Explainable AI (XAI) for Understanding PFAS Distributions at Multiple PFAS Use Sites

Complex source compositions and fate-and-transport processes for per- and polyfluoroalkyl substances (PFAS) complicate the interpretation of their distribution in the environment. Airports, landfills, chrome platers and refineries / terminals represent four common types of PFAS sources. This case study focuses on using PFAS data on these types of sites as well as hydroclimate attributes (e.g., land cover, urban density, climate, topographic, distance to source sites) to assess potential patterns and drivers. Through this, we leveraged novel data-driven chemical profiles and explainable AI (XAI) to distill salient PFAS signals and relate varying levels to industrial, climate, and hydrologic factors.

PFAS analytical data, with a focus on groundwater data for all four site types were reviewed. Initial summaries of PFAS occurrence were prepared and a geographic dataset of potential PFAS-related sites were incorporated as independent variables in a nonlinear regression analysis. Using this, potential links between hydroclimate attributes and PFAS levels were established using machine learning tree-based ensemble methods such as XGBoost and permutation-based XAI tools. A novel linear-algebra method was developed to extract the most source-like samples from each site and the reduced dataset was further interrogated with a nonlinear manifold-learning technique to identify potential PFAS signature characteristics.

Four general PFAS levels were identified, with samples in the lowest (ambient) range present on at least 61% of refinery/terminals and up to 92% of municipal solid-waste landfills, indicating that adequate assessment of ambient levels will be important for designing site-specific remediation plans. The regional driver assessment highlighted distance to specific types of sites as the most influential for predicting PFAS levels. Further, XAI indicated that airports are the most influential for predicting PFAS levels, followed by refineries/terminals, landfills, treatment plants, and chrome plating facilities. Interestingly, proximity to a water body, grassland/shrub landcover, and urban footprint were also found to be influential predictors.

Multiple types of potential source signals were identified for each site type: a) PFOA/PFHxA more prominent at landfills, b) PFBS or PFOS at chrome plating facilities, c) PFHxA and a PFOS-dominated mix with PFHxS and PFHxA at airports, and d) PFHxA/PFHpA or PFBS/PFHxA/PFHxS dominating at refineries/terminals. Although most strongly exhibited at refinery/terminal sites, a PFHxS-dominated mix with PFOS was also detected at some chrome plating and airport facilities.

Paul Hurst
WSP Canada Inc.

Paul Hurst is a Principal Senior Environmental Engineer at WSP's Ottawa, ON office with 20 years of consulting experience. His expertise spans project management, data science, and environmental site assessment (ESA), long-term monitoring, and remediation. Paul applies advanced analytics and AI / machine learning (e.g., clustering, dimensionality reduction, principal components) to fingerprint contaminants, assess source separation, and analyze fate and transport. His work integrates multiple lines of evidence to develop robust conceptual site models (CSMs) and assist clients with cost effective defensible site management.