The manuscript presents a thorough and detailed work illustrating the issues related to subjective biases during fracture collection processes. A variety of acquisition methods for fracture attributes (i.e. scan-line, window sampling, circular scan-line and topological sampling) are introduced and compared to assess their response to subjective biases during collection. The results obtained in three different 'interpretation sessions' (two workshops using images and one in the field) are then used by the authors to draw conclusions on the impact that subjective biases induce on the parameter estimation acquisition methods to and to build a protocol to reduce the effect of subjective biases during fracture data collection. The concepts will be of great interest to the Solid Earth readership since fractures play fundamental roles in many applied settings and I suggest minor corrections for the manuscript.

**Major comments**:
In my opinion, a fundamental underlying issue that has not been addressed by the authors directly relates to use of their data to draw conclusions on the accuracy in the parameter estimations of acquisition methods. In particular, do the authors have taken into consideration the possibility that the errors and uncertainties related with subjective biases can scale with the number of fractures measured in the network? Letting interpret a larger fracture dataset to participant would have reduced the uncertainties in the estimation of the fracture attributes, independently of the acquisition method used?

Although the authors clearly state that it is not in their aim to "collect sufficient fractures to represent the fracture network" and that "the tested scanlines were not designed to be statistically representative" (page 5 lines 28 – 30), at the same time they do "consider the effect of the variation on fracture statistics derived from data collected…" (page 3 lines 3 and 4) and they dedicate a Section on the "Effect of subjective bias if the derived fracture statistics" (Section 4). I fear that their conclusions on this specific matter lack of statistical robustness, because of small number of fractures in the samples.

A well-known behaviour in statistics is what is called the 'marginal error': the size (N) of a statistical sample affects the standard deviation (i.e. variability) of the same sample (Moor, D.S., McCabe G.D. "Introduction to the practise of statistics", 1999; pages 294-295; 391-392). The variability shown in this work suffers the relatively small sample size in the number of fractures interpreted by the participants. The variability in a sample (the spread of the sample distribution) matters as much as bias when building a robust and significant dataset.
Because N (i.e. the fracture sample) is the denominator of the sample st.dev. formula ($s = \sqrt{\frac{\sum_1^N (x_1 - \bar{x})^2}{N-1}}$) the st.dev. decreases as N increases It follows then that having less data in your sample gives more variation (and less precision) in the result of your statistics: it can appear that big variations occur between participants counting, for example, number of fractures, however this spread in the data reduces considerably with the number of fractures that can be counted in the network. To reduce the variability of a statistic a large sample needs to be used: large sample almost always give an estimate that is close to the true value.

I understand that reviewing the manuscript in the light this comment may take considerable time (due to re-running tests, therefore I suggest minor revisions for this manuscript; however, I advise the authors to account for these possible biases in their interpretation on fracture statistics throughout the manuscript.

**Minor comments**:
Page 2, Lines 13-14: Can you please review this sentence? As written it is not very clear.

Page 3, Line 20: Please check the use of 'Nc': should not be written as 'n-points' in concordance with nomenclature in the following sentence?

Page 4, Line 4: Please check the sentence: is there a 'where' missing between '… a technique' and 'all fractures…'?

Page 4, Line 14: I would suggest to add '… and window sampling' to the listed methods: "Trace lengths may be measured directly with linear scanlines and *window sampling*, or estimated…"

Page 5, Lines 10-11: Can you please fully explain how you measure connectivity in linear scanlines? Are you using only x- and y-connections?

Page 5, Lines 18-19: Can you please add the trending attitude for the sub-vertical joint sent? Is this a third set?

Page 5; Lines 24-25: Can you please write the size of the used circles? In this context, looking at Table 2: why did the size of circle change in different localities? How did you choose the size of the circle?

Page 5, Line 26: Missed a capital letter 'P' in Participants. For consistency, please consider changing the 'Nc' (throughout the manuscript) to 'n' or *vice-versa*.

Page 7, Line 30: 'Does not' instead of 'doesn't'.

Page 8, Lines 10-14: How do the authors assess variability? Does 'variability' refer to the statistical variance within a single scan-circle? To show the variability in your sample, I would suggest to accompany the mean values shown in the tables with variance or standard deviation. The word 'variability' has been used by the authors throughout their text, however 'variation' is never statistically evaluated.

Page 10, Lines 12-13: 'Participant 11 depicted' instead of 'depicts' and 'Participant 18 and 20' did not…' instead of 'do not', consistently with previous sentence.

Page 11, Line 29: How do the authors assess the trend? Only visually?

Page 11, Lines 31 – 32: How do you evaluate indicators for trends?

Page 12, Lines 9-10: The authors refer to two mean trace length values derived from two participants measurements, however I could not find these numbers. Can you please indicate to which table are you referring?

Page 12, Line 11: Can the authors, please, mention to what this '$R^{2}$' values stands for? Is it a coefficient of correlation?

Page 12, Line 22: Please check 'al', should it not be 'all'?

Page 12, Lines 25-26: Can you please further explain why window sampling is less subjected to biases?

Page 15, Lines 4-5: This sentence raises the question: what can be considered a 'tall geologist'? I would suggest to delete this assertion if not fully accompanied with demonstrations and scientific data.

Page 16, Line 21: Please consider changing 'won't' with 'will not'

Page 17, Lines 28-30: Please consider re-phrasing, as written the sentence is a bit obscure

Page 17, Lines 30-32: Would not this be known only after having analysed the whole fracture networks?

Page 27, Table 4: For the fracture count, it would be interesting to see the spreading of the data: i.e., the $\Delta$N (difference between Min and Max). Instead, for the trace length data do you have taken into consideration min and max within each individual/group observation? Is the Min and Max reported into the table a mean of the Min values? Similarly for the Max?

Page, 42 Figure 9: Why does the bin size vary within the same attribute?