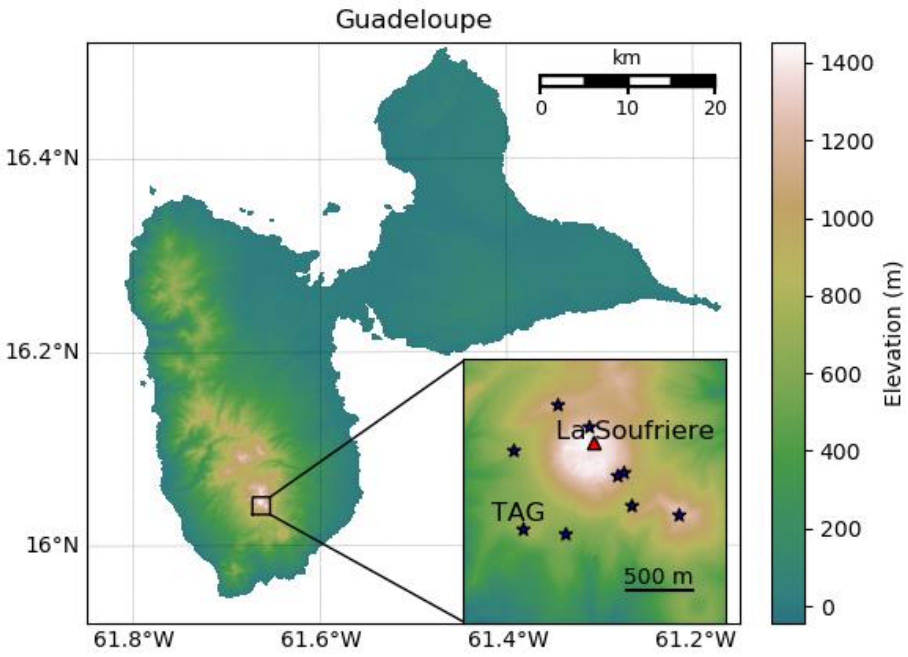
Automatic classification of seismo-volcanic signals at La Soufrière volcano, Guadeloupe

Alexis Falcin[1] (falcin@ipgp.fr), Jean-Philippe Métaxian[1,2], Jérôme Mars[3], Eléonore Stutzmann[1], Jean-Christophe Komorowski[1], Roberto Moretti[1,4]

1- Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France
2- Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, IFSTTAR, ISTerre, 38000 Grenoble, France
3- Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France
4- Observatoire volcanologique et sismologique de la Guadeloupe, Institut de physique du globe de Paris, F-97113 Gourbeyre, Guadeloupe

