Applied Machine-Learning Tools to Enhance PFAS Analytics and Support More Effective Site Management

Paul Hurst, Skyler Sorsby and Stefano Marconetto, WSP

Background/Objectives: Complex mixtures of per- and polyfluoroalkyl substances (PFAS) can enter the environment through a wide variety of sources. Once in the environment, different fate-and-transport characteristics alter the composition and spatial distribution, further increasing the number of chemical signatures observed at a given site. Additional complexities in managing PFAS sites include the ubiquitous environmental presence of PFAS in the environment at low levels, differentiating potential site impacts from background concentrations and stringent regulatory limits. This, along with the large number of PFAS compounds, makes site management, assessing sources and tracking plumes difficult or infeasible with traditional site investigation methods. Ultimately there is a burden on site owners to establish site specific knowledge to understand their relative responsibility for PFAS impacts and present data-driven boundaries for remediation and management plans.

Approach: Increasingly accessible machine-learning tools have popularized big-data approaches that “learn” the patterns of various source types (like a specific aqueous filmforming foam). New PFAS fingerprinting tools developed by WSP combine robust algorithms to characterize impacts on contaminated sites and assist in site management. The custom tools allow assessment of background levels, fate-and-transport patterns, identification of potential sources and evaluation of how those sources may be migrating and mixing across the site. Key tools used include 1) nonnegative matrix factorization which performs blindsource separation of different signatures and can estimate both their chemical profiles and highlight spatial hot spots and 2) clustering algorithms to partition the site into zones of similar impacts that may be used to target further site investigation and remediation. A novel spatially informed clustering algorithm also allows both chemical signatures and geography to inform clustering which is highly useful for large sites. The presentation will review the approach, highlight key aspects and benefits of the tools used and provide case studies from Canadian sites showing typical outputs and outcomes.

Results: Matrix factorization completed at multiple sites consistently identifies multiple potential onsite PFAS source areas and distinct chemical fingerprints. Due to the spatial patterns and chemical compositions of these fingerprints, it is likely that some of the patterns represent precursor transformations or differential transport effects rather than offsite sources. Similarly, clustering algorithms such as Hierarchical Clustering (HCA) and Gaussian Mixture Model (GMM) typically identify significant patterns in the data and successfully group like samples at inferred sources. However, patterns downgradient sometimes become less differentiated due to transformations and/or differential transport. Spatially informed clustering has also been successfully applied to partition sites into multiple data-driven focus areas for further investigation and remedial design.

Key Lessons Learned: As with other analyses, rigorous preprocessing and quality checking were important safeguards against artificial bias. For example, elevated detection limits consistently create noise in the data and a decision is required to either exclude a compound with elevated detection limits or include it but accept that signals and clusters will have artificial peaks. Machine-learning provides a highly useful additional line of evidence to traditional site investigation methods and allow for large amount of data to be synthesized rapidly. Distillation of major themes and effective use of visuals from these complex results are key to communicating findings to stakeholders. The results indicate a significant potential for the differentiation of sources, fate-and-transport patterns and ambient levels to inform effective site management and improve engineering considerations early in the remediation project life cycle.

continued on next page...

...continued from previous page

Paul Hurst

Paul Hurst is Principal Senior Environmental Engineer located in WSP's Ottawa ON office. Paul has been consulting for over 17 year and has broad experience encompassing project management, data management, data science, design, and execution of environmental site assessment (ESA), long term monitoring and remediation projects, including: conceptual site model (CSM) development (2D and 3D), remedial options evaluation, remedial action plan development, groundwater pump and treat system design, screening level risk assessments and data gap analysis.

Paul is an active participant in the Python and Data Science communities at WSP. Paul implements machine learning including development of fingerprinting code and a creation of a novel method for spatially weighted hierarchal clustering which has been applied at large PFAS projects to provide insight and multiple lines of evidence for source identification. Paul utilizes advanced data analytics, machine learning, clustering algorithms and principle components to provide details on fingerprinting, source separation, fate and transport, and linkage between media for PFAS CSMs enhanced interpretation and conclusions. The insights gained help ensure an effective and targeted approach to PFAS management and also allow efficient interpretation of large datasets. Paul provides data science expertise for projects across Canada. He supports WSPs digital data collection, management and interpretation. He regularly designs and develops data analytics tools to allow automated data visualizations (trend analysis, Mann-Kendall, etc), fingerprinting (including machine learning) and statistical outputs on large datasets. He also led Golder Canada's (now WSPs) implementation of an environmental quality database (EQuIS), which is now used globally.