FEBUS OPTICS Distributed Fiber Sensing

Integration of Machine Learning on Distributed Acoustic Sensing surveys

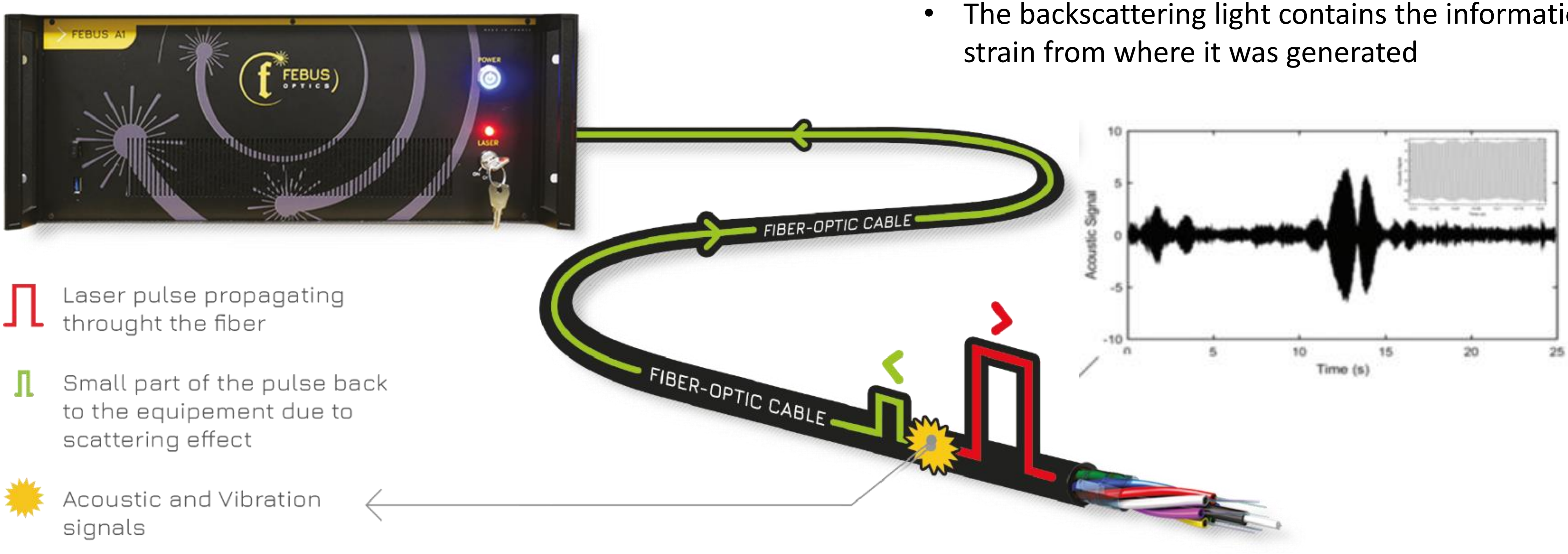
Camille Jestin, Clément Hibert, Camille Huynh, Gaëtan Calbris & Vincent Lanticq

EGU General Assembly 2020

Introduction

DAS principle

- Scattering effect occurs everywhere along the fiber
- The backscattering light contains the information of strain from where it was generated

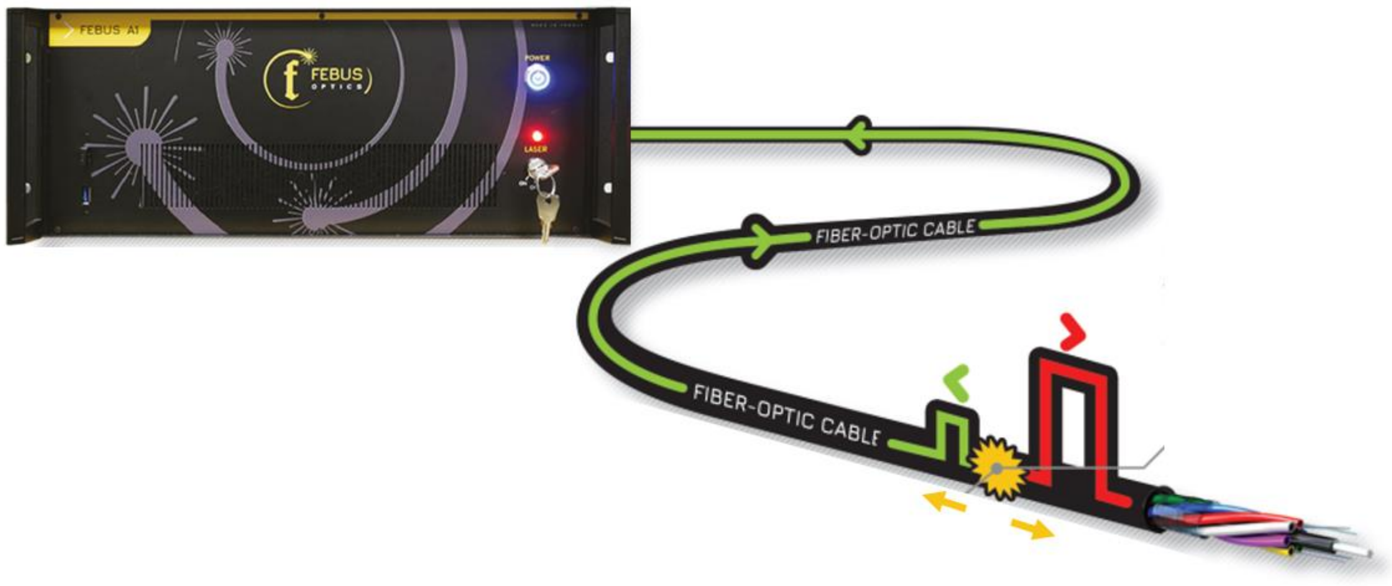


Laser pulse propagating throught the fiber

Small part of the pulse back to the equipement due to scattering effect

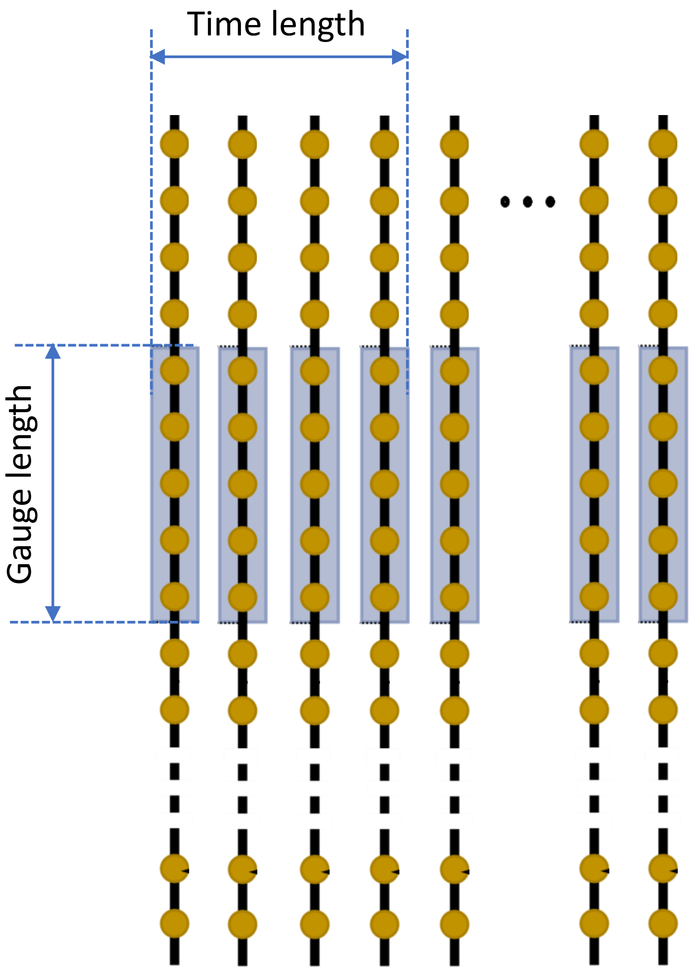
Acoustic and Vibration signals

DAS principle: Strain acquisition



- Optical phase directly related to strain applied to the fiber core over a gauge length
- Time derivative of the strain -> strain rate
- Strain-rate unit : nm/m/s

