasizing the site-specific nature of the derived empirical constants.

Your detailed comments regarding extrapolation uncertainty, scale effects, salinity sensitivity, and aquifer storage characterization have significantly strengthened the manuscript. The revisions prompted by your suggestions have improved the transparency of the uncertainty framework, enhanced the theoretical depth of the discussion, and clarified the limitations and applicability of the proposed workflow.

All revisions are clearly indicated in the tracked-changes version of the manuscript. Detailed responses to each specific comment are provided below.

Specific Comments

Comment 1 (ETQ-1):

Although the authors correctly identify this as a limitation and use a probabilistic framework to show that uncertainty increases significantly beyond the calibration limit. Authors should try to address the extrapolation uncertainty associated with applying a relationship calibrated at 0–200 meters to depths reaches 1.3 kilometer.

Response 1:

This concern has been explicitly addressed in the revised Section 4.7 of the Discussion.

In the revised text, we clearly state that application of the resistivity–permeability relationship beyond 200 m constitutes extrapolation beyond the calibrated interval. A probabilistic permeability–depth framework has been implemented to propagate regression uncertainty and progressively widen prediction intervals below the calibration limit.

We further clarify that:

- Deep predictions remain conditioned by observed CSAMT resistivity trends.
- The monotonic physical relationship between resistivity and permeability is preserved.
- However, processes not represented in the shallow dataset, such as stress-dependent fracture closure, evolving fracture connectivity, and mechanical anisotropy, may influence permeability at depth.

We explicitly state that while the probabilistic framework constrains extrapolation risk, it does not eliminate structural uncertainty, and that deep borehole validation is required to confirm predictions below the calibrated range.

Track Changes Copy: Page 41-42, Line 681-699

Clean Copy: Page 41-42, Line 680-698

"4.7 Limited and shallow borehole calibration

A key limitation of this study is the restricted depth range of the calibration dataset. The empirical resistivity–permeability relationship (Eq. 3) was derived from 116 core measurements between 0 and 200 m depth. Application of this relationship to depths approaching 1300 m therefore represents extrapolation beyond the calibrated interval.

To explicitly address this uncertainty, a probabilistic permeability–depth framework was implemented (Fig. 13). Rather than extending Eq. 3 deterministically, prediction uncertainty was propagated from the regression and progressively inflated beyond the 200 m calibration limit. Within the calibrated interval, permeability estimates remain relatively well constrained. Below 200 m, however, the 95% prediction intervals widen substantially, quantifying decreasing confidence with increasing depth.

Importantly, deep predictions remain conditioned by observed CSAMT resistivity trends and preserve the physically consistent monotonic decrease in permeability with increasing resistivity. Nevertheless, processes not captured by the shallow dataset, such as stress-dependent fracture closure, evolving fracture connectivity, and mechanical anisotropy, may modify permeability behavior at depth. While the probabilistic framework constrains extrapolation risk, it does not eliminate structural uncertainty. Deep borehole testing and stress-dependent hydraulic measurements are therefore required to validate permeability predictions below the calibration range."

Comment 2 (ETQ-2):

There is an inherent scale difference between centimeter-scale core samples and 50-meter scale of CSAMT blocks. This inherent scale explains some of the scatters in the K-ρ dataset, as a core might miss fracture corridors captured by the geophysical signal. The authors should consider expanding their discussion in Section 4.9 (Scale effects in permeability estimation) to more explicitly address the following points:

✓ Quantifying the Representative Elementary Volume (REV), The authors should emphasize that because core samples (50 mm × 100 mm cylinders) primarily capture intrinsic rock matrix properties, they may inherently underestimate the flow capacity in crystalline terrains where fractures dominate. Conversely, CSAMT inversions represent a bulk effective property that integrates both matrix and fracture contributions over a much larger volume (about 2.5 Km^2). Whereas, I believe that, explicitly stating that the core data may not reach the REV of a fractured rock mass would add theoretical depth to the explanation for the dispersion of the dataset.

