Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-231-RC2, 2020

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Fault Network Reconstruction using Agglomerative Clustering: Applications to South Californian Seismicity" *by* Yavor Kamer et al.

Leandro C. Gallo (Referee)

len.gallo@gmail.com

Received and published: 13 August 2020

General comments

The authors propose an interesting approach by applying an unsupervised learning algorithm –agglomerative clustering- that groups together unlabeled data points to extract structural information from seismicity catalogs. Their method uses an already developed -but still relatively underused within Earth sciences- clustering technique that involves the grouping of 3D spatial distribution of seismicity according to candidate active faults. Other clustering techniques have been applied as a pattern recognition in earthquake catalogs, starting with Ouillon et al. (2008) who applied the k means

method -Ansari et al. (2009), Ouillon and Sornette (2011) and Wang et al. (2013) are some examples. The difference with other clustering techniques is that hierarchical clustering uses a bottom-up approach: each observation starts in its own cluster (having 4 points), and clusters are successively merged together. The main novelty here, therefore, is not in the clustering procedure itself, but in the bottom-up approach which is, in my opinion, a valuable step-forward towards the full understanding of natural fault network modelling. The approach is first applied to a single synthetic dataset, then to a real example featuring 3,360 points (the Landers 1992 sequence) and then, to the condensation (Kamer et al., 2015) of the KaKiOS-16 catalog (Kamer et al., 2016). The flow of the paper is globally clear, well-written and figures are suitable. That said, with the aim of making the manuscript more robust, I wish the authors had made a stronger effort to validate the application of the technique before its application to observed seismicity data. Its application to a single synthetic experiment is practical for making the whole workflow understood, however, it has no statistical significance in terms of method's sensitivity. Being that the synthetic experiment features a relatively small number of data points, I would rather advise the authors to apply the technique to a larger number of models featuring a different number of faults with diverse characteristics or orientations –without prior knowledge this would be computationally inexpensive. Assessing discrepancies between the true and the inferred plane segments in a number large enough would then allow statistically meaningful results that, in my opinion, would make the whole manuscript more robust. I will not insist to modify the current version, but I would urge to at least think about this before you send this off to the printing press. Having said that, I can recommend publication after some modification, as I believe that the aims and approach are valid and also has relevance to a number of applications within the geosciences (esp. seismic hazards and structural geology/geomatics, but also to the extraction of planar facets in digital outcrop models).

Specific comments

L. 52: The contribution of source code to this section as supplementary materials –or open-access code repositories like GitHub or Zenodo- would boost scientific progress and reproducibility. L. 53: I don't see this subsection appropriate for the "methods" section. L.86-88: The criterion applied for merging two clusters involves the minimum squared Euclidean distances, was this criterion chosen for any particular reason? Is there any other metric to use instead for clustering? I'm thinking about the Eigen-based parameters of the covariance matrix. It would be valuable some extra explanation. L. 110, Figure 2: for those who are unfamiliar with the method, the hierarchical, binary cluster tree is most easily understood when viewed graphically. It would be helpful for the understanding of those who are not familiar to add the associated dendrogram to this figure.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-231, 2020.