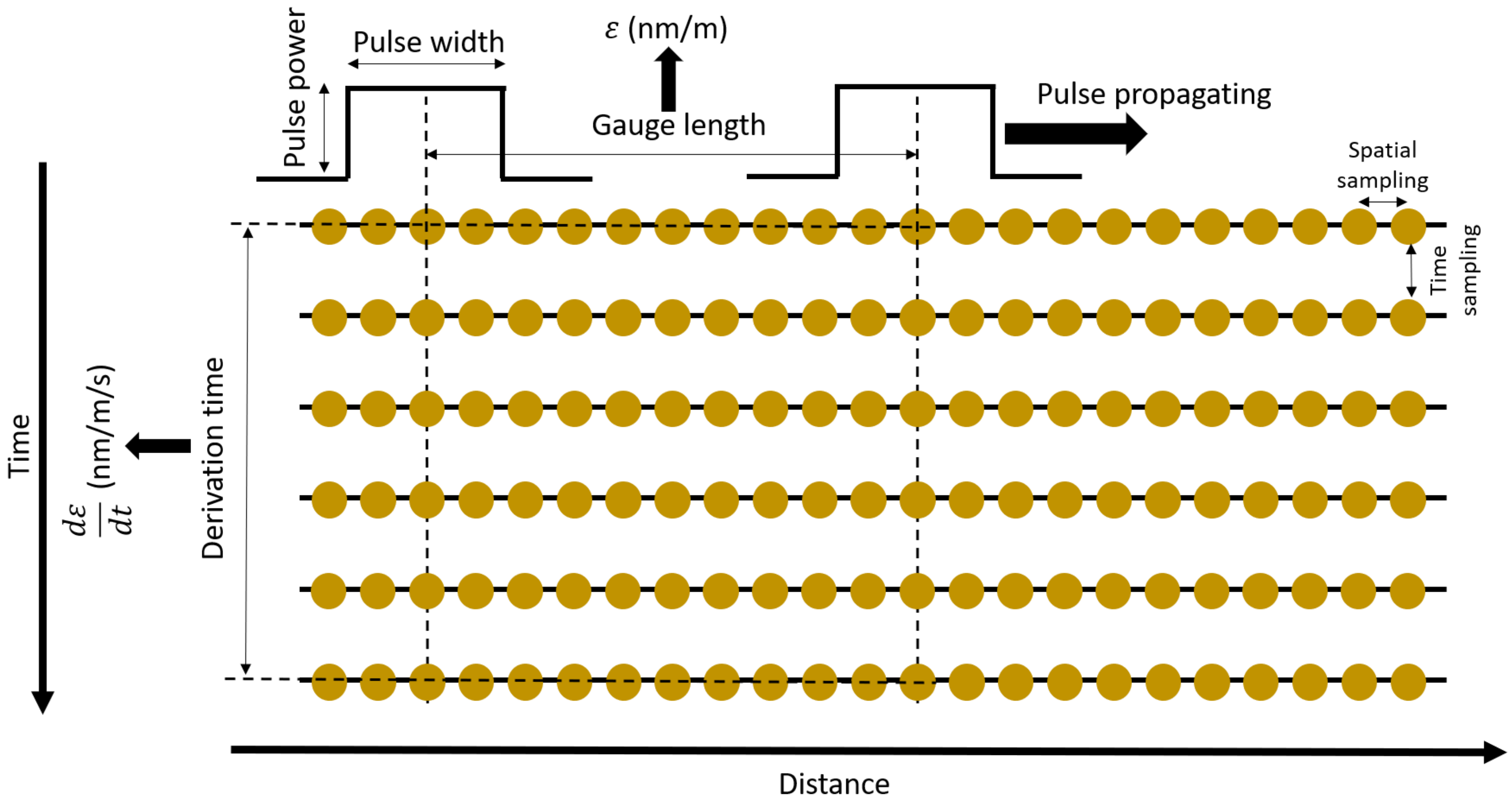
$$\Delta\phi = \frac{4\pi n G \xi}{\lambda \varepsilon}$$



DAS principle: Acquisition parameters

- Parameters to adjust :
  - Fiber distance
  - Optical power
  - Pulse width
  - Pulse rate frequency
  - Spatial sampling resolution

  - *Gauge length*
  - *Derivation time*



Machine Learning applied to DAS surveys

Context of the study

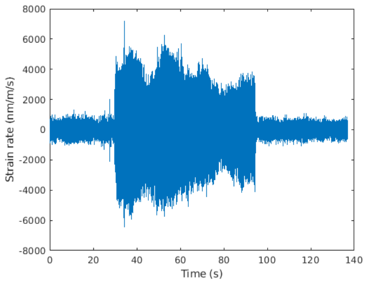
Pipeline monitoring for intrusion detection:

- Third party works detection and location using DAS is commonly applied in different contexts
- Challenge in identifying the origin of the signal:
  - Necessity of pattern recognition for relevant alarm.
  - Source and amplitude analysis for determining the threat at the pipeline neighbourhood.
  - The source identification must be fast, accurate and robust.
  - For its application to DAS data, the used method must be able to handle a big amount of data.

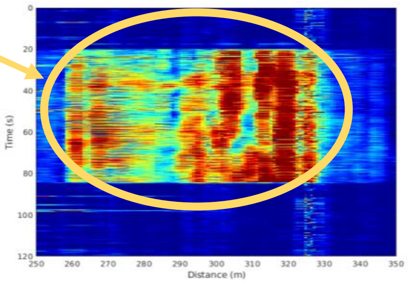
Solution: A Machine Learning algorithm enabling Classification of patterns before the alert release

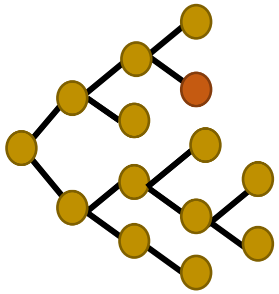
The data processing chain

Detection of signal of potential interest → **Machine Learning with Random Forest algorithm** → Signals Classification → Decision of alert release

