We thank the reviewer for the useful comments and the time spent reading the paper. As for the first reviewer, 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.

Introduction: it is clear from the reading that the presented workflow is specifically intended to provide useful parameters to be considered in DFNs. For this reason, much space is dedicated to the limitations of these and the benefits that could be drawn from them. However, only one paragraph is dedicated to the true problems addressed in the paper, that is, to give statistical robustness to some of the major bias (four identified by the Authors) concerning the determination of fracture set/network in DOMs. This section should be expanded by considering more potential bias (for example subjective bias in the data collection/analysis or bias due to inefficient algorithms for the automatic extraction of planes) and an improved consideration of the relevant literature. Relevant papers to be mentioned are, among the others; Baecher, 1983, Zeeb et al., 2013; Zhang, 2016; Watkins, 2018, Andrews et al., 2019, Eppes, 2024.

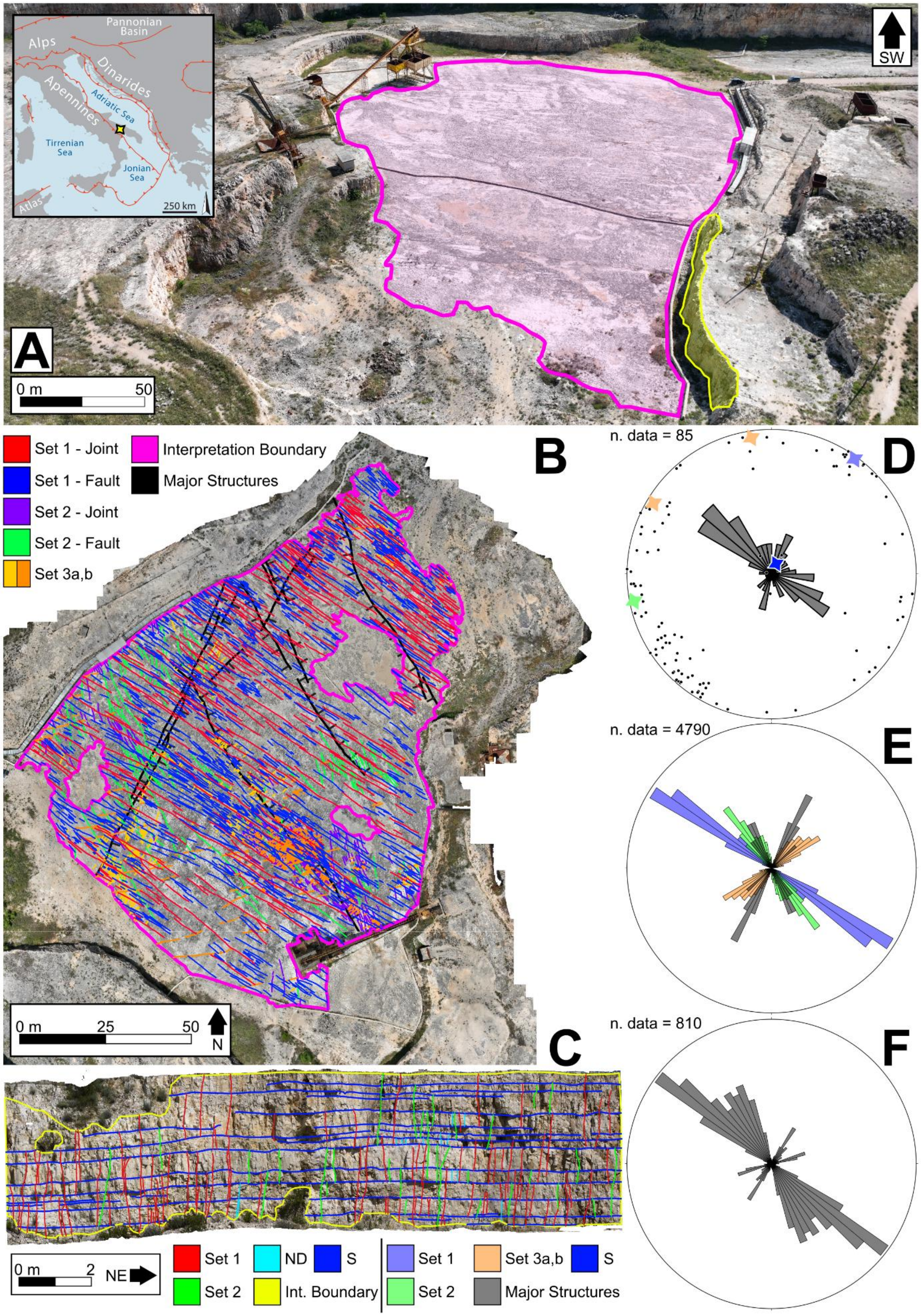
Done. We thank the reviewer for the useful comment and references. We agree that we didn't consider subjective biases and biases related to, for example automatic feature extraction method. We added a phrase to include them in the introduction:
Revised (from line 85):

