

We thank prof. Jie Liu for her constructive feedback. In the following, we will address all the comments and explain how we intend to modify the manuscript following the reviewer's suggestions. Reviewer's comments are reported in black and the replies in blue, line numbers refer to the submitted version of the manuscript.

I. The practical motivation for defining an upper REA bound is not well established.

The paper presents the determination of an upper bound of the REA range as a key novelty. However, the practical motivation for defining such an upper bound is not well established. The REV/REA literature has historically focused on the lower bound, largely because the engineering requirement is to identify the smallest scale at which a property is still representative. The need for a statistically defined upper bound is far less clear: a scale larger than the lower bound remains representative in principle, and grid resolution in reservoir models is typically constrained by geological boundaries (e.g., lithological contacts, fault zones) rather than by the upper limit of statistical homogeneity.

If the upper limit of REA is intended to define the maximum cell size in numerical modeling (Lines 67–72), the result for the Lilstock outcrop (0.3 m) is counterintuitive when viewed against Fig. 2b, even though it may be justifiable on purely statistical grounds. Using the maximum cell size of 0.3 m to perform numerical modeling of this area (Fig. 2b) is not reasonable from the perspective of a numerical modeller. The paper would benefit from a more critical discussion of why an upper REA bound, derived solely from P21 statistics, should be preferred over, or even considered alongside, geologically based cell-size decisions.

We thank the reviewer for this comment. While the upper limit has not been specifically addressed in previous literature, this does not mean it is unimportant. In fact, as acknowledged by the other reviewer, it is important to establish boundaries for the representative area or volume. The fact that a particular outcrop yields counterintuitive results is dependent on its specific characteristics and does not undermine the validity of the method itself. The method serves to establish a quantitative baseline, allowing the modeler to subsequently choose the best discretization strategy.

II. The manuscript's organizational logic needs improvement.

1) Placing a subsection entitled "Data collection" within the "Method" section is confusing.

Done. The section "Data collection" has been moved before the "Method" section, as section 3. The number of the other sections has been consequently adjusted.

2) In Line 161, Eq. [4] is referred to, yet Eq. [4] is not presented until Line 216, several subsections later.

Done. We removed Eq. [4] from line 161.

3) Figures 4 to 7 are mentioned in Section 3 (Method), whereas these figures are explained in Section 4 as results.

Done. We removed the references to figures presented in the results from the method section, in particular from line 190 and 277

4) The Method section should introduce the concepts and methods used in the study using purely mathematical language and expressions. The subsequent section can then describe data acquisition (including the studied sites) and implementation details, accompanied by necessary schematic figures, but without forward references to figures that belong to the Results section. By this standard, the implementation of the analyses has not been explained clearly.

As explained in the previous comments, the section “Data collection” has been moved after the “Method” section as a standalone section, and we removed all the forward references.

5) In the paragraph from Line 324 to Line 332, the discussion order is confusing: Fig. 9C and Fig. 9D are discussed first, followed by Fig. 10, and then Fig. 9A and Fig. 9B.

Done. The figure 9C has been cited before figures 9A and 9B because in the sentence we specifically addressed the group size = 5 configuration. Figure 9 goes from group size = 3 to group size = 6, so putting the group size = 5 as the first plot, because it's the first nominated in the text doesn't seem logic. We have modified line 324 to respect the order in which the figures appear:

Original (line 324):

Levene test results show that the tested groups remain representative up to a range of five factors (Fig. 9C).

Revised:

Levene test results show that the tested groups remain representative up to a range of five factors (Fig. 9A, B, C).

III. The manuscript requires much more concise writing.

Some statements are repeated several times across different paragraphs or sections, for example, regarding the calculation of P21. The sentence in Line 175 ("P21 is not calculated using ...") is unnecessary.

More broadly, the manuscript contains many unnecessarily detailed interpretations and self-explanatory expressions, which greatly reduce readability. The two paragraphs from Line 145 to Line 158 serve as a representative example:

In the first paragraph (Line 145), the rationale that random sampling ensures independence is unnecessarily elaborated for a specialist audience. The boundary-handling procedure is also described in excessive detail, including a bullet-like enumeration within a sentence and an explicit justification for allowing tangency. Furthermore, the explanation that larger scan areas lead to more frequent boundary rejections and that areas larger than the interpretation boundary cannot be placed is largely self-explanatory. This paragraph could be substantially shortened without any loss of methodological rigor.