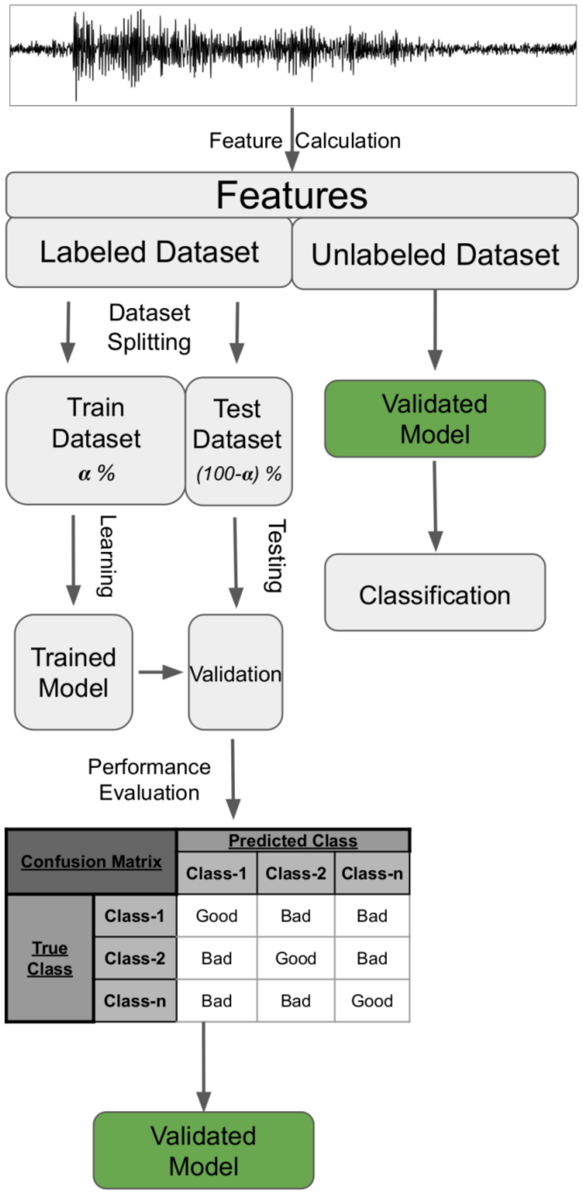
EGU General Assembly

Data



- **La Soufrière :** Active volcano located on the island of Basse-Terre in Guadeloupe in the Lesser Antilles.
- **Last Major Eruption :** 1976-77, required the evacuation of more than 70,000 people for six month.
- **Recent Activity :** Lowest measured level in 1990 [1], Increasing activity from 1992 to the highest level of seismic energy on April 27, 2018, with the strongest Volcano-Tectonic earthquake felt (M4.1) since the phreatic eruption of 1976 [2].
- **Detection/Classification :** Detection mostly automatic using a STA/LTA algorithm, **Manual Classification**
- **Three main classes :** **Volcano-Tectonic** (VT) events (78 % of the dataset), **Long-Period** (LP) events (2%) and **Nested** (20%) events.

Method



Features	Ref.	Features	Ref.
Statistic Features		**Shape descriptors features**	
Length	1	Rate of attack	19
Mean	2	Rate of decay	20
Standard deviation	3	min/mean and max/mean	21, 22
Skewness	4	Energy descriptors :	
Kurtosis	5	Signal Energy	23
i of central energy	6	maximum	24
RMS bandwidth	7	average	25
Mean skewness	8	standard deviation	26
Mean kurtosis	9	skewness	27
		kurtosis	28
Entropy Features		Specific values :	
		min	29
Shannon entropy[a]	10 to 12	max	30
Rényi entropy[b]	13 to 18	i of min	31
		i of max	32
		threshold crossing rate	33
		silence ratio	34

Feature set used to represent the signals [3]

