International Symposium on Grids & Clouds (ISGC) 2023 in conjunction with HEPiX Spring 2023 Workshop

Contribution ID: **48** Type: **Oral Presentation**

K2I - Automated anomaly detection in the chemical footprint of surface water using machine learning

Wednesday, 22 March 2023 16:40 (20 minutes)

River and lake water is a major resource for drinking water, food production and various industrial and agricultural purposes and it hosts or feeds many sensitive ecosystems. Therefore assuring the absence of potentially harmful chemicals is a vital issue for environmental and economic sustainability. Tens of thousands of different chemicals are present in fluctuating amounts in surface water bodies, many naturally occurring, others emitted human activities. Laboratories tasked with monitoring the water quality by using targeted analysis can only detect a small subset of the chemical substances present in the water. Therefore so called Non-Target-Screening (NTS) is increasingly used by labs to perform more comprehensive monitoring. This procedure typically relies on liquid chromatography in combination with high-resolution mass spectrometry (LC- HRMS). These produce a large number of signals which are difficult to evaluate for human researchers, especially when many samples are involved. We employ data processing techniques and pattern recognition methods like Autoencoders to structure the data and perform anomaly detection. A high degree of variance in measurements and processing workflows results in low comparability of data from different laboratories, which we alleviate with data alignment processes. The K2I project aims at fostering collaboration between laboratories and research institutions working towards the goal of advanced automated water quality monitoring. A joint platform for uploading and processing raw LC-HRMS data including a cloud based Datalake and processing pipeline is being developed. A standardized processing workflow is being established which is enhanced by anomaly detection to speed up the discovery of unusual changes in water bodies.

The measurements consist of signal peaks that correspond to a specific retention time (RT) in the chromatograph and a certain mass to charge ratio (m/z) which is determined by the mass spectrometer after the substances have been ionized. These peaks can be scanned for unknown combinations of RT and ⊠/⊠, indicating the presence of so far unregistered chemicals and for strong or recurring signals that have not been attributed to a known cause yet. Neural networks like Autoencoders can be trained on historical data, to recognize common components and then spot deviations from these normal patterns. It can be advantageous to narrow down the source of emission to compare measurements taken at different locations, both in the same water body (different sites at the same river or lake) and in separate waters. Thus, the combination of LC-HRMS data from different sampling locations and laboratories, which can be enhanced with spatial and temporal coordinates and additional information such as known environmental influences could be used to more effectively notice and track micro pollutants in surface water across larger regions.

Primary authors: Mrs PAUW, Viktoria (Leibniz Rechenzentrum); Mr HAYEK, Mohamad (LRZ); Ms SHOJAEI, Elham (LRZ); Mr BADER, Tobias (Landeswasserversorgung (Ulm)); Mr HACHINGER , Stephan (LRZ)

Presenter: Mrs PAUW, Viktoria (Leibniz Rechenzentrum)

Session Classification: Artificial Intelligence (AI)

Track Classification: Track 10: Artificial Intelligence (AI)