In addition to objective biases related to outcrop geometry or sampling methods, subjective biases introduced by the interpreter should also be considered (Andrews et al., 2019). In the specific context of automatic feature extraction, it is also important to account for biases inherent to the algorithms themselves, including the potential for extracting artifact features.

Figure 2: the correspondence of colours between pavement/wall and steroplots is somewhat lacking; for example, red bars in E are missing (or are they placed together with blue one?). Colours of Set 1 and 2 are too similar, please use more contrasted colours. Field-data in D are shown without set differentiation but, as you also state in the text, fieldwork is the best way to distinguish fracture sets on the basis of their geological characteristics. In addition to the raw data, the cluster division proposed by the fieldwork should also be provided. In fig Fig 2C, the distinction between set 1 and 2 on the 2D orthoimage of the wall seems to be impossible because of lack of strike information. Later (lines 776-795), the Authors present a statistical correlation between height and length based on the assumption that traces mapped on the pavement and on the wall can be associated in ordered pairs from the shortest to the longest. Maybe in this particularly

simple geological setting this assumption could be plausible, however I think it is necessary to manually check the corresponding sets by mapping the 3D fractures traces and fit the 3D planes. I understand that, due to the flatness of the wall, it could be hard to accomplish for mostly visible traces but it can be do at least for some of them. This procedure can provide a significant validation of your workflow.

Done. We thank the reviewer for the comment, which gives us the opportunity to clarify an aspect we mistakenly took for granted. Starting with Figure 2, we agree that the colors chosen for Set 1 and Set 2 on the vertical wall are not sufficiently contrasted. We have now added the medoid orientation for each cluster in the filed data stereoplot. Each medoid is colorized with the color of the set according to the legend.



Pannonian Basin
Alps
Dinarides
Apennines
Adriatic Sea
Tirrenian Sea
Jonian Sea
Atlas
250 km
SW
A
0 m
50
Set 1 - Joint
Set 1 - Fault
Set 2 - Joint
Set 2 - Fault
Set 3a,b
Interpretation Boundary
Major Structures
B
0 m
25
50
N
n. data = 85
D
n. data = 4790
E
n. data = 810
F
C
0 m
2
NE
Set 1
Set 2
ND
Int. Boundary
S
Set 1
Set 2
Set 3a,b
Major Structures
S

The fractures were digitized on the vertical orthomosaic while taking the point cloud into account. Each vertical trace was associated with a plane characterized by dip and dip direction extracted from the point cloud. Some fractures on the vertical wall could not be reliably associated with a fracture plane and were therefore excluded from digitization. Additionally, on the vertical wall, we were unable to distinguish fractures with centimeter-scale displacement from those without. Consequently, we grouped features such as joints and faults into broader categories—e.g., combining both into a more general Set 1. We added a paragraph in section 5 to clarify this aspect:

Revised (from line 486):

The digitization of fractures on the vertical TS-DOM is supported by the corresponding PC-DOM. By integrating TS-DOM and PC-DOM data, each digitized fracture trace can be associated with a best-fit plane derived from the point cloud. This approach enables the assignment of fracture traces to specific fracture sets. Fractures on the vertical wall that could not be reliably linked to a fracture plane were excluded from the digitization process.

Line 212: You miss to specify the resolution of the terrestrial photogrammetric acquisition. Based on figure 3A the mean point spacing is around 0.5 pts/mm, which corresponds to ca. 2 mm of resolution (GSD). Therefore, the wall and pavement datasets are different in resolution (4 mm vs. 2 mm) possibly influencing the fracture mapping. I think the Authors should consider this in the methodology section.