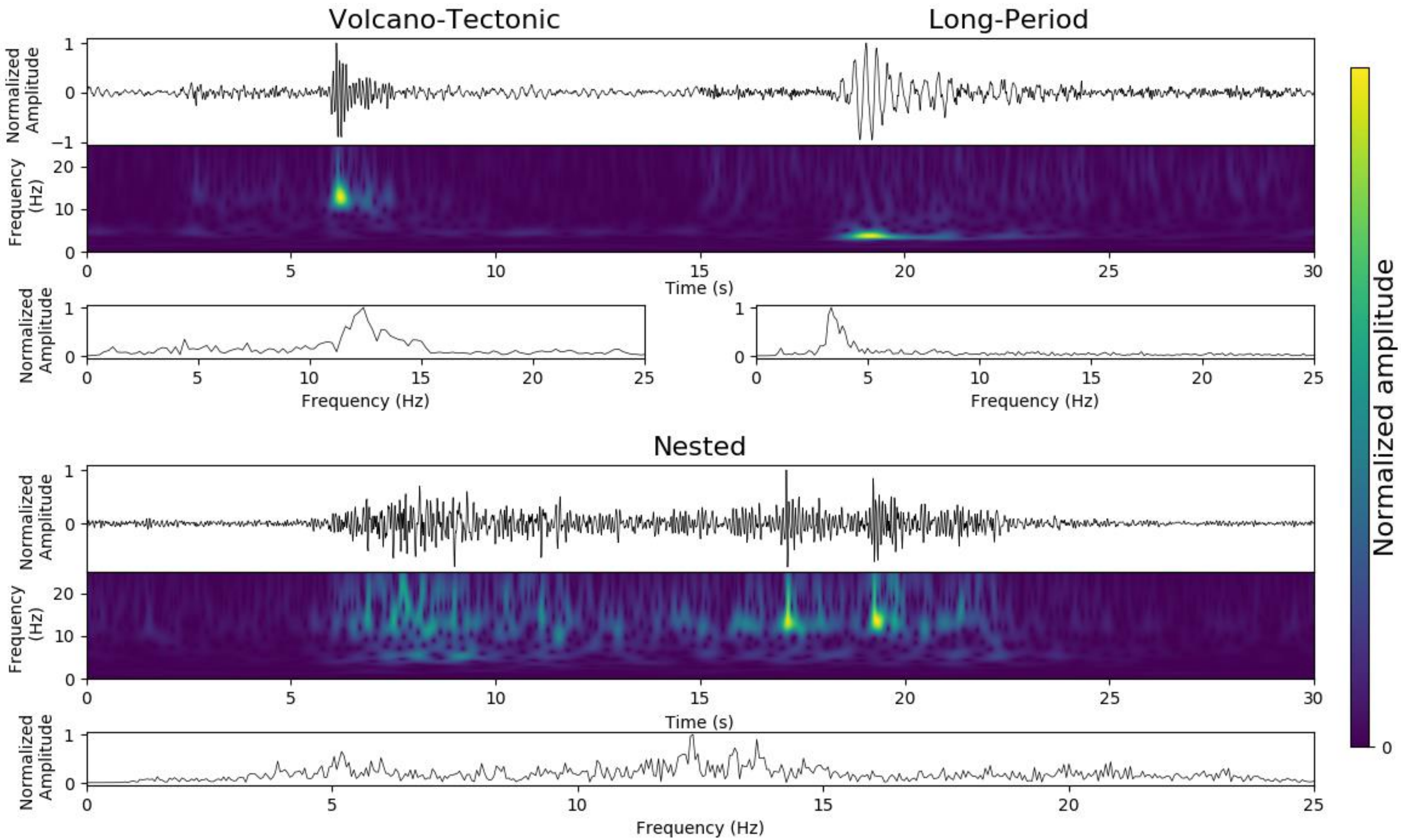
Metrics

$$Accuracy = \frac{\#\ Good\ Prediction\ Class_i}{\#\ Total\ True\ Class_i}$$

$$Precision = \frac{\#\ Good\ Prediction\ Class_i}{\#\ Total\ Predicted\ Class_i}$$

$$Overall\ Accuracy = \frac{\#\ Good\ Prediction}{\#\ Total\ Event\ Test\ Dataset}$$

Observatory Classes



Example of waveform filtered between 0.8-25 Hz, Spectrogram and Fourier Spectrum of Volcano-Tectonic events (top-left), Long-Period events (top-right) and Nested events (bottom) recorded at the station TAG on the vertical component

Observatory Classification Performance

Classification using the data recorded by TAG station and labelled in 3 classes by the observatory

- **Volcano-Tectonic events :** High frequency content (between 5 and 20 Hz). A characteristic peak observed between 10 and 15 Hz for La Soufrière. Very impulsive P wave arrival. Brittle failure events associated to stress changes due to magma movements [4].

