Underground River Monitoring from Seismic Waves with a Random Forest Algorithm

Anthony Abi Nader, Julie Albaric, Clément Hibert, Jean-Philippe Malet, Marc Steinmann

Underground River Monitoring from Seismic Waves with a Random Forest Algorithm

Anthony Abi Nader[1], Julie Albaric[1], Clément Hibert[2], Jean-Philippe Malet[2], Marc Steinmann[1]

[1] Laboratoire Chrono-environnement – Université Bourgogne Franche-Comté (UMR 6249)

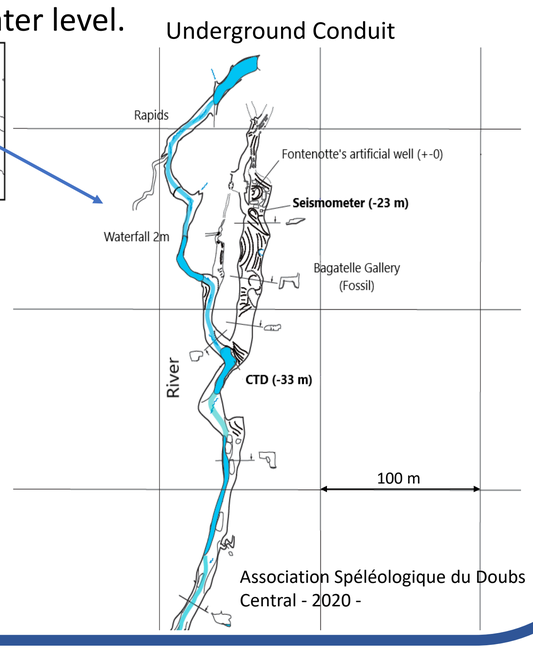
[2] Institut Terre et Environnement de Strasbourg – Université de Strasbourg (*UMR* 7063)

1- Introduction

Groundwater storages are usually inaccessible and therefore their surveillance can become complicated. As a complement to traditional methods, seismic noise analysis was suggested to monitor ground water storage[1]. In the following we present a machine learning based method, the **Random Forest Algorithm** [2], used in our case to detect similarities within seismic signals in order to find signals corresponding to **flooding**, and in a next step predict the underground river water level. This method shows promising outcomes for this kind of applications due to the strong signature of the water level on the noise.

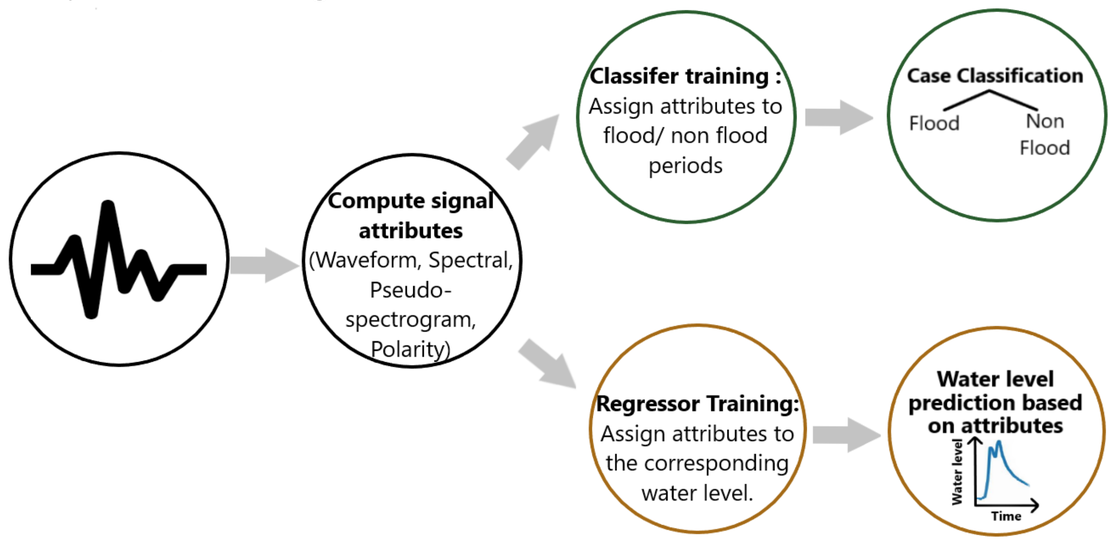
2- Site and Instrumentation

The studied site is the **Fourbanne karstic aquifer**, located in the Jura mountains and monitored since 2014[3]. The underground conduit is accessible through a drilled well and is instrumented by two 3C seismometer (at the surface and 20 m depth) and a hydrological probe giving the underground river water level.



