

EGU2020-10234

<https://doi.org/10.5194/egusphere-egu2020-10234>

EGU General Assembly 2020

© Author(s) 2020. This work is distributed under the Creative Commons Attribution 4.0 License.



## Automatic classification of seismo-volcanic signals at La Soufrière of Guadeloupe

Alexis Falcin<sup>1</sup>, Jean-Philippe Metaxian<sup>1</sup>, Jérôme Mars<sup>2</sup>, Eléonore Stutzmann<sup>1</sup>, Roberto Moretti<sup>1,3</sup>, and Jean-Christophe Komorowski<sup>1</sup>

<sup>1</sup>Institut de Physique du Globe de Paris, Université Sorbonne-Paris-Cité, CNRS, Paris, France

<sup>2</sup>Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France

<sup>3</sup>Observatoire Volcanologique et Sismologique de Guadeloupe, Institut de Physique du Globe de Paris, Gourbeyre, France

Seismic activity at La Soufrière volcano of Guadeloupe is composed of various transient signals, which are classified manually by the Observatoire Volcanologique et Sismologique de Guadeloupe (OVSG-IPGP) considering waveforms recorded at several stations. Although five main types of signals are recognized in the data analysis by the observatory (Moretti et al., 2020), only three main classes readily distinguishable on seismic traces during the daily analytical protocol have been catalogued: Volcano-Tectonic events, Long-Period events and Nested events, each related to a distinct physical process.

Automatic classification of seismo-volcanic signals of La Soufrière was performed by using an architecture based on supervised learning, available at [github.com/malfante/AAA](https://github.com/malfante/AAA). Seismic waveforms are transformed into a large set of features (34 features for each representation domain) computed from three representation domain of the signal (time, frequency, quefrequency). The resulting vectors of features are then used for the modeling. We are using the Random Forest Classifier algorithm from the scikit-learn library.

At first, we trained the model with the dataset given by the OVSG consisting of 845 available labeled events (542 VT, 217 nested and 86 LP) recorded in the period 2013-2018. We obtained an average classification rate of 72 %. We determined that the VT class includes a variety of signals covering the LP, Nested and VT classes. Reviewing in details the waveforms and the spectral characteristics of the signals belonging to the 3 classes we then introduced Hybrid events and also defined a monochromatic class (so-called Tornillo) of LP signals, thus matching the full description of signals provided in Moretti et al. (2020).

Then, using the new information, a new model was trained with 5 classes and tested. We obtained a much better classification average rate of 84 %. The classification is excellent for Nested events (93 % of accuracy and precision) and Tornillo events (93% of accuracy and precision). The classification of VT events (90% accuracy, 89% precision) and LP events (86% accuracy, 82% precision) were also very good. The most difficult class to recognize is the Hybrid class (64 % accuracy, 69 % precision). Hybrid events are often mixed with VT and LP events. This may be explained by the nature of this class and the physical process that includes both a fracturing and a

resonating component with different modal frequencies.

Machine learning is a powerful tool to handle large datasets. From a dataset built manually, the processing we applied allowed to obtain a reliable automatic classification by refining class definitions. This has important implications for observatory data processing during unrest and eruptive activity.