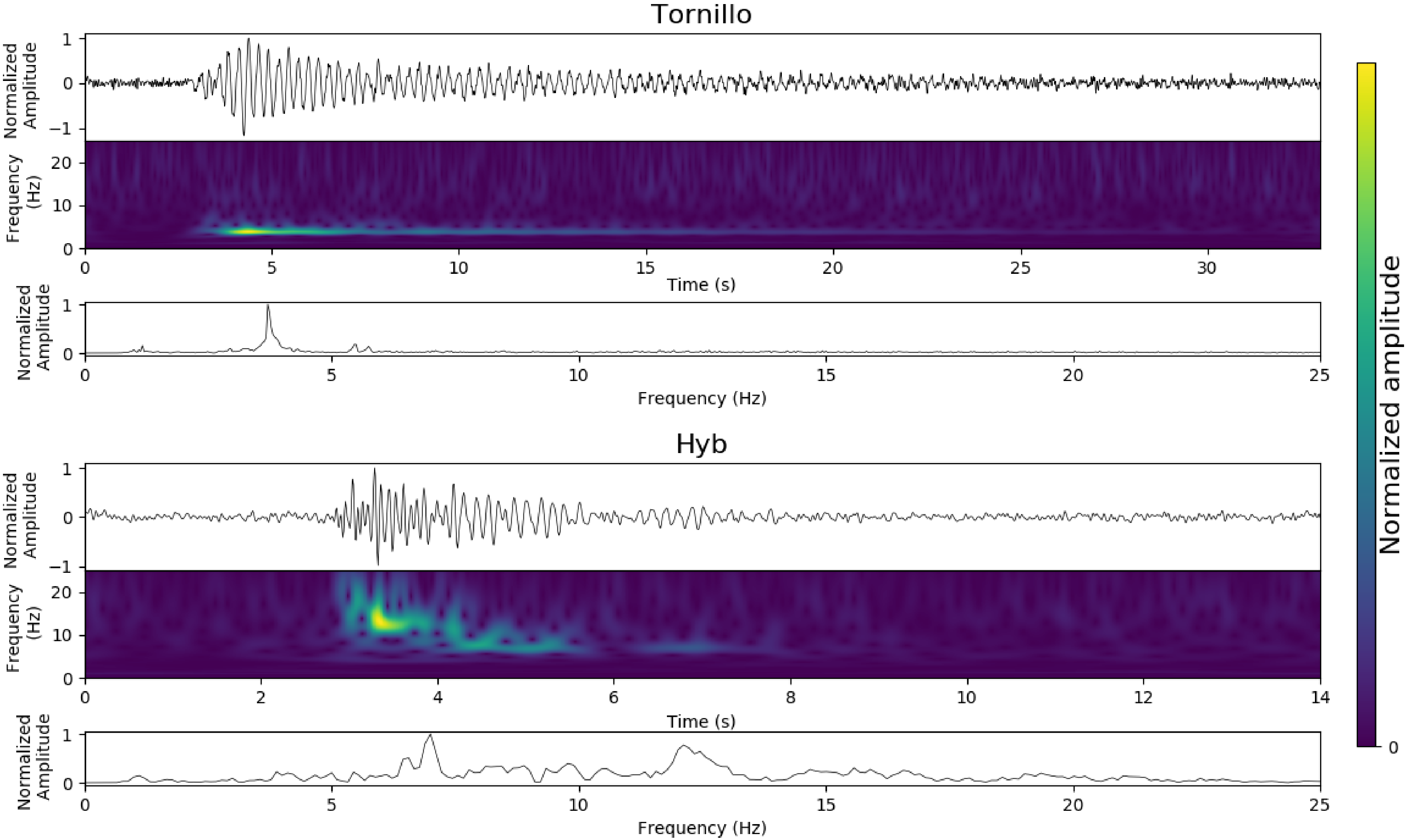
- **Long-Period events :** Fairly narrow spectral content around 4 Hz. Waveforms characterized by an emerging arrival of P waves, the S phase is not identifiable. Related to resonances of fractures, dykes, conduits or cavities during propagation of fluids (magmatic or hydrothermal) [5][6].

- **Nested events :** Small seismic packets in which events appears in the coda of each other. Not concomitant or precursor to a particular phenomenon [2]. Consist in a sequence of several seismic events with very short inter-times, with very often >6 seismic events in a short sequence (10s) [2][6]. Frequency content is pretty broadband (5-20 Hz) but most of the energy is in the same band as VT events. Source process not well understood. Events specific to La Soufrière volcano.

		Predicted Class			Accuracy (%) :
		Nested	LP	VT	
True Class	Nested	32	4	72	29,63
	LP	6	22	48	28,95
	VT	16	4	251	93,25
Precision (%) :		59,26	73,33	67,65	
Overall Accuracy (%) :				72,67 +/- 1,01	

Confusion matrix with a rate of 50-50% between the train and test dataset, mean score obtained after 10 trials

New Classes



Example of Waveform filtered between 0.8-25 Hz, Spectrogram and Fourier Spectrum of Tornillo events (left) and Hybrid events (right) recorded at the station TAG on the vertical component

Refined Classification Performances

		New Class					
		VT	LP	Nested	Tor	Hyb	Total
Old Class	VT	364	39	1	0	138	542
	LP	15	36	1	26	8	86
	Nested	125	22	28	2	40	217
	Total	504	97	30	28	186	845

