RamanFest 2023
Paris | FRANCE

Thomas Bocklitz[1,2,3]

Shuxia Guo, Oleg Ryabchykov

[1] Institute of Computer Science, Faculty of Mathematics, Physics & Computer Science, University Bayreuth Universitaetsstraße 30, 95447 Bayreuth, Germany;

[2] Leibniz Institute of Photonic Technology (Leibniz-IPHT), Albert-Einstein-Straße 9, 07745 Jena, Germany;

[3] Institute of Physical Chemistry and Abbe Center of Photonics (IPC/ACP), Friedrich-Schiller-University, Helmholtzweg 4, D-07743 Jena, Germany

Thomas.bocklitz@uni-jena.de

Photonic Data Science:
Data pipelines for modeling of Raman effect related data

Raman spectroscopic techniques are increasingly used in various disciplines such as chemical analytics, life science and medicine. This increase in Raman based applications is being driven by improvements in measurement techniques and instrumentation, but also by the development of data science methods. When data science is applied to Raman effect related data the aim is to extract high-level information and knowledge from subtle data differences. The high-level information depends on the task and the sample, e.g., disease types, tissue types and other properties of the samples such as concentrations of constituents. Raman spectroscopy and its related methods have several advantages, for example they can be used as non-destructive fingerprinting techniques. To unlock their full potential the whole spectroscopic data lifecycle needs to be studied. This includes aspects like data generation, data modelling and data archiving. In particular, experimental design, sample size planning, data pre-treatment, data pre-processing, chemometric and machine learning based data modelling, model transfer methods and transfer learning are important. All procedures are sequentially combined in a data pipeline that standardizes the vibrational data and extracts reliable high-level information.

Here, we present our studies researching a standardized data analysis pipeline for biomedical Raman spectra [1] and describe studies dealing with the comparability of Raman spectra between conditions and setups [2,3]. Additionally, we describe our efforts to generate a vibrational spectroscopic data base (VibSpecDB) within the NFDI4Chem project [4].

References

[1] S. Guo, J. Popp, and T. Bocklitz, Nature protocols, vol. 16, no. 12, pp. 5426–5459, 2021 doi: 10.1038/s41596-021-00620-3.

[2] S. Guo et al., Analytical Chemistry, vol. 92, no. 24, pp. 15745–15756, 2020, doi: 10.1021/acs.analchem.0c02696.

[3] S. Mostafapour, T. Dörfer, R. Heinke, P. Rösch, J. Popp, and T. Bocklitz, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, p. 123100, 2023, doi: https://doi.org/10.1016/j.saa.2023.123100.

[4] C. Steinbeck et al., Research Ideas and Outcomes, vol. 6, p. e55852, 2020, doi: 10.3897/rio.6.e55852.