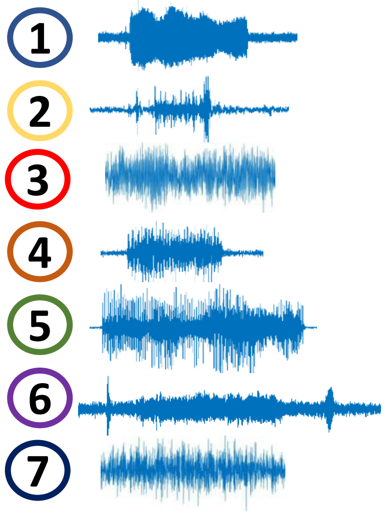


Spectral content between 5Hz and 95Hz





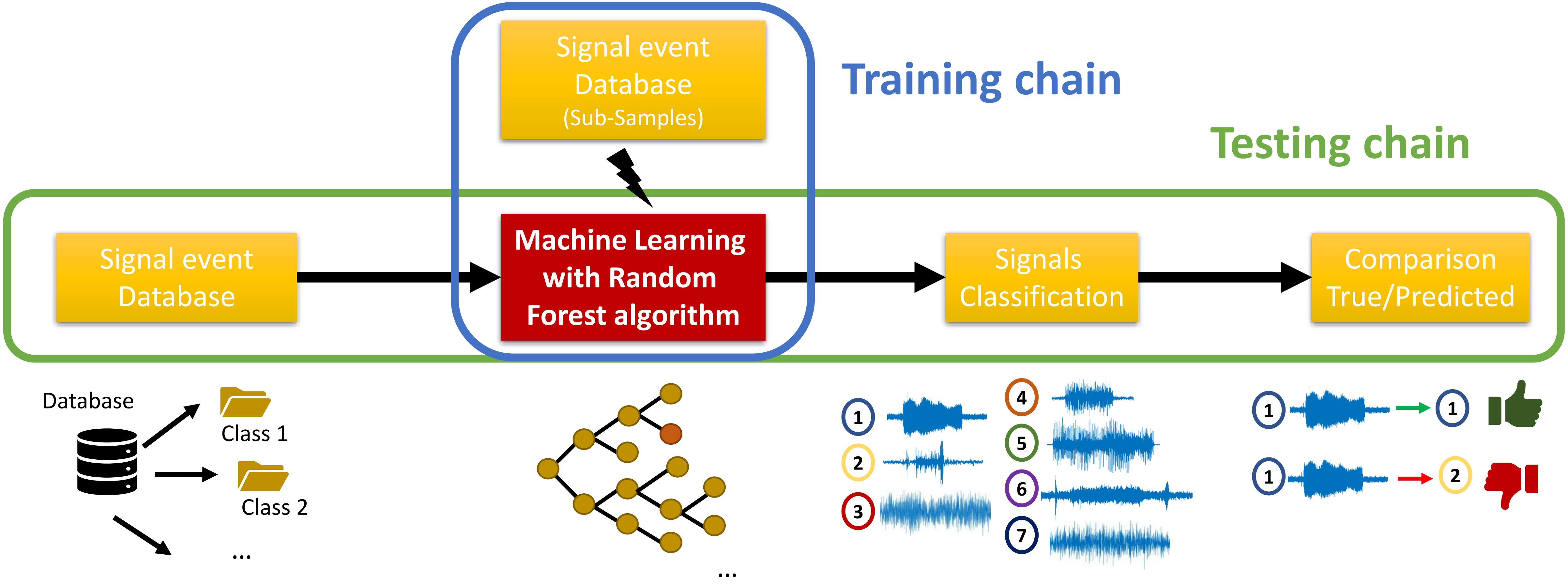
...



IF
Identified class = Risk
AND
Corresponding energy band >threshold:
ALERT

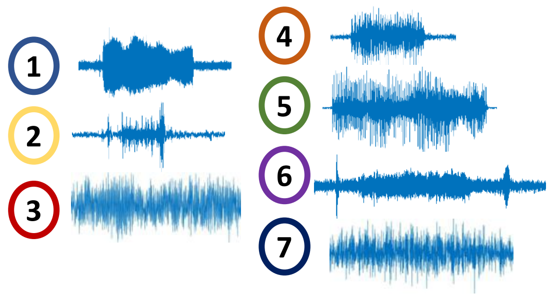
Machine Learning applied to DAS surveys

The training and testing chains



Test of the use of **the supervised classifier named *Random Forest* algorithm,** an ensemble learning method based on the use of decision trees

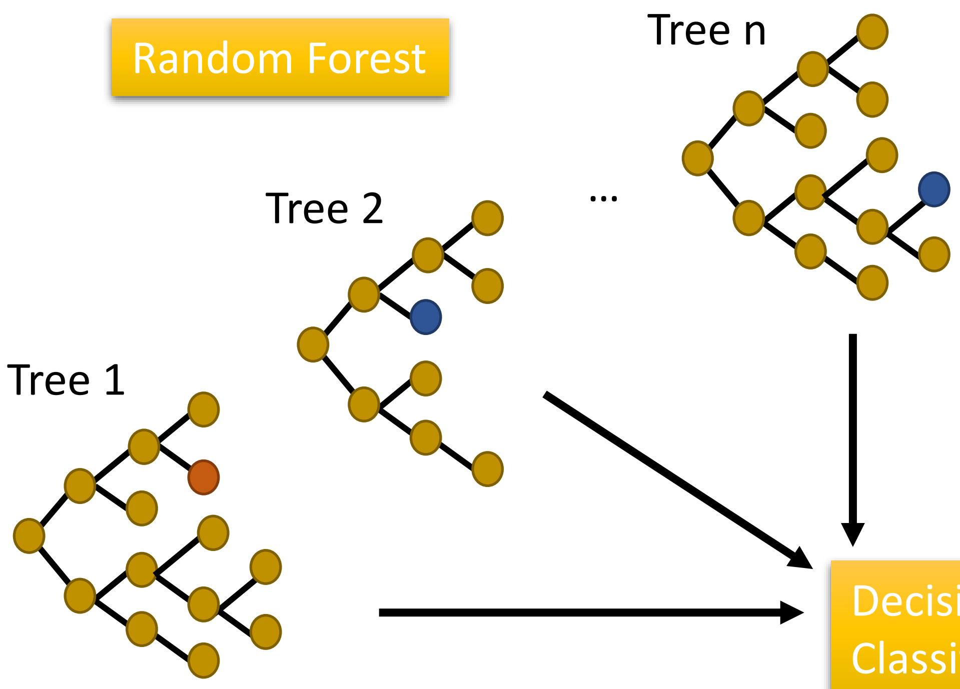
Training set:
Class determination



Discriminating attributes:

- **Duration**
- **skewness**
- **f_{max}**
- **Kurtosis**
- **Spectral properties**
- **50+ features**

Random Forest



Decision = 1
Classification

Machine Learning: Use of tens of attributes

Waveform Attributes: 23

1. Duration
2. Max/Mean ratio
3. Max/Median ratio
4. Ascending/Descending time ratio
5. Kurtosis of raw signal
6. Kurtosis of signal envelop
7. Skewness of raw signal
8. Skewness of signal envelop
9. Number of peaks in autocorrelation function
10. Energy in 1st third part of autocorrelation function
11. Energy in remaining part of autocorrelation function
12. Ratio of 11 and 10

13-17. Energy of the signal filtered in 5-10Hz, 10-30Hz, 30-50Hz, 50-75Hz and 75-99Hz

18-22. Kurtosis of the signal filtered in 5-10Hz, 10-30Hz, 30-50Hz, 50-75Hz and 75-99Hz

23. RMS between decreasing part of the signal and $I(t) = Y_{max} - \frac{Y_{max}}{t_f - t_{max}} t$

Spectral Attributes: 17

24. Mean of the Discrete Fourier Transform (DFT)
25. Max of the DFT
26. Frequency at the maximum DFT
27. Frequency at the centroid
28. Central frequency of the 1st quartile
29. Central frequency of the 3rd quartile
30. Median of the normalized DFT
31. Variance of the normalized DFT
32. Number of peaks in normalized DFT
33. Number of peaks (>0.75 DFT_{max})

34-37. Energy in [0, 1/4]Nyf, [1/4, 1/2]Nyf, [1/2, 3/4]Nyf, [3/4, 1]Nyf

38. Spectral centroid
39. Gyration radius
40. Spectral Centroid width

Pseudo-Spectrogram Attributes: 17

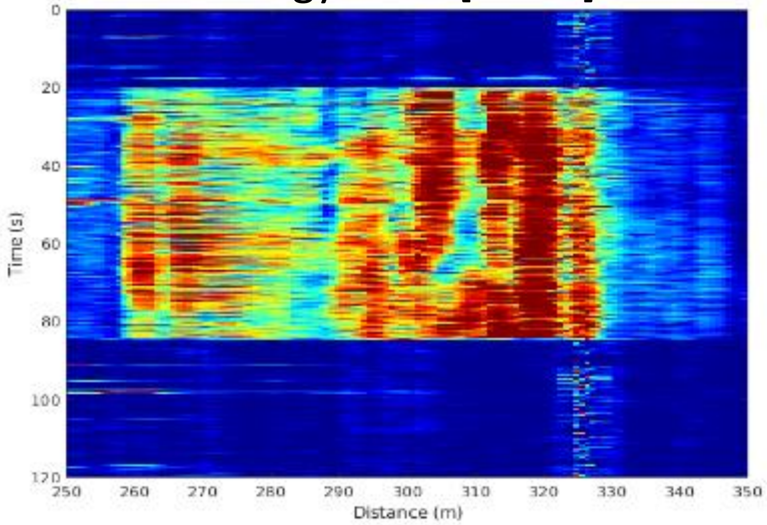
41. Kurtosis of max of all DFTs as a function of time
42. Kurtosis of median of all DFTs as function of time
43. Mean ratio between max and mean of all DFTs
44. Mean ratio between max and median of all DFTs
45. Number of peaks in the curve of temporal evolution of DFTs max frequency
46. Number of peaks in the curve of temporal evolution of DFTs mean frequency
47. Number of peaks in the curve of temporal evolution of DFTs median frequency
48. Ratio between 45 and 46
49. Ratio between 45 and 47
50. Mean distance between max and mean of all DFTs as function of time
51. Mean distance between max and median of all DFTs as function of time
52. Number of peaks in the curve of centroid frequency spectrum DFT
53. Number of peaks in the curve of max frequency spectrum DFT
54. Ratio between max frequency and centroid frequency DFTs
55. Mean distance between 1st quartile and median of all DFTs as function of time
56. Mean distance between 3rd quartile and median of all DFTs as function of time
57. Mean distance between 3rd and 1st quartiles of all DFTs as function of time