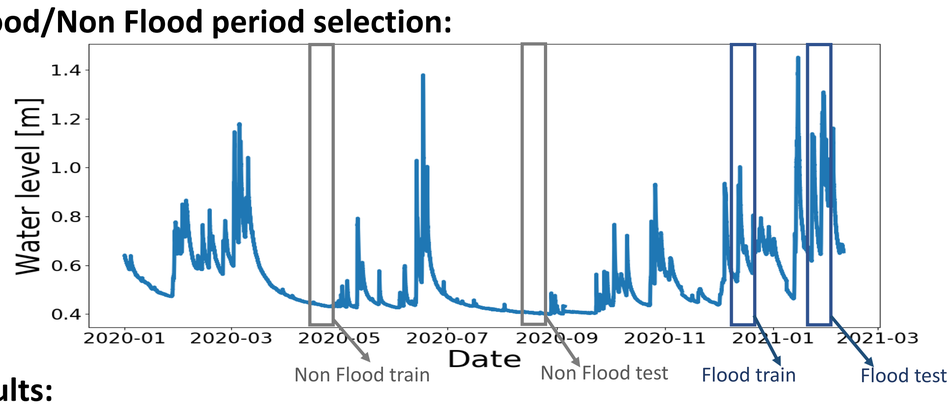
3- Method : Random Forest Algorithm

The method consists in the calculation of 60 attributes for the chosen dataset of seismic signals that are then used in the training for the computation of decision trees [4]. For the **Classification algorithm**, the training is done by assigning attributes a case (flood or non flood), for the **Regression algorithm**, the attributes are assigned to the corresponding water level. The objective is to only use seismic signal in order to detect floods and predict the underground river water level.



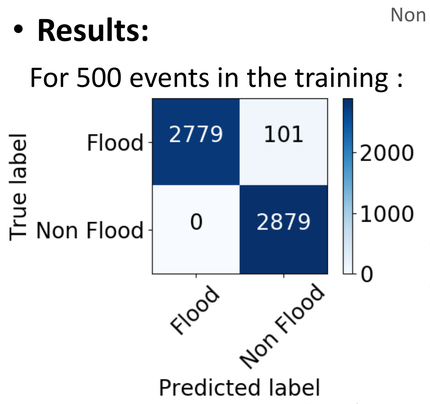
4.1- Results: Classification

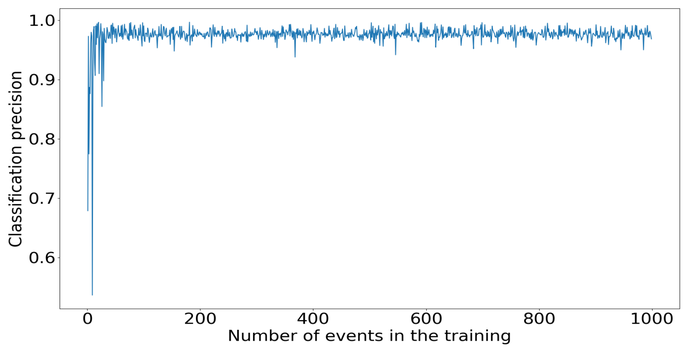
- **Objective:** Detect seismic signals corresponding to flooding.
- **Settings: -**Event= Set of 60 attributes calculated for a 5 minutes signal.
  -Number of events for the testing sets : 2879 (five-days signal sampled each five minutes with 50% overlap)
  -These images are the results obtained using signals from the station in depth.
- **Flood/Non Flood period selection:**