New classes assigned after reviewing

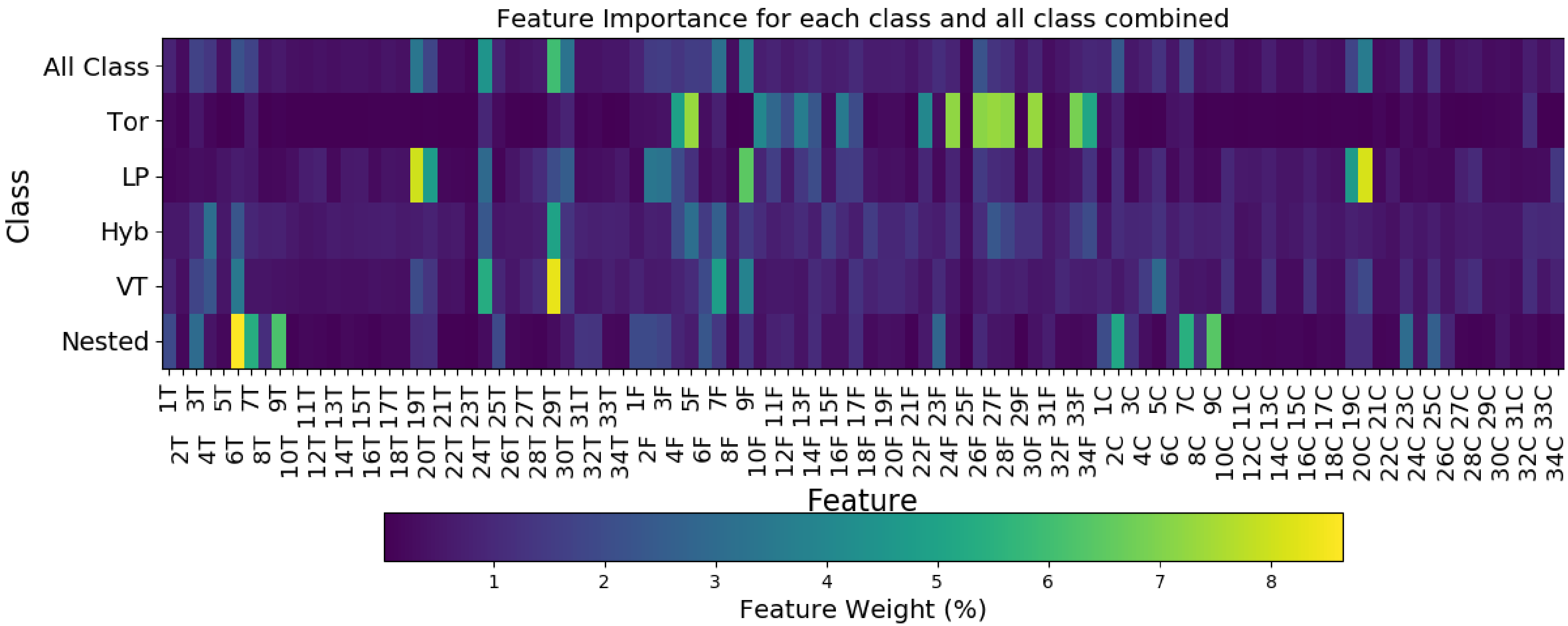
		Predicted Class					Accuracy (%) :
		Nested	Hyb	LP	Tor	VT	
True Class	Nested	14	0	0	0	1	93,33
	Hyb	0	59	7	0	26	64,13
	LP	0	4	42	1	2	85,71
	Tor	0	0	1	13	0	92,86
	VT	1	22	1	0	227	90,44
Precision (%) :		93,33	69,41	82,35	92,86	88,67	
Overall Accuracy (%) :				83,95 +/- 1,50			

- **<u>Hybrid events :</u>** Characteristics of both VT and LP events. High frequency impulsive arrival between 10-20 Hz. Coda dominated by lower frequency waves around 5 Hz. Halfway between VT and LP with fragile fracturing processes producing high frequencies and then the propagation of fluid responsible for resonance phenomena producing low frequencies [7]. Clear continuum between LP and hybrid [8]. Simple fracturing process with a very slow rupture velocity [9].