Done. We thank the reviewer for this correction. The resolution of the vertical orthomosaics is 2 mm/pixel (specifically, 2.06 mm/pixel). We have now added this information at the end of the paragraph on line 220.

We do not consider this slight difference in resolution significant enough to affect the digitization process, as the length of the smallest detected features is on the order of some decimeters, approximately an order of magnitude larger than the orthomosaic resolution. However, we agree that if the orthomosaics were used to digitalize structures at the millimeter scale, such a resolution difference could indeed have a noticeable impact.

Revised (from line 220):

The resulting photogrammetric model has a resolution of approximately 2 mm/pixel.

Line 255-261: it seems that the Authors inverted the reference to the figure panel 3C and 3D.

Done. Thank you for the correction. We fixed the reference

Line 257-261: Why the Authors propose a comparison between a low-quality photo acquisition from commercial drone DJI Mini 3 Pro when they used a professional DJI Mavic 3E for the pavement acquisition, that can provide higher quality picture? Moreover, the comparison between the two datasets is unclear because the Authors never specified how the drone dataset was acquired (outcrop camera distance, scheme of acquisition, orientation of the camera, etc). The resolution of DOM mostly depends from the drone camera quality and distance of acquisition, therefore using professional drones and flying close to the outcrop it's possible to obtain high quality DOM, comparable to the terrestrial acquisition. Moreover, as visible in the figure 3C and 3D, the drone dataset suffers less of occlusions (i.e. shadow areas) making it more appropriate for 3D tracing and mapping. This is particularly true for large and high rock walls where the terrestrial acquisition cannot provide proper data. In conclusion, the presented comparison between the two datasets is weak and doesn't provide any significant contribution in the workflow.

Done. We thank the reviewer for the comment and the opportunity to clarify this point. The purpose of the comparison was not to compare ground-based and UAV photogrammetry per se, but rather to highlight the impact of using a high-quality camera and a proper acquisition scheme versus a lower-performing camera and a simplified acquisition geometry. This was illustrated through differences in point cloud surface density.

We fully agree that a UAV equipped with a professional-grade camera and a proper acquisition scheme could achieve results comparable to the ground-based survey. Likewise, a similar contrast could have been demonstrated using a ground-based survey performed with an entry-level camera. At the time of acquisition, the DJI Mini 3 represented the least capable instrument available to us, which is why it was selected for this illustrative comparison.

We believe the comparison remains meaningful, as the three-order-of-magnitude difference in point cloud density between the two models strongly supports the need for high-quality equipment and acquisition geometry—whether ground-based or UAV—for reliable data collection.

We have slightly modified the text to avoid highlighting the comparison between terrestrial survey and UAV.

Original (lines 255-259):

As an example, in Figure 3, two PC-DOMs of the same vertical outcrop are compared, collected in two different ways to 255 obtain a different SD. The PC-DOM in Figure 3D is reconstructed from more than 400 photos collected as discussed above (terrestrial survey with fans scheme, with high end Nikon Z7 mirrorless). On the other hand, the PC-DOM in

Figure 3C is collected with a smaller dataset (150 photos) collected with the lower quality camera of a small commercial drone (DJI Mini 3 Pro).

Revised:

As an example, in Figure 3, two PC-DOMs of the same vertical outcrop are compared, collected in two different ways to obtain a different SD. The PC-DOM in Figure 3C is reconstructed from more than 400 photos collected as discussed above (fans scheme, with high end camera, Nikon Z7). On the other hand, the PC-DOM in Figure 3D is collected with a smaller dataset (150 photos) collected with a lower quality camera (DJI Mini 3 Pro).