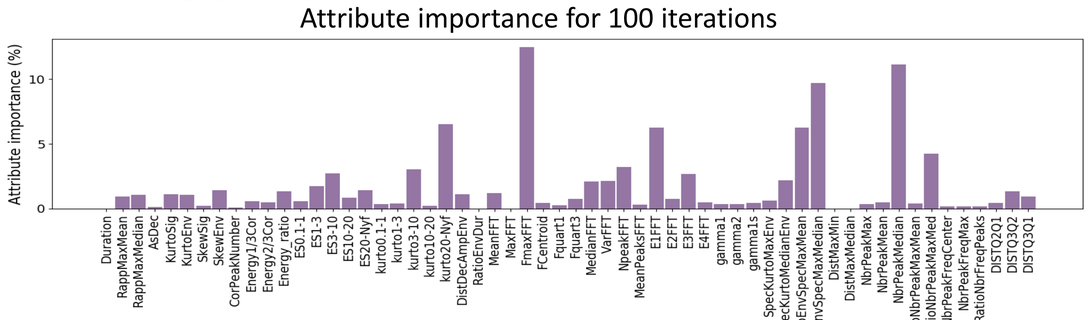
- **Results:**

For 500 events in the training :





Attribute importance for 100 iterations

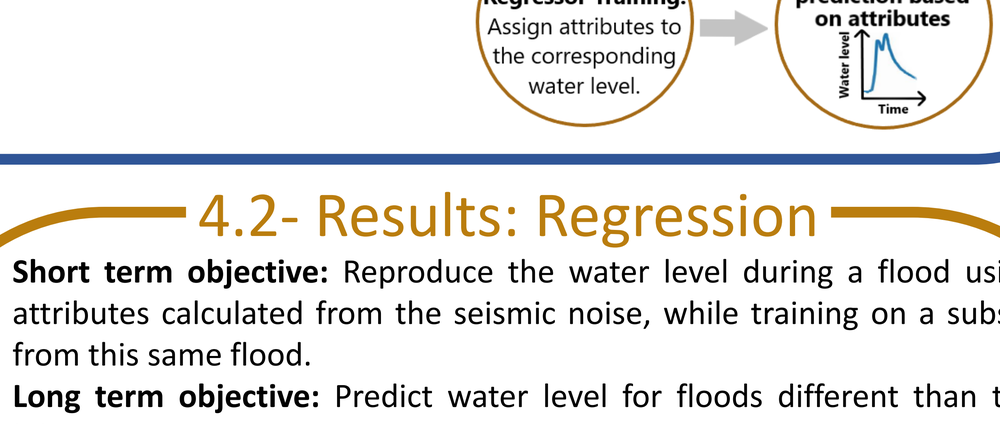


Relevant attributes: -FmaxFFT: Frequency at Max(FFT)
-RatioEnvSpecMaxMedian: Ratio Max DFT(t)/ Median DFT(t)
-NbrPeakMedian: Nbr peaks Median DFTs(t)

Station in depth: Flood detection achieved for all floods
Station at the surface: Flood detection achieved for large floods only.

4.2- Results: Regression

- **Short term objective:** Reproduce the water level during a flood using attributes calculated from the seismic noise, while training on a subset from this same flood.
  Long term objective: Predict water level for floods different than the training set.
- **Settings:** Number of events in total: 2880
  Number of events for training: 1000 (from the same flood)
- **Results:**



Water level accurately reproduced while training the algorithm on a subset from the same flood.
Water level prediction for other floods not achieved.

5- Conclusion

- The **Random Forest classifier** is used because it is a non demanding method with high performance when compared with other machine learning techniques. The algorithm is able to even **detect the smallest floods** with high accuracy using seismic signals from the station located in depth.
- For the **Random Forest Regressor**, the algorithm was able to **reproduce the hydrological signal** with precision while training on the same flood, further trials should be done in order to try to predict the underground river water level while training on attributes calculated during a flood and testing on another flood.

Acknowledgments & References

This work has been supported by the Region Bourgogne Franche-Comté. We acknowledge greatly the ASDC (Association Spéléologique du Doubs Central) for helping us to investigate and instrument the karst conduit. We acknowledge as well everyone that was implied in this project's achieval.

[1] Lecocq *et al.*, *Scientific Reports* volume 7, Article number: 14241 (2017)
[2] L. Beriman, Machine Learning volume 45, pages5–32 (2001)
[3] Cybèle Cholet, https://hal.archives-ouvertes.fr/tel-01617489v2 (2017)
[4] Hibert *et al.*, Journal of Volcanology and Geothermal Research Vol. 340 (2017)

Contact

Name: **Abi Nader Anthony**
Email: anthony.abi_nader@univ-fcomte.fr
Laboratoire Chrono-environnement
16 route de Gray, 25000 BESANÇON, France