Big brown data: analyzing field data on chemicals of emerging concern in wastewater from Nijmegen

Level: Master
Duration: 6 months
Start: April 2023 - ongoing
Project form: Data analysis, modelling
Supervision: Caterina Zillien, Ad Ragas
Contact: caterina.zillien@ru.nl

Background
We all use the toilet multiple times a day without much thinking of what happens to wastewater after flushing. However, wastewater is not just generated by using the toilet but also by using sinks, showers, washing machines, dish washers, industrial processes, car washing facilities and other activities. At the same time, most of our daily lives involve more and more chemicals in one way of the other: as a society we use pharmaceuticals, fragrances, cleaning products, food additives, plasticizers, flame retardants, anti-corrosives and so on. Many of those chemicals are being discharged with wastewater into the urban sewer system and ultimately reach the environment where they can have negative effects on wildlife (e.g. Oaks et al., 2004, Rochman et al., 2014). As different urban activities involve different chemicals, we hypothesize that the composition of wastewater regarding chemicals differs between locations within the sewer system.

Objectives and outcomes
The aim of this project is to assess differences in wastewater composition between locations in the sewer system, at different time points and sampling methods. You'll analyze datasets of sewer measurements on >200 compounds from Nijmegen to evaluate, for example, in what way industrial wastewater differs from domestic wastewater regarding which type of chemicals are present and in what concentrations. Methods could involve multivariate analysis, e.g. hierarchical clustering and principal component analysis, but are open for discussion.

Please note:
We are looking for someone with an affinity for data analysis, statistics and modelling (preferably in R). A background in chemistry could be beneficial but is not strictly necessary. This project **does not involve field or lab work** but uses real data from previous sampling campaigns.

References + further reading

Giglioli, S., Colombo, L., & Azzellino, A. (2023). Cluster and multivariate analysis to study the diffuse contamination of emerging per-and polyfluoroalkyl substances (PFAS) in the Veneto Region plain (North-eastern Italy). *Chemosphere*, *319*, 137916. https://doi.org/10.1016/j.chemosphere.2023.137916

Oaks, J. L., Gilbert, M., Virani, M. Z., Watson, R. T., Meteyer, C. U., Rideout, B. A., ... & Ahmed Khan, A. (2004). Diclofenac residues as the cause of vulture population decline in Pakistan. *Nature*, *427*(6975), 630-633. https://doi.org/10.1038/nature02317

Ringnér, M. (2008). What is principal component analysis?. *Nat Biotechnol* **26**, 303–304. https://doi.org/10.1038/nbt0308-303

Rochman, C. M., Lewison, R. L., Eriksen, M., Allen, H., Cook, A. M., & Teh, S. J. (2014). Polybrominated diphenyl ethers (PBDEs) in fish tissue may be an indicator of plastic contamination in marine habitats. *Science of the total environment*, *476*, 622-633. https://doi.org/10.1016/j.scitotenv.2014.01.058