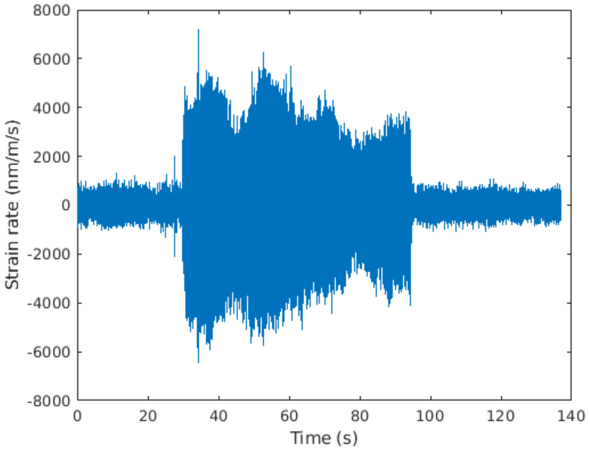
Hibert et. al, 2014, Provost et al., 2017; Hibert et al., 2017

Exemple of pipeline monitoring

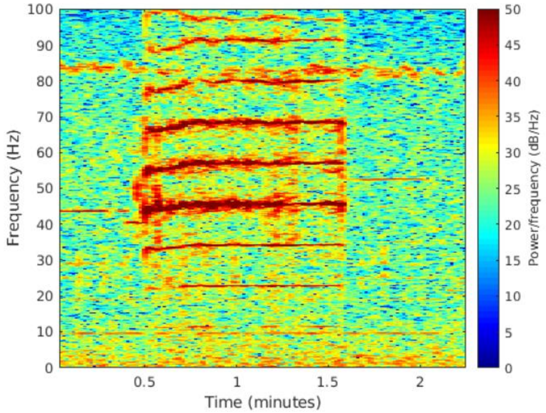
Manual Compactor

Energy band [5 - 95]Hz





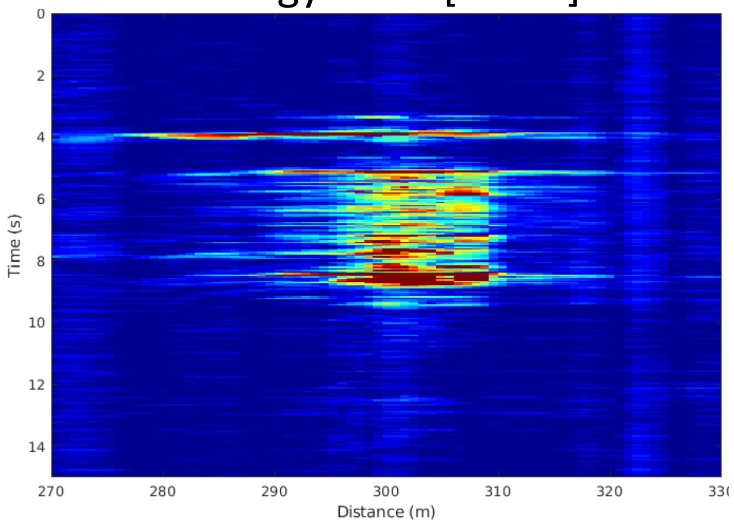
Spectrogram

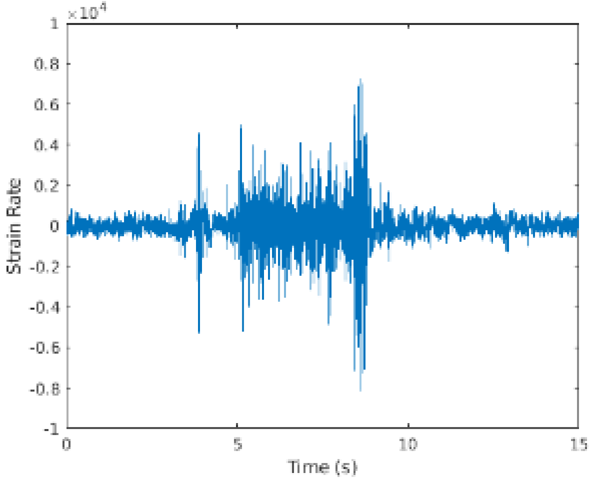


Pipeline monitoring: Third party works classification

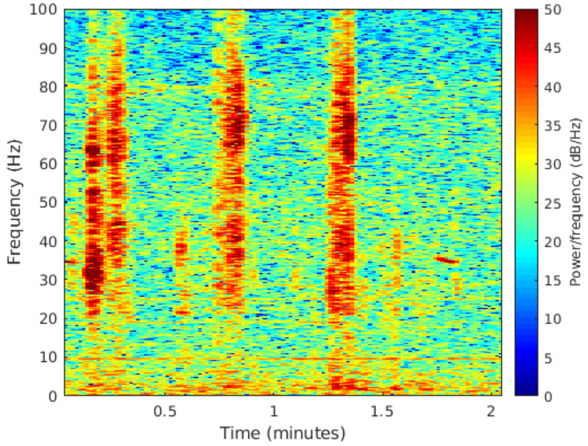
Excavation

Energy band [5 - 95]Hz



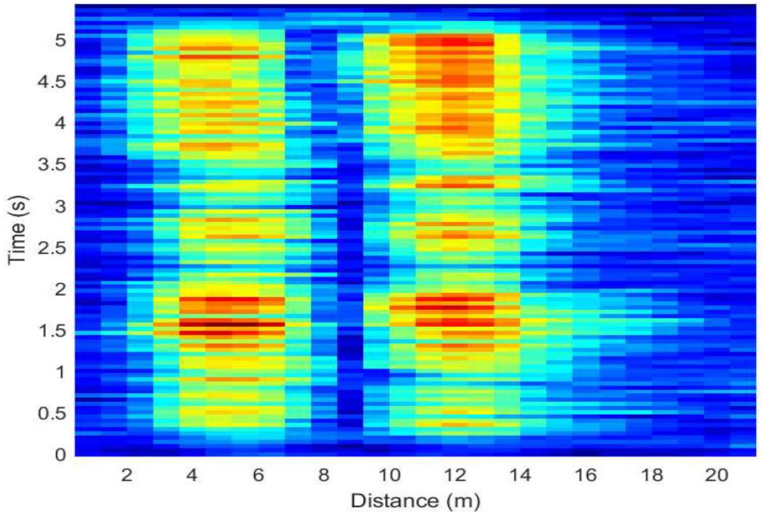


Spectrogram

