AUTOMATIC CLASSIFICATION OF SEISMO-VOLCANIC SIGNATURES

Seismic Waves
Supervised Machine Learning
Random Forest
Support Vector Machine
Feature Space
Automatic Analysis
Volcanoes Monitoring

Marielle MALFANTE[1], Mauro DALLA MURA[1], Jérôme I. MARS[1], Orlando MACEDO[2], Adolfo INZA[2], Jean-Philippe METAXIAN[3]

[1] Université Grenoble Alpes, CNRS, Grenoble-INP, GIPSA-Lab, F-38000 GRENOBLE, FRANCE

[2] Instituto Geofísico des Perú, Lima, PERU

[3] ISTerre, IRD, CNRS, Université Savoie Mont Blanc, F-73376 LE BOURGET DU LAC, France

This work is funded by DGA/MREIS and Labex OSUG@2020.

BACKGROUND

In the context of volcanoes monitoring, the evolution of **seismicity** is one indicator of **volcanic unrest**. We here propose an architecture to automatically classify volcano-seismic events into one of six seismic classes. Our system reaches 90% of accuracy.

ABOUT MACHINE LEARNING

“Field of study that gives computers the ability to learn without being explicitly programmed.”

— **A. Samuel**
1901-1990
Pioneer in artificial intelligence
Wrote the first self-learning program

“Can machine think?”

— **A. Turing**
1912-1954
Computer scientist and cryptanalyst
Father of theoretical computer science & artificial intelligence

“All models are wrong, but some are usefull.”

— **G. Box**
1919-2013
Statistician
« One of the great statistical minds of the 20th century »

CONTACT

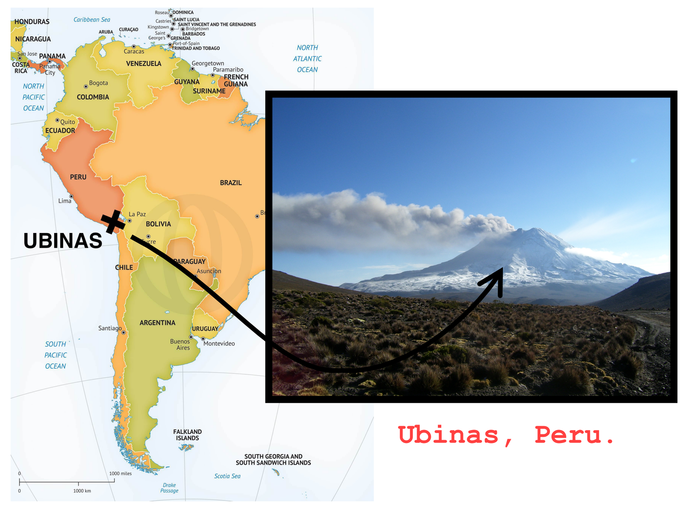
▸ **Marielle MALFANTE**
PhD Student
marielle.malfante@gipsa-lab.fr

SYSTEM ARCHITECTURE

The system we present is an automatic classifier of seismic events associated to volcanic activity. Its architecture is built upon Supervised Machine Learning and is in Figure1:

1- Dataset constitution
2- Features extraction
3- Learning & Testing

Ubinas Volcano



Ubinas is a Peruvian andesitic stratovolcano (southern Peru 16° 22' S, 70° 54' W – alt. 5676m). Considered as the most active Peruvian volcano, it is closely monitored by the Instituto Geofisico del Peru. After nearly 40 years of quiescence, Ubinas volcano erupted in 2006. Three eruptive phases occurred since 2006, respectively 2006-2011, 2013-2014 and 2016. Ubinas is monitored since 2006, with four seismic stations and tiltmeters. Data are recorded at 100Hz.

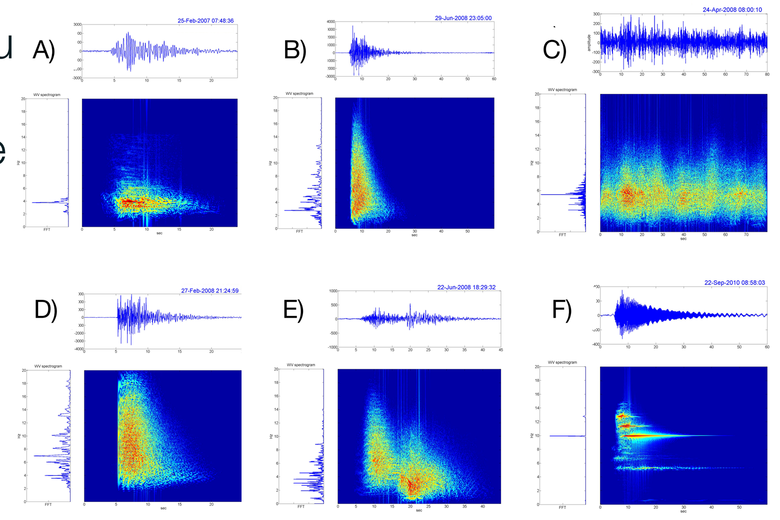


Figure 2: Signals from Ubinas. Six classes:
a) LP = Long Period
b) EXP = Explosions
c) TR = Tremors
d) VT = Volcano Tectonic
e) HIB = Hybrid (VT/LP)
f) TOR = Tornillo

