

**Dejan ŠKRELIĆ**

Ekaterina BUROV

SVI lab, Saint-Gobain Research Paris / CNRS, 39 quai Lucien Lefranc, Aubervilliers, France

[dejan.skreltic@cnrs.fr](mailto:dejan.skreltic@cnrs.fr)

## Visual Exploration of Raman Spectra (Open Source Python Toolkit)

We are witnesses of the proliferation of Raman spectroscopy in wide range of areas from industry to academic research. However, regardless of domain, visual exploration of experimental data is almost always an indispensable step allowing us to gain insight of the processes at work in our sample.

In parallel, modern Raman spectrometers allow us to record an evergrowing quantity of spectra in rather short time (as maps for example), and therefore require performant tools for analysis of these big datasets ( $\sim 10^9$  intensity values). Having such amount of data at our disposal, one naturally turns to machine-learning algorithms to extract meaning. And arguably the most widespread modern machine learning tools use python for its front-end (scikit-learn, TensorFlow, PyTorch).

Furthermore, the growing need to ensure reproducibility of results in scientific research, incites us to provide the explicit and verifiable code for all the steps undertaken during the data analysis.

With all of the above in mind, we engaged in developing a python tool for intuitive and ergonomic exploration of hyperspectral images, irrespective of spectrometer provider. In addition, we also developed several interactive tools facilitating the choice of parameters for typical spectra processing steps (baseline correction, denoising, CR correction, etc).

The main part of this talk will consist of a live demonstration of some of the capabilities of the “SpectraPy” toolkit.

If the allocated time permits, we will briefly present a typical workflow we use for the deconvolution of Raman spectra, namely the sequential use of PCA, NMF and MCR (this will be further developed in an usecase example presented by Ekaterina BUROV).

### References

- [1] D. Hunter, "Matplotlib: A 2D Graphics Environment", Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, 2007..
- [2] Hoyer, S. & Hamman, J., (2017). xarray: N-D labeled Arrays and Datasets in Python. Journal of Open Research Software. 5(1), p.10. DOI: <https://doi.org/10.5334/jors.148>
- [3] Harris, C.R., Millman, K.J., van der Walt, S.J. et al. Array programming with NumPy. Nature 585, 357–362 (2020). DOI: 10.1038/s41586-020-2649-2.

### Figures

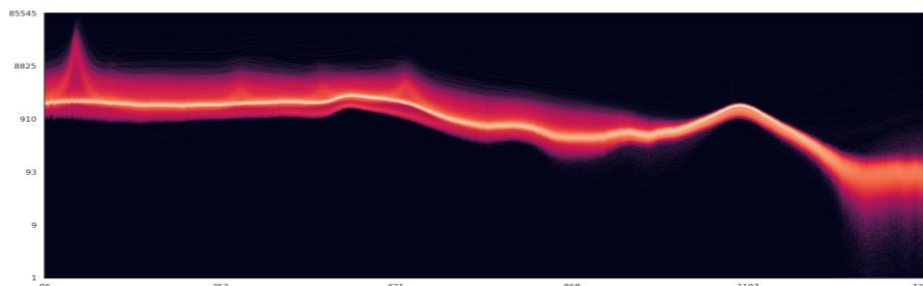


Figure 1: Histogram-like representation of ~40k spectra