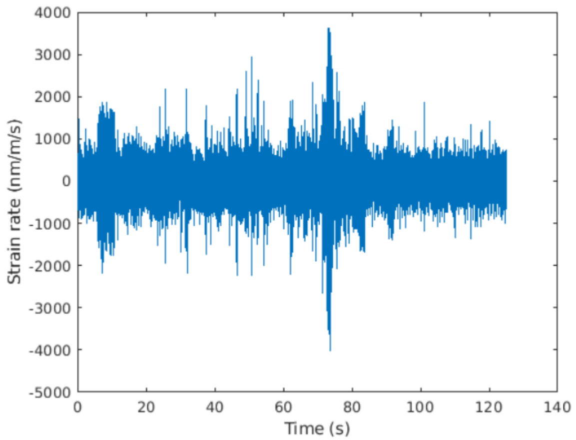


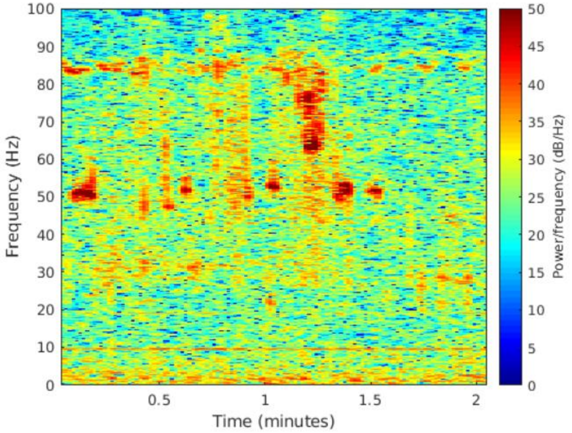
Drilling

Energy band [5 - 95]Hz



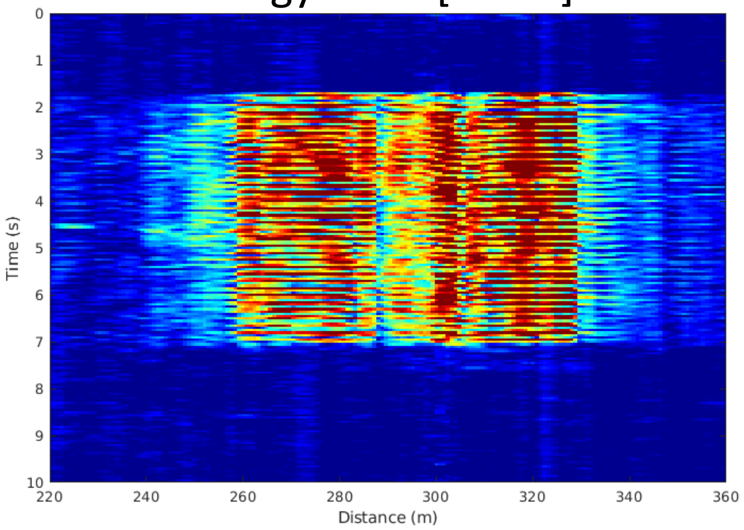
Spectrogram

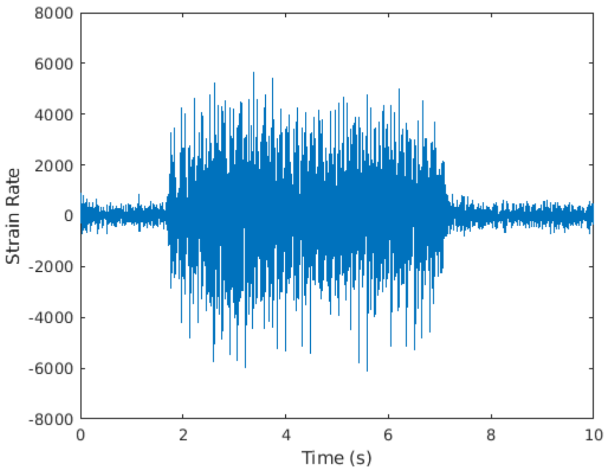




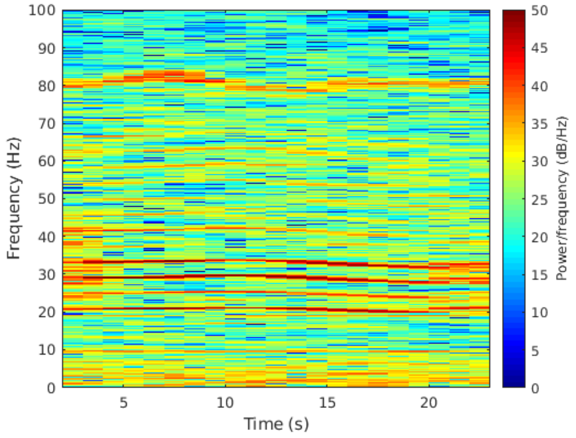
Jack hammer

Energy band [5 - 95]Hz

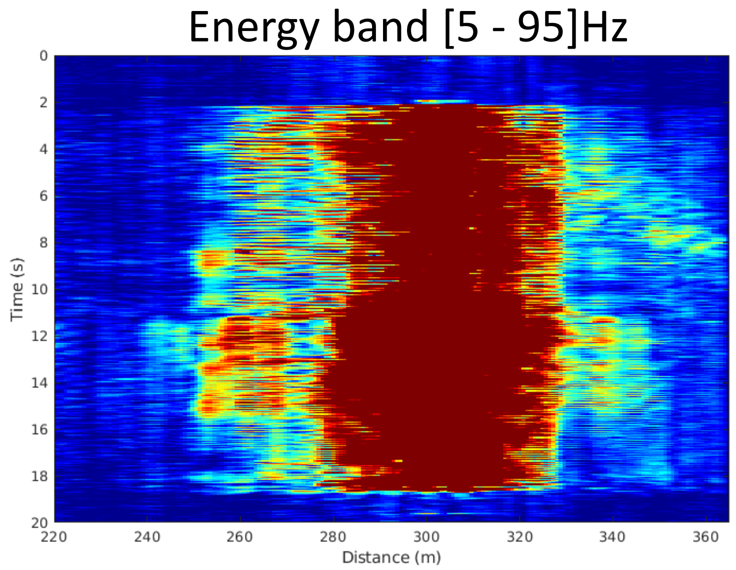


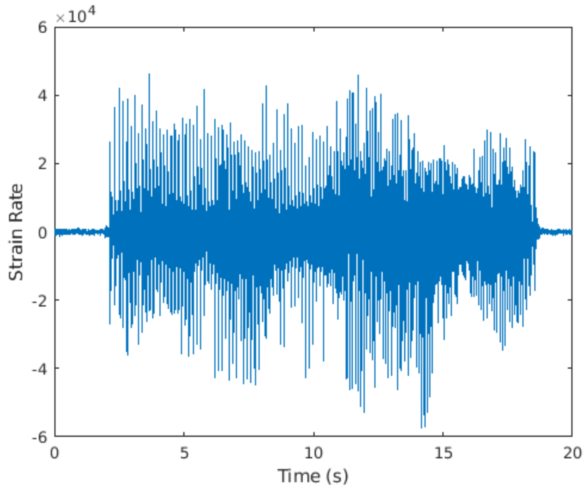


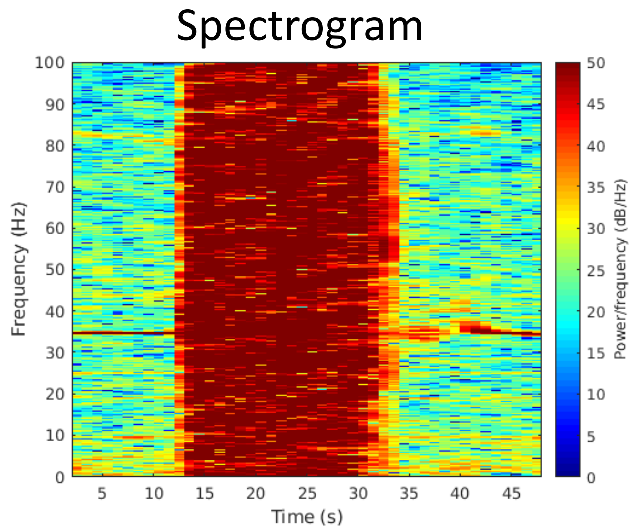
Spectrogram



Sheet pile

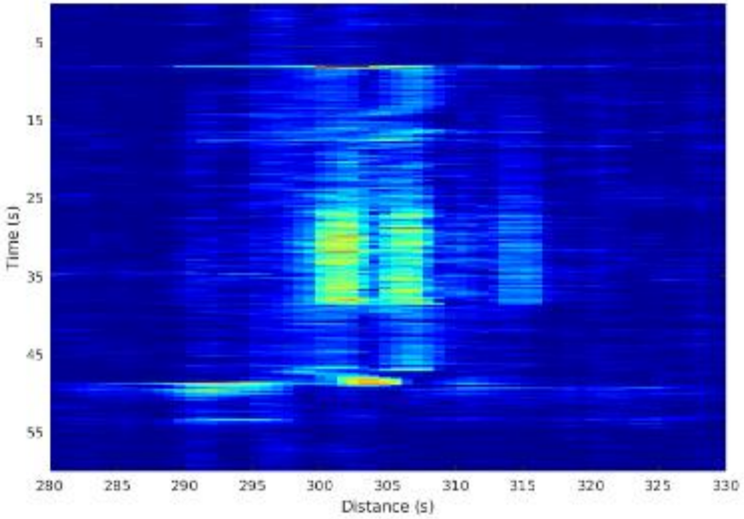

Energy band [5 - 95]Hz
Time (s)
Distance (m)