✓ The impact on statistical anomalies; I suggested relating the scatters in the K-ρ plot (Fig. 5) to different specific geological scenarios. to help readers understand that the "discrepancies" are actually physically meaningful reflections of geological heterogeneity rather than mere measurement error.

✓ Bridge the Scale Gap by, for example, recommending integrating intermediate-scale data, such as packer tests, in future work to reconcile matrix-scale measurements with field scale connectivity.

Response 2:

This issue has been comprehensively addressed in the revised Section 4.9 (Scale effects in permeability estimation).

The revised section now:

- Explicitly introduces the concept of the Representative Elementary Volume (REV).
- Clarifies that individual core samples (50 mm × 100 mm) may not reach the REV of fractured crystalline rock masses.
- Explains that core measurements primarily capture matrix-scale permeability, whereas CSAMT inversions represent bulk effective properties integrating matrix and fracture contributions over ~50 × 50 m blocks.

We now explicitly interpret the dispersion in the k–ρ dataset (Fig. 5) as a physically meaningful expression of geological heterogeneity rather than measurement error. Specific geological scenarios are discussed, including:

- Fracture corridors intersecting CSAMT blocks but not cores
- Localized fractures intersecting cores
- Variability in cementation and fracture density
- Stress-controlled aperture closure

To bridge the scale gap, we explicitly recommend integration of intermediate-scale hydraulic constraints, including:

- Packer testing
- Interval hydraulic testing
- Borehole geophysics
- Pumping tests

These additions strengthen the theoretical basis of the scale discussion and clarify the multi-scale nature of permeability estimation.

Track Changes Copy: Page 42-43, Line 715-739

Clean Copy: Page 42-43, Line 714-738

"4.9 Scale effects in permeability estimation

In addition to depth-related extrapolation, permeability estimation in this study is influenced by scale transition between centimeter-scale core measurements and tens-of-meter-scale CSAMT inversions. Core plugs (50 mm × 100 mm) primarily capture intrinsic matrix permeability and fractures intersecting the limited sample volume. In fractured crystalline systems, however, hydraulic flow is governed by connected fracture networks whose representative elementary volume (REV) may exceed the dimensions of a core specimen. Individual cores may therefore not reach the REV of the fractured rock mass, leading to systematic underestimation of bulk hydraulic conductivity.

By contrast, CSAMT inversions resolve an effective bulk resistivity over ~50 × 50 m blocks, integrating matrix and fracture contributions across a much larger volume. The empirical k–ρ relationship thus links matrix-scale permeability measurements with block-scale electrical responses. The dispersion observed in Fig. 5 is therefore not merely statistical noise but a reflection of geological heterogeneity and scale-dependent flow processes.

For example, low k values from intact granite cores may correspond to CSAMT blocks intersecting fracture corridors that enhance effective permeability at the field scale. Conversely, cores intersecting localized fractures may yield elevated k relative to the surrounding bulk medium. Variability in cementation, fracture density, and stress-controlled aperture closure further contributes to scatter.

Bridging this scale gap requires intermediate-scale hydraulic constraints. Integration of packer testing, interval hydraulic testing, borehole geophysics, and pumping tests would help reconcile matrix-scale measurements with field-scale connectivity and better constrain the effective REV of the fractured system. Such multi-scale calibration would reduce uncertainty in empirical relationships and improve permeability modeling in structurally complex crystalline terrains.”

Comment 3 (ETQ-3):

The model assumes fresh groundwater at depth based on regional data. A sensitivity analysis reveals that if fluid resistivity were halved due to salinity, the inferred permeability could increase by a factor of 2 to 18 depending on the formation resistivity, which is a critical consideration for deep-seated aquifers. In addition to the authors, in the current study, focuses on K without mentioning the aquifer storage parameters like specific yield, which are essential for a complete groundwater resource assessment. So, I suggest that the discussion regarding the sensitivity of the model to groundwater salinity and the limitations in characterizing aquifer storage should be more explicitly highlighted to ensure the results are not over generalized. Like the future integration of Nuclear Magnetic Resonance (NMR) or other logging tools to better characterize storage capacity alongside permeability.