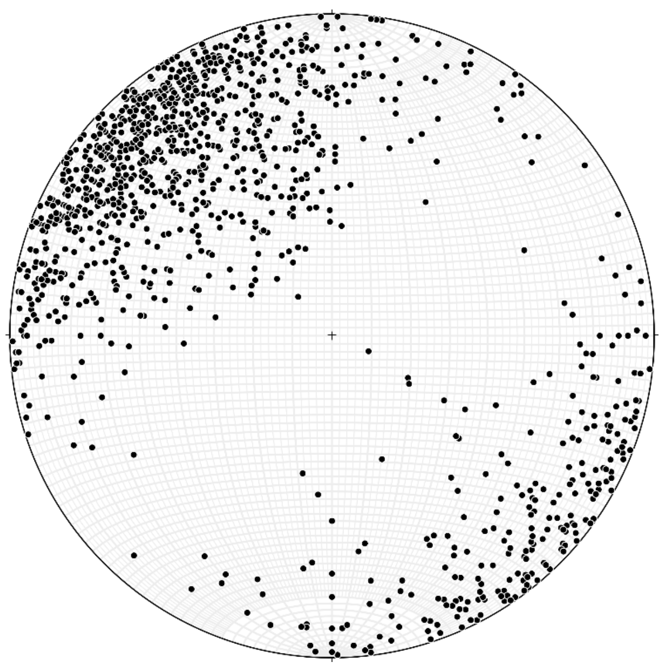
Cap 4 dealing with the semi-automated analysis of fracture orientation from point clouds. My major concerns are about the novelty of this workflow. It seems it has some similaritiy with the DSE of Riquelme et al., 2014. Please, provide a comparision. I also have doubts about its applicability if addressed to geologically complex setting. I guess that when rock masses are affected by multiple foliations, fracture sets, folding etc, it is virtually impossible to detect reliable pre-defined clusters, invalidating all the procedure. In similar cases, it is well known that semi-automatic methods for plane extraction have multiple bias affecting the quality of the data collection. I would like to invite the Authors to consider and highlight these limitations of the method and provide the warning that it can successfully used in areas with low complexity. In the other cases, manual mapping is still the more effective way to derive robust data, even taking into consideration the subjectivity bias (Andrews et al., 2019).

We thank the reviewer for this comment. Regarding the comparison with the method proposed by Riquelme et al., we believe that such a comparison would be ill-posed. Riquelme and co-authors propose a method for the automatic extraction of 2D features directly from point clouds. In contrast, our contribution focuses on a supervised calibration procedure intended to enhance the performance of such automatic methods.

A more appropriate comparison would be between the method developed by Riquelme et al. and FACETS (Dewez et al., 2016), the automatic feature extraction algorithm implemented in CloudCompare. We chose to work with FACETS because it is already integrated into CloudCompare; however, our calibration procedure is generalizable and could also be applied to the method proposed by Riquelme and co-authors.

Concerning the applicability of our approach to more geologically complex case studies, we argue that, from a methodological point of view rather than a geological one, the complexity is not primarily related to the number of fracture sets or the presence of foliations. Instead, it depends on the size of the point cloud patches that can be identified as planar features. Smaller patches require a higher octree level, which increases the risk of fitting 2D planes to noise. In this respect, the selected case study is particularly challenging, as many fracture planes are not continuous along their trace, and orientation measurements are based on small, isolated point clusters.

To further support the effectiveness of our calibration procedure, we include in this reply a stereoplot resulting from the application of the automatic feature extraction algorithm to the entire outcrop without any prior segmentation of the point cloud.



Cap 8, about fracture intensity. The Authors propose to use the Representative Elementary Area (REA) that is the area above which the value becomes independent from the position and scan area size. They use a lower threshold of REA defined as the minimum hexagon area where no significant difference is detected between the mean and standard deviation of $P21$ obtained at that area and at the next step. Due to the unequal sample size of different scan areas, the Authors propose a qualitative approach based on the difference between the interquartile range (deltaIQR) of two subsequential $P21$ samples, where the REA is reached when deltaIQR stabilizes around 0, displaying a plateau in the diagram of its variation. Given that in scan windows with size close to that of the outcrop the representativity is compromised by the too small sample size, it is also well-known that windows smaller of the average fracture dimension cannot correctly capture the geometrical features. The new proposed calculation gives a statistical confirmation of the prevoius approaches. These suggest that the size of the area required for a representative quantification of fractures depends on both fracture average length and number density (Rohrbaugh et al., 2002; Zeeb et al., 2013; Zhang, 2016; Eppes, 2024). I'm not entirely convinced by the plateau window indicated by the Authors. It could be widened or narrowed a bit without significantly affecting the statistical variations. In anycase, a range of 5-12 m is very large and fall in the interval derived from the standard approach. In general, the case study seems very homogeneous, so it is easy to see how the average P21 is stable for almost all samples. It would be more interesting to see the calculation on more heterogeneous outcrops. The question is whether it is useful to use a statistically averaged REA in heterogeneous outcrops, rather than adapting the window to geological variations?