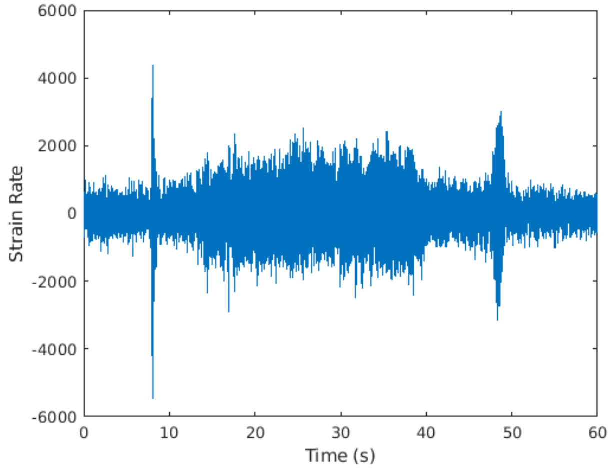

$\times 10^4$
Strain Rate
Time (s)


Spectrogram
Frequency (Hz)
Time (s)
Power/frequency (dB/Hz)

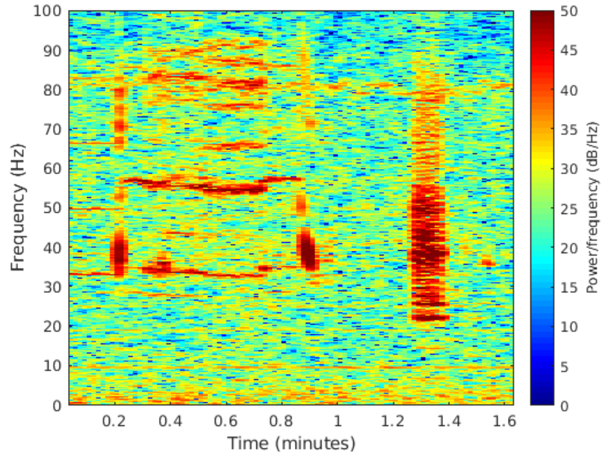
Circular saw

Energy band [5 - 95]Hz



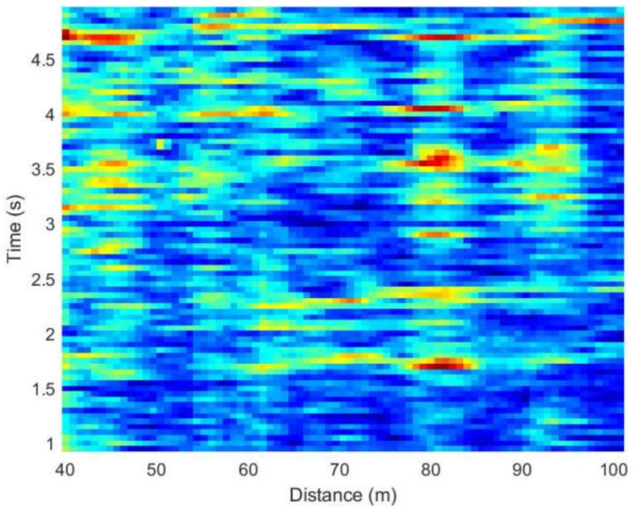


Spectrogram

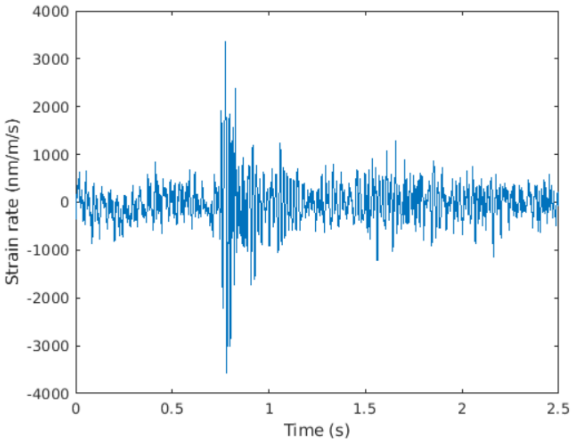


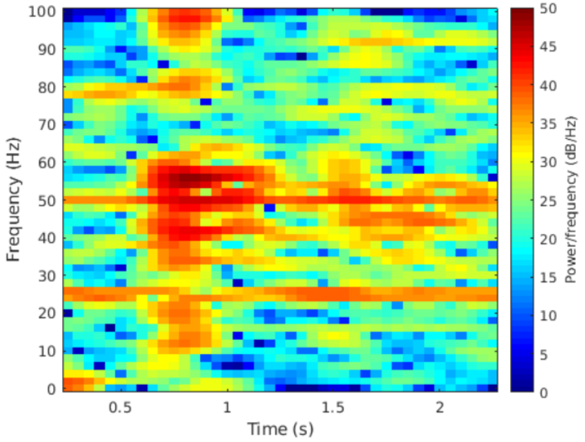
Transportations

Energy band [5 - 95]Hz



Spectrogram





Then:

- In this study, we work on 7 classes of event, numbered from 1 to 7.

- Because DAS acquisition can generate traces every few meters along fibres of tens of kilometres, **two methods are used for classification** using Random Forest algorithm:

    1. The first one is **signal** based: The algorithm is using each single trace/station for the signal classification.
    2. The second one is **event** based: A cluster of stations, identified as recording the same event, is used by the algorithm for the source signal classification. The majority of votes will release the final ID of the event.

- Three parameters are used to check the efficiency of the pattern ID using Machine Learning: Precision, Recall and Accuracy

Quality Control parameters

		Predicted	
		Negative	**Positive**
Actual	Negative	True negative	False positive
	Positive	False negative	True positive

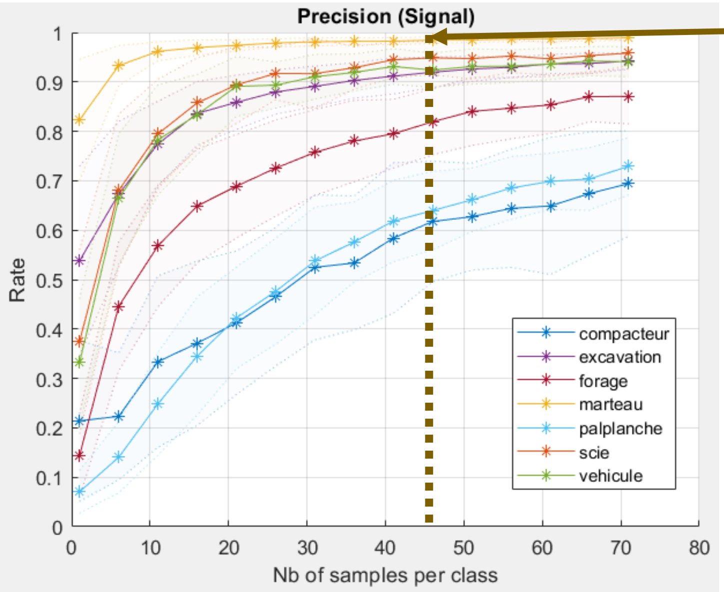
Accuracy $\frac{True\ positive+True\ negative}{Positive+Negative}$

Recall $\frac{True\ positive}{True\ positive+False\ negative}$

Precision $\frac{True\ positive}{True\ positive+False\ positive}$

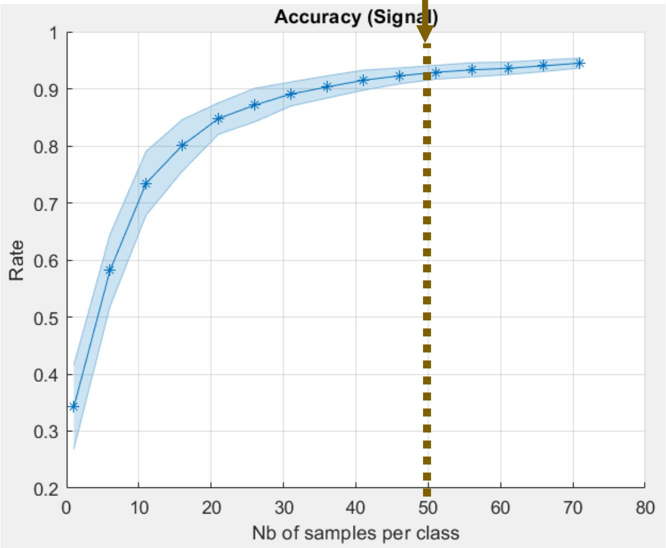
F1 Score $2 * \frac{Precision * Recall}{Precision+Recall}$