- **<u>Tornillo events :</u>** or monochromatic long-period events are a subcategory of LPs. Emerging wave arrival, a duration of a few tens of seconds, an almost sinusoidal signal and a coda which decreases very slowly and almost linearly. Characteristic peak around 4 Hz. Resonance last longer than the LP. Self-oscillations of fluid filling a cavity. [10][11]

Confusion matrix with a rate of 50-50% between the train and test dataset, mean score obtained after 10 trials

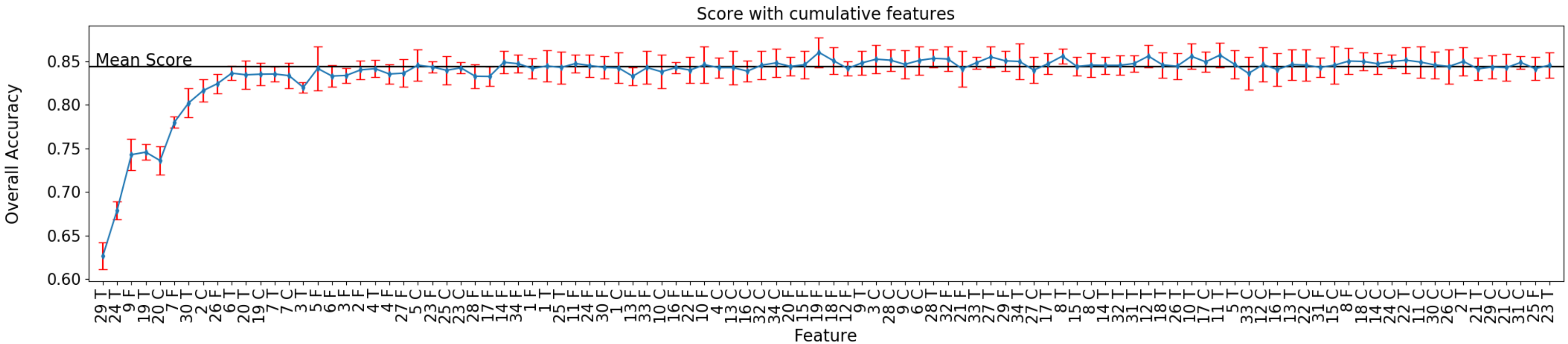
Features Importance



Feature importance *in percentage for the classification with* ***all the classes*** *and classification* ***one class vs other classes;*** *the numbers of the labels refer to the feature table, the letters to the representation domain (T : time, F : frequency, C : quefrency)*

Best features depend on the class we want to recognize

Cumulative Features Performance



Mean overall accuracies with the n most important features cumulated; Mean score obtained after 10 trials; Black line shows the mean score obtained with 102 features after 10 trials

18 features are enough to obtain mean score obtained with 104 features

Conclusion an Prospect

- Good recognition **increased from 72 % to 84 % after data class reviewing**, Machine learning helps to build a robust catalogue
- **Able to recognize two classes hardly distinguishable by OVSG** during daily data analysis protocole
- Poor **Hybrid** recognition suggests that this class is a **continuum between LP and VT** events
- The **most important features depend on the class** we want to recognize
- A **well chosen subset of features** (18/102) **is sufficient** to obtain substantially identical scores than with the whole feature set

- Use the Guadeloupe seismic network to make a **multi-station analysis**
- Feature exploration on another volcano
- **Unsupervised classification** to see how machine learning can discriminate the different signals
- **Implement** the model in observatory for real-time monitoring

References

- [1] Jessop et al., 2019, hal-02395880
- [2] Moretti et al., 2020, 10.1016/j.jvolgeores.2020.106769
- [3] Malfante et al.,2018, **10.1029/2018JB015470**
- [4] Chouet et al., 2013, 10.1016/j.jvolgeores.2012.11.013
- [5] Chouet et al., 1988, **10.1029/JB093iB05p04375**
- [6] Ucciani, 2015
- [7] Lahr et al., 1994, 10.1016/0377-0273(94)90031-0
- [8] Neuberg et al., 2000, 10.1016/S0377-0273(00)00169-4
- [9] Harrington et al., 2007, **10.1029/2006GL028714**
- [10] Gomez et al., 1997, 10.1016/S0377-0273(96)00093-5
- [11] Konstantinou et al. , 2015, 10.1016/j.pepi.2014.10.014