Done. We understand that the reviewer has some reservations about the choice of the interval of sizes where P21 is stable. It is the downside of qualitative methods where the choice is left to the user. At the same time, we think that a qualitative method is still better than a quantitative one where the underlying assumptions are violated (sample size, normality, homogeneity of variance). About the application to more heterogeneous outcrops, it is beyond the scope of this paper, and we are currently working on a future paper about this topic. We added a sentence in the discussions to highlight this problem:

Revised (from line 969):

We also recognize that adopting a more qualitative approach may introduce subjectivity in the selection of window size, but still having an order of magnitude for the REA (and hence for REV) is important in modelling studies.

Line 763: Authors write "$P21$ REA can be safely calculated only for Set 1 fractures, because, as highlighted in Section 2, in some areas only Set 1 fractures can be digitized, while Set 2 and Set 3 are drowned by the quarrying related fractures". If this is true, which is the usefulness of this analysis? Authors declare that they only may safely map one set of fracture, whereas the rest of dataset is somehow masked, hidden, in any case not representative. So, what is the validity of the P21 dataset of only part of the factures)? Same question is valid for the others parameters (topology, H/L ratio, …)?

Some parameters are inherently influenced by the conditions of the outcrop. The P21 analysis, for example, is valid for Set 1 fractures. However, as correctly noted, the datasets for Set 2 and Set 3—derived from the pavement—are incomplete, making it unreliable to calculate a representative elementary area (REA) for P21. This is not a limitation of the method itself, but rather a limitation of the specific case study. Even the use of alternative methods would yield similarly unreliable results due to the incomplete dataset.

This limitation does not apply to all parameters. For the length distribution, we were able to gather a substantial amount of data for all fracture sets, as shown in Tables 5, 6, and 7. The same holds for the height distribution, with the exception of Set 3.

Regarding the topological analysis and backbone extraction, it is explicitly stated in the Discussion section that the current backbone geometry is likely to change if the complete fracture network was available. This uncertainty is acknowledged and discussed in the context of the limitations imposed by partial exposure (starting at line 927).

Line 859: Authors claim for a “very high quality of our outcrop, with perfectly exposed horizontal and vertical surfaces”, but even in this case of exceptional outcrop, they calculate very few statistically robust parameters. A conclusion should therefore be to honestly say that an approach inclined towards statistical calculation, like the one proposed, has very little chance of being applicable in outcrops.

Done. In accordance with Reviewer 1’s suggestions, we have removed “very” and similar subjective terms. We acknowledge that the outcrop is not perfect. However, the parameters we did not calculate are not related to the applicability of our methodology, but rather to the limitations of the available data.

Specifically, the absence of Set 3 fracture traces on the vertical wall makes it impossible to calculate height distribution parameters for Set 3 and H/L ratio—regardless of the method used. Similarly, the limited presence of Set 2 and Set 3 fracture traces on the pavement prevents a reliable calculation of representative P21 values. While it would have been technically possible to distribute grids to the areas where Set 2 and Set 3 are present, the restricted areal extent would not have allowed us to investigate sufficiently large windows. This remains true even though we could have used windows larger than the mean trace length, following the criteria proposed by the reviewer in other comments.

Therefore, we are not in a position to assess whether the P21 values obtained for these sets would be representative, or whether the representative elementary area (REA) has been reached.

In any case all measured parameters are listed in Table 1, and we believe that they are the majority.

