Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-231-AC2, 2020

Natural Hazards and Earth System Sciences
Discussions
Open Access
EGU

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

Yavor Kamer et al.

yaver.kamer@gmail.com

Received and published: 15 October 2020

Thank you for taking the time to review our paper and providing detailed suggestions. Below is our response to your comments and description of the modifications we made to address them.

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

Printer-friendly version

Discussion paper

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.

We agree with you that the synthetics provided previously did not allow for conclusions about the sensitivity and robustness of the method. We have therefore supplemented the synthetics section with a more elaborate study where we gradually increase the sampling of a ground truth fault network under different background noise levels and investigate the method's clustering performance using the Rand index. This is now covered in section 3.1 and the results are provided in Figure 5.

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.

Based on your suggestions we have made publicly available the codes for the agglomerative clustering and the codes for the generation and evaluation of the synthetic sensitivity analysis here: https://www.mathworks.com/matlabcentral/fileexchange/81193. The link is included in the "Code availability" section.

L. 53: I don't see this subsection appropriate for the "methods" section.

We have moved this part to a separate section after the introduction.

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.

Our selection of the Ward's criterion was motivated by its characteristic of producing

regular sized clusters. This is important for the atomization procedure because we want all clusters to have the same potential to merge and grow into bigger structures. Initially we also investigated using the Mahalanobis distance with single linkage, and using the Gaussian associated with the location uncertainty of each event without atomization. These methods were not successful in reconstructing the synthetic networks in the presence of background noise; hence we focused our attention on atomization using the Ward creation. We have added the following sentence to the method section.

"While there are many different linkage methods and distance metrics, here we have chosen to use the Ward's criterion because it produces clusters with regular sizes. This is important for the atomization procedure as we want clusters to have similar potentials to merge and grow into bigger structures."

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.

We have added the dendrograms for both datasets to Figure 2.

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

Printer-friendly version

Discussion paper