Response 3:

This concern has been explicitly addressed in revised Sections 4.11 and 4.13.

Section 4.11 – Salinity Sensitivity

We clarify that the assumption of predominantly fresh groundwater conditions is an important model constraint. While regional hydrochemical data support fresh conditions down to ~1000 m, no direct fluid data exist below ~1 km.

A detailed sensitivity analysis now quantifies that halving fluid resistivity (ρf) would increase inferred permeability by approximately:

- Factor ~2 at 1000 Ωm
- Factor ~7 at 2000 Ωm
- Up to factor ~18 at 3000 Ωm

We explicitly state that:

- The model should not be generalized to saline deep aquifers without recalibration.
- Deep hydrochemical sampling and downhole conductivity logging are required to constrain salinity-related uncertainty.

Section 4.13 – Storage Characterization

We now clearly distinguish between permeability and aquifer storage parameters (specific yield, specific storage, storativity).

The manuscript explicitly states that:

- The present study characterizes hydraulic flow potential rather than total groundwater resource capacity.
- High permeability does not necessarily imply high storage.
- The absence of deep pumping tests and porosity logs limits quantitative storage estimation.

To avoid overgeneralization, we recommend integration of:

- Nuclear Magnetic Resonance (NMR) logging
- Borehole porosity tools
- Interval hydraulic testing
- Aquifer-scale pumping tests

These additions strengthen the hydrogeological completeness of the discussion and ensure the results are properly contextualized.

Track Changes Copy: Page 44-47, Line 759-781 and 808-828

Clean Copy: Page 44-47, Line 758-780 and 807-827

"4.11 Salinity effects and uncertainty in deep fluid properties

The influence of factors beyond lithology, particularly groundwater salinity, on CSAMT-derived resistivity warrants careful consideration. Electrical resistivity is sensitive to porosity, fracture density, mineral alteration, fluid saturation, and fluid resistivity (ρf). In this study, the empirical k–ρ relationship was calibrated using core samples from 0–200 m across six boreholes under predominantly fresh groundwater conditions. Regional hydrochemical data from the Geological Survey of China (800–1000 m depth) consistently indicate low salinity, supporting the assumption of fresh groundwater within the investigated interval. However, no direct fluid data are available below ~1 km, and the assumption of fresh conditions at greater depths represents an important model constraint.

To evaluate this uncertainty, a sensitivity analysis was performed. If fluid resistivity were reduced by 50% due to increased salinity, formation resistivity would decrease proportionally, resulting in higher inferred permeability values when substituted into Eq. (3). Because permeability increases exponentially with decreasing resistivity in the fitted model, this effect is nonlinear and resistivity-dependent. Specifically, halving ρf increases inferred k by approximately a factor of 2 at 1000 Ωm, 7 at 2000 Ωm, and up to 18 at 3000 Ωm. These results indicate that salinity effects are modest in low-resistivity formations but may significantly influence permeability estimates in highly resistive deep crystalline units.

Accordingly, the permeability model should be interpreted as valid under the assumption of predominantly fresh groundwater conditions and should not be generalized to saline deep aquifers without recalibration. Future investigations should incorporate deep borehole fluid sampling, downhole conductivity logging, and hydrochemical analyses to directly constrain fluid resistivity and reduce salinity-related uncertainty in deep permeability predictions."

"4.13 Limitations of storage characterization

A complete aquifer assessment requires evaluation of both permeability and storage parameters, including specific yield, specific storage, and storativity. This study primarily focuses on delineating spatial variations in permeability (*k*), referred to here as hydraulic flow potential, using CSAMT-derived resistivity calibrated with borehole data. While this approach provides robust insights into transmissivity patterns and fracture-controlled flow pathways, it does not directly quantify aquifer storage capacity.