Line 865: Authors write: “The integration of facets and traces (collected both on horizontal and vertical outcrops) allows a complete characterization of fracture network parameters, unlike other approaches that rely on the analysis of only one of these two datasets (e.g. Ortega et al., 2006; Boro et al., 2014; Martinelli et al., 2020; Smeraglia et al., 2021).” However, Authors must consider they contradict what has just been said, that is they have failed to provide many, if not most, of the fracture parameters due to outcrop conditions. I also find it unfair to attribute incompleteness (which subtly suggests poor quality) to previous works that did not use a method similar to the one described by the Authors. Each of the cited works provides description of their approach, placing it in the existing literature and highlighting limitations. The Authors, rather than discrediting previous works, should focus on emphasise merits and limitations of their own research.

Done. This sentence has been partially modified following a comment by reviewer 1. We would like to make it clear that we do not intend to discredit other authors, and we apologize if we have given this impression through a poorly formulated sentence. Here is a revised version of the sentence:

Original (lines 865-867):

The integration of facets and traces (collected both on horizontal and vertical outcrops) allows a complete characterization of fracture network parameters, unlike other approaches that rely on the analysis of only one of these two datasets (e.g. Ortega et al., 2006; Boro et al., 2014; Martinelli et al., 2020; Smeraglia et al., 2021).

Revised:

The integration of facets and traces (collected both on horizontal and vertical outcrops) allows a complete characterization of the parameters listed in Table 1, while other approaches rely on the analysis of facets or traces only (e.g. Ortega et al., 2006; Boro et al., 2014; Martinelli et al., 2020; Smeraglia et al., 2021).

Cap 11. I suggest the Authors to avoid the continuous use of terms such as "robust" or "rigorous analysis" in contrast to what done in the past. It seems to read that the Authors discover now how to manage DOM fracture data. This is not the case. The paper has some merits that I recognize and that should be rightly highlighted. However, it also has many limitations, as highlighted by the Authors themselves. I therefore ask to review the way in which this discussion is presented.

We acknowledge the reviewer's observation and have revised the discussion accordingly, taking into account the comments provided by both Reviewer 1 and Reviewer 2. In the revised version, we have avoided overly assertive terms such as 'robust' or 'rigorous' when referring to our approach.

Line 886: Why "unfortunately"? I suggest to avoid moralisms. Moreover, why these papers among the others? Practically the entire community takes the same assumptions. I guess that this is simply an aspect not considered in much of the previous research. I suggest to the Authors to highlight the novelty of their statistical approach. So, these sentences need adjustment so as not to be misleading.

Done. We apologize for the poor wording of the sentence. We have removed the term "unfortunately" which is indeed unnecessary. On the other hand, we would like to highlight that carrying out a "rigorous analysis" includes being self-critical, considering both pros and cons of the method. In any case, although we originally cited a few recent papers as examples, we agree that this practice is common and does not require specific references. Here is the revised version of the bullet point:

Original (lines 885-887):

Testing the fitted orientation distributions with goodness-of-fit tests, instead of assuming circular symmetry and a Fisher distribution without a proper statistical test as (unfortunately) is a common practice in structural geology (e.g. Bisdom et al., 2014; Smeraglia et al., 2021; Menegoni et al., 2024; Panara et al., 2024, just to cite some recent papers).

Revised:

Rather than assuming circular symmetry and fitting a Fisher distribution without prior statistical verification, our approach explicitly tests the fitted orientation distributions using goodness-of-fit tests. This provides a more robust and statistically grounded assessment of fracture set orientation parameters.

Line 928-936: here the Authors seem to make explicit the main problem of the presented approach based on a statistical validation of each parameter. The robustness of the statistical analysis is effective only if with truly complete fracture mapping along the entire outcrop. The presence of even small hole in the dataset (i.e. debris patches or not perfectly exposed walls) can invalidate the entire results. The Authors need to highlight these limitation and the repercussions on the applicability of the method in other settings.

We thank the reviewer for this comment. However, we respectfully disagree with the statement that topological analysis is invalid due to the presence of no-data zones. If this were the case, it would imply that most published studies involving fracture network topology—many of which are based on incompletely exposed outcrops—would also be considered invalid. The limited extent of the outcrop itself often represents a more significant constraint. We agree that if the no data zones cover most of the outcrop surface or their size match the outcrop scale, probably the analysis would be compromised. But this is an outcrop selection problem.

