Geophysical Research Abstracts
Vol. 16, EGU2014-14941-1, 2014
EGU General Assembly 2014

Fault Network Reconstruction using Agglomerative Clustering: Applications to South Californian Seismicity

Yavor Kamer (1), Guy Ouillon (2), Didier Sornette (3), and Jochen Wössner (1)

(1) Swiss Seismological Service, ETH, Zürich, Switzerland, (2) Lithophyse, Nice, France, (3) Department of Management, Technology and Economics, ETH, Zürich, Switzerland

We present applications of a new clustering method for fault network reconstruction based on the spatial distribution of seismicity. Unlike common approaches that start from the simplest large scale and gradually increase the complexity trying to explain the small scales, our method uses a bottom-up approach, by an initial sampling of the small scales and then reducing the complexity. The new approach also exploits the location uncertainty associated with each event in order to obtain a more accurate representation of the spatial probability distribution of the seismicity.

For a given dataset, we first construct an agglomerative hierarchical cluster (AHC) tree based on Ward's minimum variance linkage. Such a tree starts out with one cluster and progressively branches out into an increasing number of clusters. To atomize the structure into its constitutive protoclusters, we initialize a Gaussian Mixture Modeling (GMM) at a given level of the hierarchical clustering tree. We then let the GMM converge using an Expectation Maximization (EM) algorithm. The kernels that become ill defined (less than 4 points) at the end of the EM are discarded. By incrementing the number of initialization clusters (by atomizing at increasingly populated levels of the AHC tree) and repeating the procedure above, we are able to determine the maximum number of Gaussian kernels the structure can hold. The kernels in this configuration constitute our protoclusters. In this setting, merging of any pair will lessen the likelihood (calculated over the pdf of the kernels) but in turn will reduce the model's complexity. The information loss/gain of any possible merging can thus be quantified based on the Minimum Description Length (MDL) principle. Similar to an inter-distance matrix, where the matrix element di,j gives the distance between points i and j, we can construct a MDL gain/loss matrix where mi,j gives the information gain/loss resulting from the merging of kernels i and j. Based on this matrix, merging events resulting in MDL gain are performed in descending order until no gainful merging is possible anymore.

We envision that the results of this study could lead to a better understanding of the complex interactions within the Californian fault system and hopefully use the acquired insights for earthquake forecasting.