

Publication

Enhancing PFAS data integrity

In a recent article for the *Journal of Hazardous Materials Advances*, Haley & Aldrich's Jialin Dong and her coauthors examine how a more-standardized AI-driven approach can help users work with environmental data, especiallyper- and polyfluoroalkyl substances (PFAS) data, more efficiently.

The article focuses on a persistent challenge in PFAS investigations: Environmental data are often maintained by different states in different formats, which makes comparisons difficult and requires substantial manual data cleaning and review. It also analyzes PFAS distribution across the United States and different environmental matrices, providing context for geographic trends and background concentrations. This challenge is especially relevant forgovernment agencies that manage environmental databases, as well as for clients that are conducting PFAS investigations and need background concentration data for a site or region. Private-sector organizations involved in extensive data collection may also benefit from the approach.

Dong's work shows how integrating these datasets and applying a large language model (LLM)-based approach can reduce manual effort, improve data evaluation, and provide more useful background context before deeper investigation. She wrote the article in collaboration with colleagues from the University of California, Irvine: Sean D. Young, Zixin Hu, and Christopher I. Olivares.

Ultimately, the article suggests that a more automated and standardized method for PFAS data review could save time, improve consistency, and make environmental contaminant data easier to use across jurisdictions.

Read the full article, "Enhancing PFAS data integrity: An LLM-based FAIR+Environmental principle for improved evaluation of environmental contaminants and related constituent databases."