In our study, we explicitly addressed the impact of no-data zones in several ways. For length and height distributions, we identified censored fractures using B-nodes and applied a survival analysis approach to correct for censoring bias. From a topological perspective, we accounted for B-nodes by identifying and removing them to avoid underestimating the connectivity index. These steps were taken specifically to mitigate the limitations associated with partial data and ensure that the results remain as representative as possible.

Line 960-961: why “arbitrary” and why “without defining a proper representative sampling area”? Previous studies constrain the size of the window area on the fracture average length and the number density as described by, among the others, Zeeb et al., 2013, Zhang, 2016, etc The use of the REA is a novelty in the DOM analysis but it’s not the unique reliable method. Previously defined standards for the scan window definition are based on more empirical data rather than purely statistical approach, i.e. performing multiple tests in different geological contexts, with changing operators, outcrop conditions, varying fracture intensity homogeneity, which often doesn’t match the ideal one described in this paper.

Done. We apologize for the misleading statement regarding the methodologies cited. It is correct that a correlation between mean fracture trace length and the minimum scan area required for a representative fracture intensity calculation was provided in the referenced works. Nevertheless, defining the mean trace length without defining the statistical distribution is meaningless (as demonstrated by Benedetti et al, 2025). For instance, the REA would not exist at all for a strictly fractal (power-law) length distribution. In any case we have revised the paragraph, removing the misleading statement and placing greater emphasis on the specifics and contributions of our own method.

Original (lines 960-963):

Areal fracture intensity $P21$ is quite often calculated using scan areas of arbitrary size, without defining a proper representative sampling area (e.g. Bisdom et al., 2014; Menegoni et al., 2024; Panara et al., 2024). To our knowledge, only Martinelli et al. (2020) presented an analysis allowing to define the minimum REA where fracture intensity can be mediated to ensure a proper continuum-equivalent description (Bear, 1975).

Revised:

Areal fracture intensity is often estimation using methods based on scan lines, scan areas or circular scan line (Rohrbaugh Jr. et al., 2002; Zeeb et al., 2013). These methods provide a minimum scan area size for a representative estimation of P21 based on the mean fracture trace length. In this contribution we proposed a different approach, based on the concept of Representative Elementary Area to try to quantify the range of scan area size in which fracture intensity can be mediated to ensure a proper continuum-equivalent description (Bear, 1975).

Conclusions don’t fit the Introduction themes, and they present main results in a too local way. I suggest to better discuss the improvements of the new statistical approach that, tahnks to the related algorithms, allows to improve the determination of the fracture parameters. On the other hand, the important limits of applicability of the methodology should be highlighted even in cases of top-quality outcrops.

Done. The conclusions have been revised in accordance with the suggestions of both Reviewer 1 and 2. We attach the modified version of the conclusions for reference:

Revised:

In conclusion, this paper presented a series of quantitative methodologies to characterize fracture network geometry from Digital Outcrop Models (DOMs). Among all the parameters required to fully characterize a fracture network we focused on those required to generate 3d stochastic DFN models, that are: Orientation parameters, Topological relationships, length and height distribution parameters, H/L ratio and P21:

- Orientation data are collected through a semi-automatic workflow, divided into cluster via a clustering algorithm (k-medoid) and tested for the goodness-of-fit to a Fisher distribution. Alternatively, the Kent distribution parameters are also provided. This procedure allows subjectivity to be removed from the assignment of dip/dip direction data to a specific fracture set and supports the choice of meaningful orientation parameters through the implementation of statistical tests.
- Topological relationships are calculated including the interpretation boundary, this allows to: (i) to define B nodes and exclude them from the connectivity index (CI) calculation (ii) to identify censored fractures in an automatic way. Backbone extraction highlights the presence of large, connected clusters in the network. Crosscutting and abutting relationships between different fracture sets are quantified through directional topology.
- The approach developed to deal with censoring bias provides as a result a set of fully specified distributions (all parameters are explicit) corrected for censoring. The best model among the initial selection is defined through a graphical approach and a series of statistical distances.
- Estimating H/L was not possible without introducing some assumption, even for the best exposed set and in presence of both horizontal and vertical exposures. Therefore, we opted to make our assumption as transparent and possible, and testing it with regression analysis.
- P21REA is calculated with a qualitative approach, to avoid violating the underlying assumption of more formal statistical tests.