Geophysical Research Abstracts
Vol. 19, EGU2017-4373, 2017
EGU General Assembly 2017

Automated classification of seismic sources in a large database: a comparison of Random Forests and Deep Neural Networks.

Clement Hibert, André Stumpf, Floriane Provost, and Jean-Philippe Malet

Institut de Physique du Globe de Strasbourg - CNRS UMR 7516, University of Strasbourg/EOST, 5 rue Descartes, 67084 Strasbourg (hibert@unistra.fr)

In the past decades, the increasing quality of seismic sensors and capability to transfer remotely large quantity of data led to a fast densification of local, regional and global seismic networks for near real-time monitoring of crustal and surface processes. This technological advance permits the use of seismology to document geological and natural/anthropogenic processes (volcanoes, ice-calving, landslides, snow and rock avalanches, geothermal fields), but also led to an ever-growing quantity of seismic data. This wealth of seismic data makes the construction of complete seismicity catalogs, which include earthquakes but also other sources of seismic waves, more challenging and very time-consuming as this critical pre-processing stage is classically done by human operators and because hundreds of thousands of seismic signals have to be processed. To overcome this issue, the development of automatic methods for the processing of continuous seismic data appears to be a necessity. The classification algorithm should satisfy the need of a method that is robust, precise and versatile enough to be deployed to monitor the seismicity in very different contexts.

In this study, we evaluate the ability of machine learning algorithms for the analysis of seismic sources at the Piton de la Fournaise volcano being Random Forest and Deep Neural Network classifiers. We gather a catalog of more than 20,000 events, belonging to 8 classes of seismic sources. We define 60 attributes, based on the waveform, the frequency content and the polarization of the seismic waves, to parameterize the seismic signals recorded. We show that both algorithms provide similar positive classification rates, with values exceeding 90% of the events. When trained with a sufficient number of events, the rate of positive identification can reach 99%. These very high rates of positive identification open the perspective of an operational implementation of these algorithms for near-real time monitoring of mass movements and other environmental sources at the local, regional and even global scale.