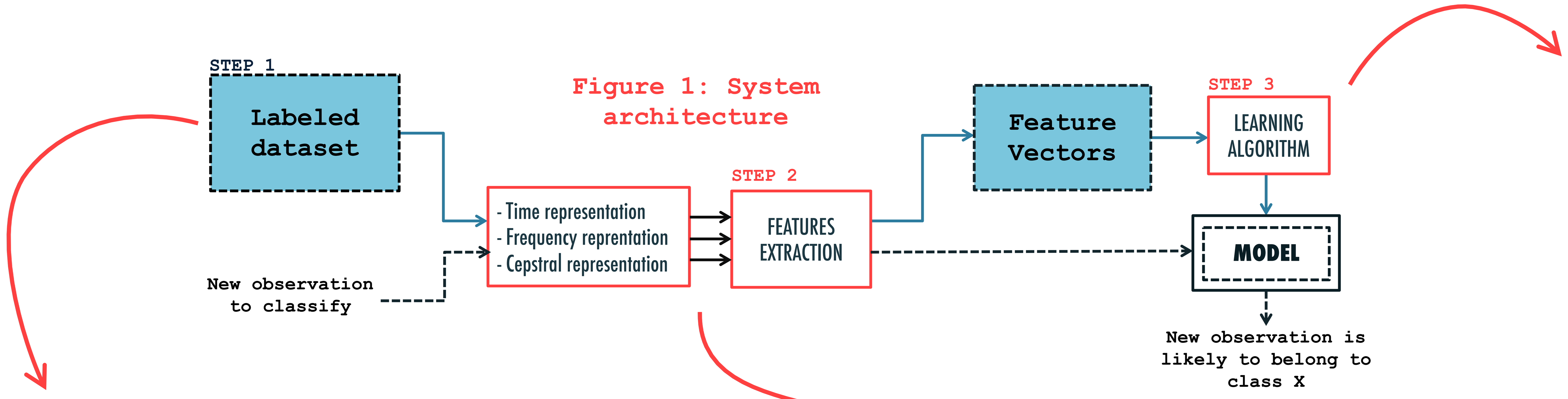


Figure 1: System architecture

3 LEARNING & TESTING

By using the feature vectors in a learning algorithm, a model is built (Figure 5). This is known as the **learning** or **training** phase. In this work we used **Random Forest**[1,2] (RF) or **Support Vector Machine**[3] (SVM) algorithm. Learning is done on a fraction **α** of the dataset (learning rate) and is tested on the remaining (1-**α**) observations. This process is known as cross validation.

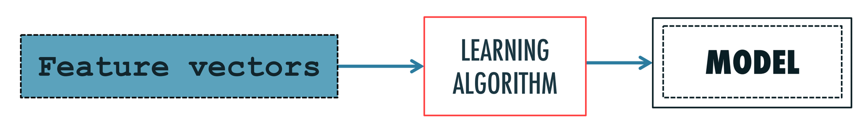


Figure 3: Learning

1 DATASET CONSTITUTION

Considered number N_i of observations for each class i, each observation corresponds to a **volcano-seismic event** sampled at 100Hz. 3125 seismic events are considered, recorded between 2006 and 2011 (observations energy are normalized).

Each observation has manually been assigned to its class by experts. The dataset is then said to be ‘labeled’.

Ref.	Description	N_i	Δ_t (in days)	Mean length
LP	Long period	800	201	49s
TR	Tremors	800	8	27min48s
EXP	Explosions	154	1396	51s
VT	Volcano-tectonic	800	1958	24s
HIB	Hybrid	466	1647	34s
TOR	Tornillo	105	632	40s

Table 1: Labeled dataset considered for this study

2 FEATURES EXTRACTION

Feature extraction is about extracting relevant information that will represent each observation. Representing each observation in the feature space instead that by time serie, greatly improves classification results.

Contribution 1 ▸ Features used in this work come from an extensive state of the art in signal representation for classification purposes in seismic and acoustic fields. Those 32 features will generally describe the signal. Especially, we use mean values, standard deviations, kurtosis, skewness (high order moments), entropies, threshold crossing rate, central frequency, etc.

Contribution 2 ▸ Features are extracted from three different representations of the observation: the time serie x(t), the spectrum $X(f) = \mathbb{TF}(x(t))$, and the spectrum of the spectrum $X(q) = \mathbb{TF}(X(f))$. This last domain is known as cepstral domain and is originated from speech processing to highlight harmonic properties of a signal.

Each observation is represented by a feature vector of dimension 96.

RESULTS

Main results include a validation of our architecture with 90% of correct classification and the analysis of errors made by the model. Results between Support Vector Machine and Random Forest are similar.

	TIME	FREQUENCY	CEPSTRAL	ALL FEATURES
ACCURACY	86.1 ± 1.0 %	83.0 ±1.0 %	79.4 ± 1.0 %	**90.1 ± 0.9 %**

Table 1: Accuracy results for various feature sets. Training algorithm: SVM, kernel rbf, C=10, gamma = 0.01.

PREDICTED CLASS \ TRUE CLASS	LP	TR	EXP	VT	HIB	TOR
LP	218	16	0	1	6	1
TR	17	223	0	1	1	0
EXP	0	0	40	5	0	0
VT	1	0	6	219	11	7
HIB	3	1	0	12	121	0
TOR	0	0	0	1	0	24

Table 2: Confusion matrix associated to the best model.

PROSPECTS

▸ Automatic analysis of continuous signals and application for volcano monitoring.
▸ Anomaly Detection with unsupervised learning
▸ Lowering the labelisation constraint with semi-supervised models.

BIBLIOGRAPHY

[1] **Induction Decision Tree**, J.R. Quinlan, *Machine Learning*, 1986

[2] **Random Forest**, L. Breiman, *Machine Learning*, 2001

[3] **A training algorithm for optimal margin classifiers**, B.E. Boser, I.M. Guyon, V.N .Vapnik, *Proceedings of the fifth annual workshop on Computational learning* 1992