First approach: use of the same number of samples for each class

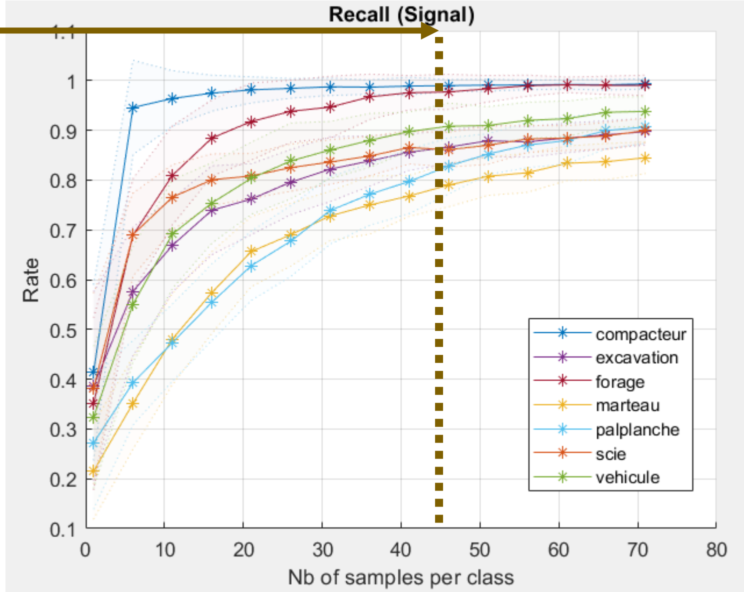


1 - False alarm rate for each class

The first classifier was trained using 50 samples of each class

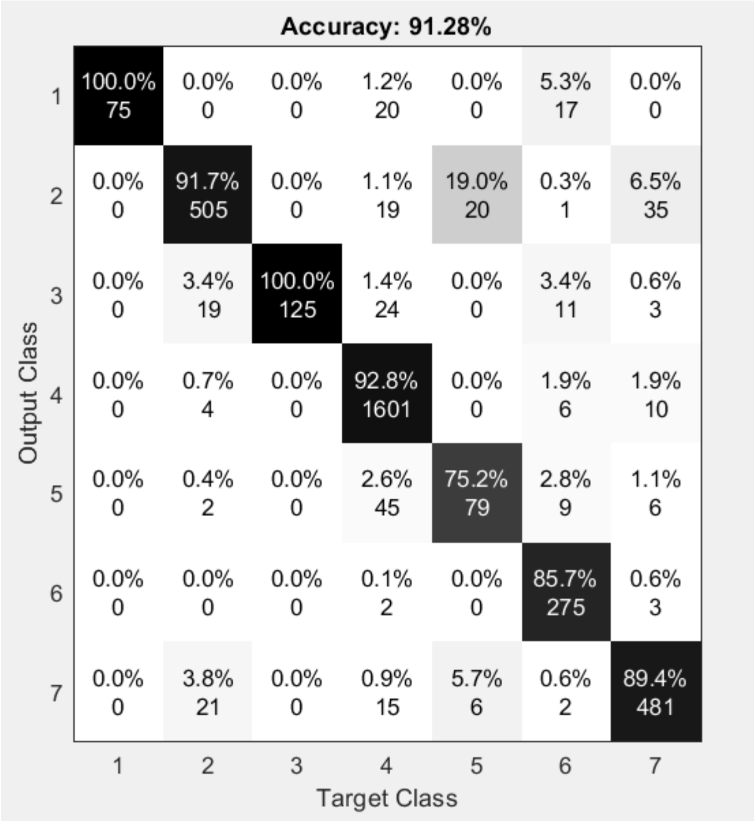


Global good classification rate

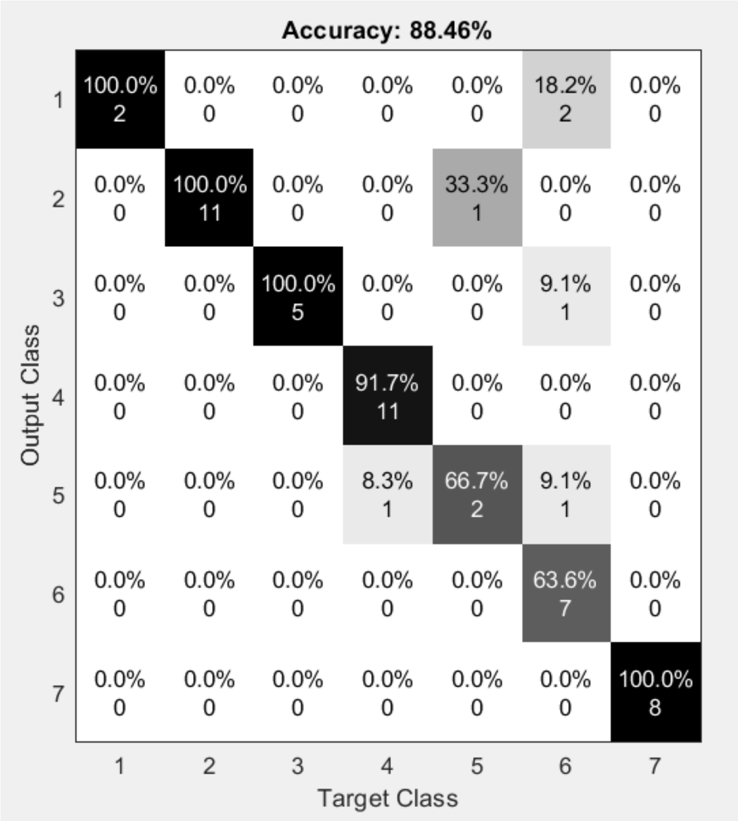


1 – Lack of detection rate for each class

First approach: use of the same number of samples for each class



Confusion matrix for signal



Confusion matrix for event

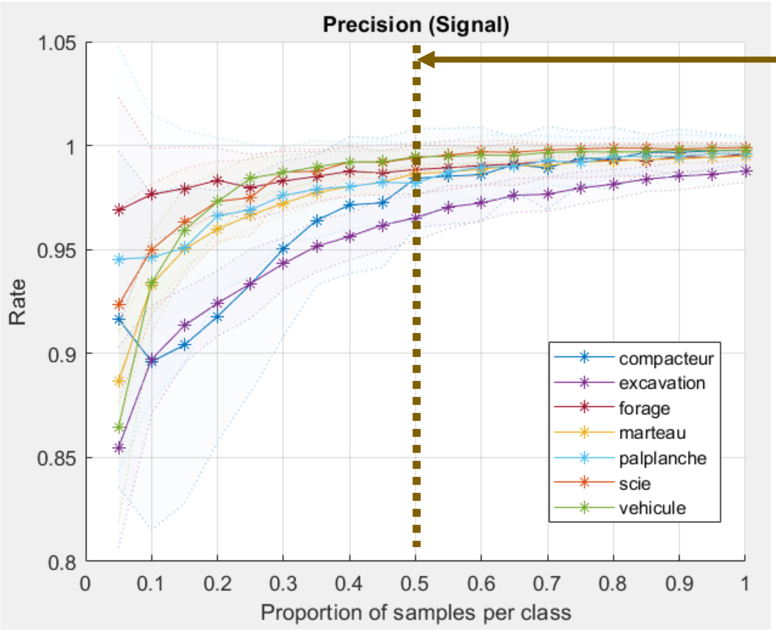
Classes:

1. Manual compactor (75)
2. Excavation (551)
3. Drilling (125)
4. Jack hammer (1726)
5. Palplanche (105)
6. Circular saw (321)
7. Transportation (538)