Permeability and storage represent distinct hydrogeological properties. High *k* zones do not necessarily imply high storage if fracture porosity is limited, and conversely, formations with significant porosity may exhibit substantial storage despite moderate permeability. Due to the absence of deep pumping tests, drawdown analyses, and detailed porosity logs, storage parameters could not be independently constrained. As such, the present results characterize relative transmissivity and hydraulic connectivity rather than total extractable groundwater volume.

To avoid overgeneralization, the model outputs should therefore be interpreted as indicators of flow potential rather than comprehensive groundwater resource capacity. Future studies should integrate Nuclear Magnetic Resonance (NMR) logging, borehole geophysical porosity tools, interval hydraulic testing, and aquifer-scale pumping tests to better constrain storage properties alongside permeability. Such multi-parameter characterization would enable more rigorous evaluation of sustainable yield and groundwater resource potential in fractured hard rock systems."

Final Comment:

Finally, the paper is scientifically sound and provides a valuable workflow for deep groundwater assessment. So, My Final Decision is Accept with Minor Revisions Thank you for the opportunity to review this impactful work.

Response:

We sincerely thank you for your positive assessment and for recognizing the scientific rigor and practical contribution of this work. We are especially grateful for your acknowledgment that the proposed workflow provides a valuable framework for deep aquifer characterization in complex hard rock terrains.

Your constructive suggestions have substantially improved the clarity, robustness, and transparency of the uncertainty analysis and methodological discussion.

We deeply appreciate the time and expertise you dedicated to reviewing our manuscript and thank you again for recommending it for publication with minor revisions.

Referee # 7 Comments (Report 2):

I found this work remarkable and challenging, as it explores the use of the CSAMT technique to estimate permeability at great depths, and thus determine which rocks are suitable for high-potential aquifers. The authors have properly described the workflow required to achieve this type of estimation. I liked the comparative analysis in the introduction of the different geophysical techniques for this specific purpose, and their respective capabilities and limitations were clearly referenced. Furthermore, every step is justified and can be perfectly reproduced by other practitioners interested in implementing this proposed methodology. In my opinion, this manuscript fits the scope of this journal because it covers a relevant methodology that is not commonly used. I appreciated the authors' justification or listing of the potential limitations of their methodology and their explanation of how to address these limitations using probabilistic modelling. However, they propose future work involving deep boreholes (>500 m) for sampling and pumping tests to validate the extrapolated k-values. The manuscript was easy to read, well supported by tables and figures, and the references were appropriate and up to date. I believe that scientists working in aquifer characterisation at great depths, geothermal studies, etc. will be pleased to read this study. As the authors have addressed all the comments and suggestions from previous reviewers, I have decided to publish the manuscript in its current form.

Response:

We sincerely thank the Anonymous Referee for the thorough and highly positive evaluation of our manuscript. We greatly appreciate the recognition of the methodological rigor, reproducibility, and clarity of the workflow presented in this study. We are particularly grateful for the acknowledgment of the comparative analysis of geophysical techniques, the transparent discussion of limitations, and the integration of probabilistic modeling to address extrapolation uncertainty.

We also appreciate the referee’s constructive reflection regarding the importance of future deep borehole investigations (>500 m) to validate extrapolated permeability estimates. As highlighted in the revised Discussion (Sections 4.7, 4.11, and 4.16), we have explicitly clarified the depth-related uncertainties and the need for deeper hydraulic and hydrochemical validation to strengthen long-term confidence in deep aquifer characterization.

We are thankful for the referee's recommendation for publication and for recognizing the potential relevance of this methodology for deep aquifer exploration and geothermal studies.

