Geophysical Research Abstracts
Vol. 18, EGU2016-12895, 2016
EGU General Assembly 2016

Automated classification of seismic sources in large database using random forest algorithm: First results at Piton de la Fournaise volcano (La Réunion).

Clément Hibert (1), Floriane Provost (1), Jean-Philippe Malet (1), André Stumpf (1,2), Alessia Maggi (1), and Valérie Ferrazzini (3)

(1) Institut de Physique du Globe de Strasbourg - CNRS UMR 7516, University of Strasbourg/EOST, 67084 Strasbourg, (2) Laboratoire Image, Ville, Environnement - CNRS UMR 7362, University of Strasbourg, 3 rue de l'Argonne, 67000 Strasbourg , (3) Observatoire Volcanologique du Piton de la Fournaise, Institut de Physique du Globe de Paris, 4 RN3 - Km 27 97418 La Plaine des Cafres, La Réunion.

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. 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, that 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. 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.

We propose a multi-class detection method based on the random forests algorithm to automatically classify the source of seismic signals. Random forests is a supervised machine learning technique that is based on the computation of a large number of decision trees. The multiple decision trees are constructed from training sets including each of the target classes. In the case of seismic signals, these attributes may encompass spectral features but also waveform characteristics, multi-stations observations and other relevant information. The Random Forests classifier is used because it provides state-of-the-art performance when compared with other machine learning techniques (e.g. SVM, Neural Networks) and requires no fine tuning. Furthermore it is relatively fast, robust, easy to parallelize, and inherently suitable for multi-class problems.

In this work, we present the first results of the classification method applied to the seismicity recorded at Piton de la Fournaise volcano by the OVPF permanent observatory between May 2007 and January 2010. We selected a dozen of seismic signal features that characterize precisely its spectral content (e.g. central frequency, spectrum width, energy in several frequency bands, spectrogram shape, spectrum local and global maxima) and its waveform (e.g. duration, ratio between the maximum and the mean/median of the envelope amplitude, envelope kurtosis and skewness, polarization). This preliminary study shows that the classification accuracy is high, and insensitive to sampling permutations of training/validation sets. We also discuss the portability of the method in other contexts, and particularly for the detection of landslide hazards in mountainous areas.