For all studied events:

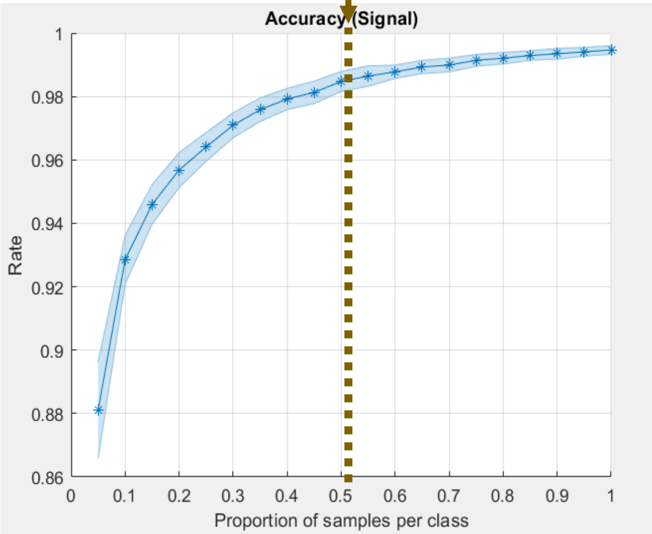
Classification with this algorithm is **88.46% correct** with an accuracy of 91.28%

Second approach: Training samples are taken proportional to their natural distribution occurrences

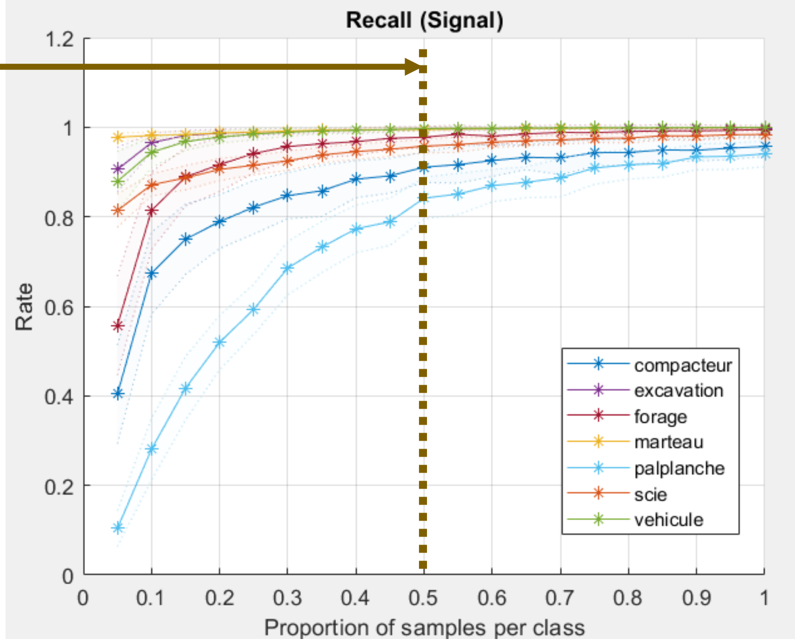


1 - *False alarm rate for each class*

Our classifier was trained using a half of the total dataset

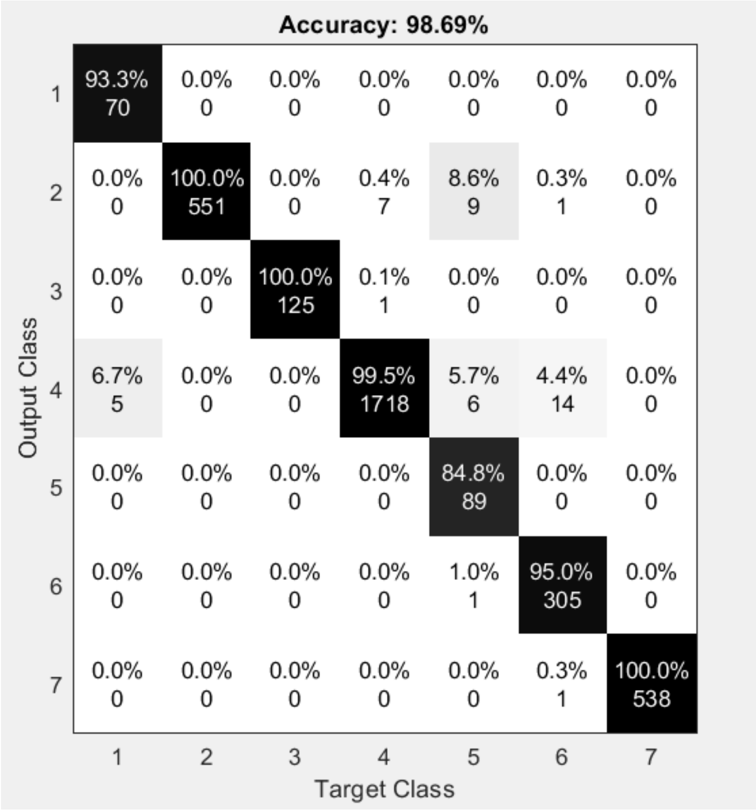


Global good classification rate

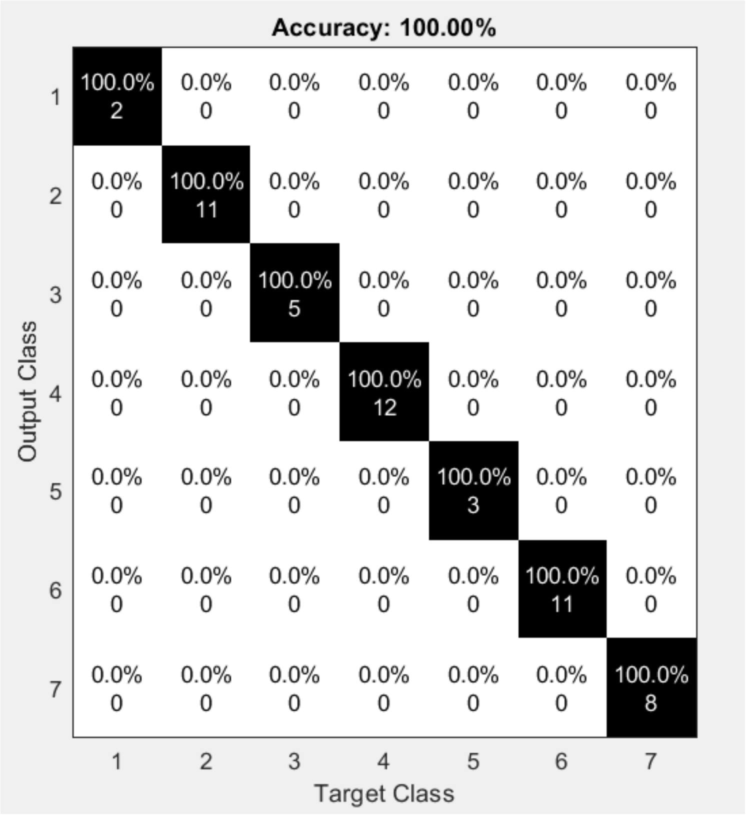


1 – *Lack of detection rate for each class*

Second approach: Training samples are taken proportional to their natural distribution occurrences



Confusion matrix for signal



Confusion matrix for event

Classes:

1. Manual compactor (75)
2. Excavation (551)
3. Drilling (125)
4. Jack hammer (1726)
5. Palplanche (105)
6. Circular saw (321)
7. Transportation (538)

For all studied events:
Classification with our algorithm is **100% correct** with an accuracy of 98.69%

Conclusions

- Random Forest algorithm appears to be relevant (fast and robust) for the classification of acoustic events recorded with DAS.

- Tests on other field sites are under process to demonstrate the efficiency of our Machine Learning method on different contexts.

- Different fields of application of this algorithm are possible: intrusion detection along pipelines, in perimeters, seismic event detection and classification (volcanoes, glaciers, etc.).

- Tests on data processing in flux for real-time event detection and classification are under process.

www.febus-optics.com

Technopole Helioparc - 2 av. Président Pierre Angot - 64000 Pau - France
+33 (0)5 24 36 45 82 - info@febus-optics.com