Referee # 8 Comments (Report 3):

General Comment

Comment:

Dear Authors, I've been asked to review manuscript 'Novel insights into deep groundwater exploration by geophysical estimation of hard rock permeability' in view of the interactive public discussion. The paper is a valid contribution to the field of aquifer permeability estimation using a methodology that involves an applied geophysical technique, controlled source audio frequency magnetotellurics (CSAMT), to model deep subsurface electrical resistivity, and correlation with aquifer permeability using calibration in boreholes, offering a methodological framework that can be adapted to other sites. This is a common approach in applied geophysics: using geophysical properties as proxy for the estimation of another property of interest. However, it is important to keep working on the establishment of common methodologies for the approach, and common prediction ability estimations. In that regard, this paper is helpful in its uniqueness and capacity of being applied on similar conditions. However, I think it could be improved in the model prediction ability estimation, and on some details. Therefore, I recommend it for publication with revisions that are the following:

Response:

Dear Anonymous Referee,

We sincerely thank you for your thoughtful and constructive evaluation of our manuscript. We greatly appreciate your recognition of the study's contribution to deep aquifer permeability estimation using CSAMT and borehole calibration, as well as your acknowledgment of its potential applicability to similar geological settings.

We fully agree with your observation regarding the importance of establishing robust and standardized methodologies for prediction capability assessment in hydrogeophysical studies. In response to your valuable suggestion, we have strengthened the manuscript by incorporating a formal model validation framework, including leave-one-out cross-validation (LOOCV) and leave-one-well-out cross-validation, along with quantitative RMSE evaluation of predictive performance. These additions provide a clearer assessment of the model's prediction ability within the calibrated permeability range and enhance methodological transparency and reproducibility.

In addition, we have carefully addressed the specific technical details noted in your review to improve clarity, consistency, and figure presentation throughout the manuscript. All revisions are clearly indicated in the tracked-changes version.

We are grateful for your insightful comments, which have helped us improve the rigor and overall quality of the manuscript.

Specific Comments

Comment 1:

Change term 'depth penetration' to 'penetration depth'.

Response 1:

The term "depth penetration" has been replaced with "penetration depth" throughout the manuscript.

Track Changes Copy: Page 5, 6, 38, 49; Line 109, 116, 128, 625, 876

Clean Copy: Page 5, 6, 38, 49; Line 109, 116, 128, 624, 853

Comment 2:

On Figure 3, Figure 5, Figure 6, and Figure 7:

- rotate orientation of labels on x-axis, so the text becomes horizontal;
- rotate orientation of labels on y-axis, so the text becomes horizontal (also do it on Figure 4);
- rotate orientation of axis title on y-axis, so the text readable from bottom to top (also do it on Figure 4);

Response 2:

The orientation of axis titles and tick labels on both x- and y-axes has been adjusted accordingly in Figures 3–7 to improve readability.

Track Changes Copy: Page 16, 20, 22, 24, 25; Line 298, 379, 401, 431, 434

Clean Copy: Page 16, 20, 22, 24, 25; Line 297, 378, 400, 430, 433

Comment 3:

Figure 3, Figure 6, Figure 7, Figure 8, Figure 9, Figure 10, Figure 11, and Figure 12, rotate orientation of labels on colour scale, so the text becomes horizontal;

Response 3:

The orientation of the colour scale labels has been revised in Figures 3 and 6–12. All colour bar labels are now horizontal to enhance clarity and consistency.

Track Changes Copy: Page 16, 24, 25, 27, 28, 29, 31, 33; Line 298, 431, 434, 463, 466, 483, 500, 526

Clean Copy: Page 16, 24, 25, 27, 28, 29, 31, 33; Line 297, 430, 433, 462, 465, 482, 499, 525

Comment 4:

On Section 3.5, agreement is measured using a percentage match equation for each point in space, as part of the validation of the obtained empirical model. Since the empirical model was obtained performing a regression with 116 data points, I think the paper would improve if a leave-one-out cross-validation was conducted, and the RMSE of that prediction was calculated. This will help validate the empirical model and will allow determining its prediction ability, within the range of measured permeability values.

Response 4:

We fully agree that incorporating a formal cross-validation procedure strengthens the assessment of model predictive capability. Accordingly, Section 3.5 has been substantially revised to include both leave-one-point-out cross-validation (LOOCV) and leave-one-well-out validation.

