Solid Earth Discussions

Solid Earth Discuss., referee comment RC1
https://doi.org/10.5194/se-2021-45-RC1, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on se-2021-45

Anonymous Referee #1

Referee comment on "Investigating spatial heterogeneity within fracture networks using hierarchical clustering and graph distance metrics" by Rahul Prabhakaran et al., Solid Earth Discuss., https://doi.org/10.5194/se-2021-45-RC1, 2021

Dr. Prabhakaran and co-authors here present a novel approach to spatial network analysis of fractures. This approach involves quantifying hierarchical clustering based on similarity of four statistics: fingerprint distance, D-measure, NetLSD, and portrait divergence. Hierarchical clusters are identified based on the similarity of areal sub-samples of a large fracture map (the example here being the Lilstock anticline, UK). The authors show maps of their results, noting apparent spatial autocorrelation in all analyses except for NetLSD.

The manuscript is mostly well written and appears to me to be mathematically sound. However, despite using several highly sophisticated statistics, the method relies on at least two rather qualitative and/or arbitrary choices: (i) the sampling window size and overlap amount, which would seem to exert an effect on the degree of the aforementioned spatial autocorrelation, and (ii) the selection of the number of clusters present, which should sensitively affect the resulting map patterns.

Moreover, the utility of the approach remains elusive, despite some time spent on it in the introduction and discussion. These four statistics are novel, but their meaning is fairly obscure. The main output of the technique is a map that highlights spatial variation in these statistics, which indeed could be useful. But the same types of patterns show up on maps of more inuitive statistics, like intensity (Figure 9, 14, 19). What does the technique tell you about the spatial arrangement of fractures, fracture connectivity, fracture drivers, etc., that exceeds the explanatory power of more conventional approaches?

Use of the term “clusters” (Line 96-99) is non-standard. Marrett et al. (2018), which the present paper cites, defines a cluster as a place where fracture spacing is smaller than it is elsewhere; as such, a uniformly spaced fracture pattern would lack clusters entirely. This definition is consistent with the framework described earlier in the present study (Line 26-7). Why the change in usage? I suggest using a different term for the output of the present approach that is not already widely used in pattern-quantification literature.

Line-by-line comments:

6: S/V agreement. Also 8, 10, 11, 289, 344.

21: ambiguous pronoun “they”

24: I suggest abandoning this acronym. It is only used once more in the paper (Line 396) and furthermore it does not stand for its definition.

27: See definition comment above; also, in this framework I believe that “irregularly spaced” is a synonym for “clustered” and “regularly spaced” is a synonym for “periodically spaced.”

42: None of these drawbacks is unique to 1D analysis.

57: I am not an expert on stationarity, but I don't think this is true. Can't an irregular pattern nevertheless be stationary, meaning have attributes that do not change with position, above a certain size-scale?

66: Also “called” nodes?

92: What do undirected and weighted mean?

Figure 2: The construction of this figure is very difficult to understand until we read about Algorithm 1 below.

106: HC undefined acronym until Line 250, and even then it should be made explicit.

124-30: I would add synkinematic cementation (Hooker and Katz, 2015, Am. J. Sci.) to this list.

125: “etc” inappropriate—if more studies have made the point and need to be listed, list

them; else put “e.g.” and then your chosen, representative references.

Table 1. Are these Edges and Nodes in the primal or dual graph sense?

141: I don’t understand: how is curvature illustrated in these plots?

176: Greater regularity: can you justify this? A long, thin rectangular block would be highly “regular” and yet have a shape factor near zero.

179: “distribution of block-face regions”—you mean areas?

203: “w.r.t” seems needlessly curt; why not spell it out? Also 424.

258: “HC clusters” redundant

260: “decisions on the height”—see major comment above about the arbitrariness of this decision. You put a tremendous amount of effort into quantifying aspects of the spatial arrangement, and then seem to make an arbitrary decision as to how to bin your clusters.

297: “correctly” a loaded term without more explanation.

298: I would omit “clear” and let the reader judge. Also “clearly” in 315.

Figure 18 caption: acronyms not used in figure?

342: “unsupervised”—see major comment above about arbitrary choices. The user has to make the call about what qualifies as a cluster here; I don’t think this approach matches the spirit of unsupervised learning.

357: “striking” for “trending”

360: “background-variation”—such an interpretive term, especially considering the only external clustering control you have eliminated, to some degree, is faulting.

386: "we will tackle”—this is inappropriate. Anyone who wants to extend your work is entirely free to do so. I am sure you are not intentionally “marking your territory,” but we need to be especially cognizant of this kind of thing, especially in these days of pre-prints, self-archiving, predatory journals, and similar avenues whereby unscrupulous workers can “jump the gun” on an idea before proper peer review.

391: “inherent non-stationarity” see comment on Line 57

392: define “disassortativity”

397: “stationarity” for “stationariness”?

397: “have to be made based on hard data”—see major comment on arbitrarily choosing the number of clusters, above. Have you made your decisions based on hard data?

408: define “modularity”

411: What is graph partitioning, and is that a worthy goal?

418: “hierarchical” for “hierarchically”?