In the second paragraph (Line 154), the comparison with grid sampling is introduced via a cumbersome, literature-review-style sentence, when a simple "Compared to grid sampling" would suffice. The subsequent justification for equal sample sizes, while correct, expands into a general discussion of statistical power and the risk of misleading comparisons, complete with a citation. A single sentence stating that equal sample sizes were maintained to ensure comparability of statistical tests would convey the same message more efficiently.

I recommend a thorough revision of these and similarly verbose passages throughout the manuscript. The focus should be on reporting what was done and why in the briefest possible terms that a domain expert would require, eliminating over-justification and redundant explanation.

Done. We thank the reviewer for the comment. Shortening and removing certain sentences will certainly help to improve the readability of the manuscript. Below is a list of the sections of text that have been modified or removed:

Original (lines 145-147):

The independence of observations within each factor is ensured by employing a random sampling with replacement strategy. This approach guarantees that the selection of one sampling unit does not influence another, thereby satisfying the statistical requirement for independence regardless of the underlying spatial correlations of the fracture network. In this approach, scan area centres are randomly positioned within the interpretation boundary. If a scan area intersects the interpretation boundary, it is rejected and randomly repositioned until the conditions are met, that are: (i) complete inclusion into the

interpretation boundary, or (ii) tangency, that is permitted to ensure equal sampling probability across the outcrop. This step represents the most computationally demanding part of the workflow. Increasing the scan area radius leads to more frequent rejections due to boundary intersections, and scan areas larger than the interpretation boundary are systematically rejected.

Revised:

To ensure the independence of observations, we employ a random sampling with replacement strategy, positioning scan area centers randomly within the interpretation boundary. Any scan area that intersects the boundary is rejected and randomly repositioned until it is fully included or tangent. Handling these boundary conditions represents the most computationally demanding part of the workflow.

Original (lines 154-158):

With regard to grid sampling strategy, for example that employed in Casiraghi et al. (2025b), or Martinelli et al., (2020) random sampling allows the same number of P_{21} data to be collected for every scan area radius. Even if it is not an underlying assumption, equal sample size is a desirable condition when it comes to testing different groups of data. Sample size is related to the power of a statistical test and comparing the test results of samples with different sample sizes will produce misleading results (Kozak and Piepho, 2018).

Revised:

Compared to grid sampling (Martinelli et al., 2020; Casiraghi et al., 2025b), random sampling allows the same number of data points to be collected for every scan area radius. Equal sample sizes were maintained across all radii to ensure the comparability and reliability of the statistical tests (Kozak and Piepho, 2018).

We decided to substitute the P21 calculation definition at lines 173-175 with the definition at lines 74-76 and consequently delete lines 74-76.

Original (lines 74-76):

P21 is computed by placing sampling areas of arbitrary shape (e.g., squares, hexagons, or circles) and size within the outcrop boundary, measuring the total cumulative length of fractures (or fracture segments when only partially contained), and dividing this by the area of the sampling window.

Revised: deleted in the revised version.

Original (lines 173-175):

P21 values are straightforwardly obtained by calculating the total sum of length of the fracture traces (complete or partial) inside the circular scan area and dividing it by the scan area surface area. P21 is not calculated using estimators (e.g. Mauldon et al., 2001).

Revised:

P21 is computed by placing sampling areas of arbitrary shape (e.g., squares, hexagons, or circles) and size within the outcrop boundary, measuring the total cumulative length of fractures (or fracture segments when only partially contained), and dividing this by the area of the sampling window.

We removed the description of the rationale of the presented methodology from the introduction, as it is already present in the “Methods” section.

Original (lines 83-88):

Each parameter is assessed using a different statistical test and an associated diagnostic plot. The shape of the distribution is evaluated using the Shapiro–Wilk test (Shapiro and Wilk, 1965), under the assumption that P21 values follow a normal distribution within the REA range. Test results are checked with normal probability plots. Variance equality among groups of P21 values obtained with different scan area sizes is tested using Levene’s test (Brown and Forsythe, 1974), and results are visually inspected with residual plots (Kozak and Piepho, 2018). Differences in mean values between groups are assessed through one-way ANOVA test (Stahle and Wold, 1989).

Revised: deleted in the revised version.

Original (lines 275-276):

Having described the core features of the presented statistical method, it is now possible to describe the testing procedure for REA range definition.

Revised:

The testing procedure for defining the REA range is executed as follows:

IV. Miscellaneous issues

1) It is not necessary to include P21 in the title, nor in the abstract.

Done. We removed “P21” from the title and the abstract.

2) Fig. 11 could be plotted in a single panel using narrow bars.

We thank the reviewer for the comment. We tried to plot all the box plots in a single panel, but the figure has become difficult to read, particularly the x-axis.