Specifically:

- A sequential LOOCV procedure was implemented using all 116 paired resistivity–permeability observations.
- In each iteration, one data point was excluded, the regression recalibrated using the remaining 115 observations, and permeability predicted for the excluded point.
- Prediction error was quantified using the root mean square error (RMSE), explicitly defined in Eq. (5).
- The resulting $RMSE_{LOOCV}$ was 2.36 mD.

Given the measured permeability range (0.01–19.9 mD), this value indicates stable predictive performance within the calibration domain. For the representative points listed in Table 3, LOOCV-predicted values differ only marginally from those obtained using the full regression, confirming parameter stability.

To further assess spatial transferability, leave-one-well-out cross-validation was performed. In this approach, all data from one borehole were excluded, the regression recalibrated using the remaining wells, and permeability predicted for the omitted borehole. This yielded:

$RMSE_{well}$ = 2.78 mD.

The modest increase relative to pointwise LOOCV reflects geological heterogeneity rather than model instability, demonstrating that the empirical resistivity–permeability relationship generalizes reasonably well across different boreholes and lithological domains.

These revisions are included in:

Track Changes Copy: Page 34-35, Line 552-580

Clean Copy: Page 34-35, Line 551-579

“3.5 Validation of predicted vs. measured permeability

……..Because the empirical model was derived from 116 paired measurements, predictive capability was evaluated using leave-one-point-out cross-validation (LOOCV). Each observation was excluded sequentially; the regression was recalibrated using the remaining 115 data points; and permeability was predicted for the excluded point. Prediction error was quantified using the root mean square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(k_i - k'_i)^2} \qquad (5)$$

where k_i denotes the measured permeability, k'_i represents the predicted value for the excluded observation, and n = 116 is the total number of paired resistivity–permeability data points used in the cross-validation procedure. This procedure yielded:

$$\mathrm{RMSE}_{\mathrm{LOOCV}} = 2.36 \text{ mD} \qquad (6)$$

Given the observed permeability range (0.01–19.9 mD), this error indicates stable predictive performance within the calibration domain. For the representative points listed in Table 3, LOOCV-predicted permeability values differ only slightly from those obtained using the full regression model, indicating that the fitted relationship is not strongly influenced by individual observations.

To further assess spatial transferability, a leave-one-well-out cross validation was conducted. In this approach, all data from one borehole were excluded, the regression was recalibrated using the remaining wells, and permeability was predicted for the omitted well. Using the same RMSE formulation (Eq. 5), this yielded:

$$\mathrm{RMSE}_{\mathrm{well}} = 2.78 \text{ mD} \qquad (7)$$

The modest increase in error relative to pointwise LOOCV reflects geological heterogeneity rather than model instability. This result demonstrates that the empirical resistivity–permeability relationship generalizes reasonably well across different boreholes and lithological domains.

Together, percentage agreement, leave-one-point-out validation, and leave-one-well-out validation demonstrate that the CSAMT-derived empirical model provides robust and transferable permeability estimates within the measured range. Although localized deviations occur, likely due to structural anisotropy and fracture-controlled flow, the regression is not dominated by single data points or individual wells, supporting its application for regional-scale groundwater assessment."

And discussed further in:

Track Changes Copy: Page 48, Line 833-840

Clean Copy: Page 48, Line 832-839

"4.14 Optimizing borehole placement for CSAMT calibration

………..Cross-validation results demonstrate that calibration quality depends more on spatial distribution than on borehole number. Leave-one-out cross-validation (LOOCV) of the 116 paired measurements yielded an RMSE of 2.36 mD, indicating that the regression is not overly sensitive to individual data points. Leave-one-well-out cross-validation resulted in an RMSE of 2.78 mD, showing that the empirical relationship maintains reasonable predictive capability when applied to an entirely excluded borehole. The modest increase in error reflects geological heterogeneity rather than model instability."

We thank the referee again for the constructive recommendations. The revisions prompted by your comments have significantly strengthened the methodological rigor and clarity of the manuscript.