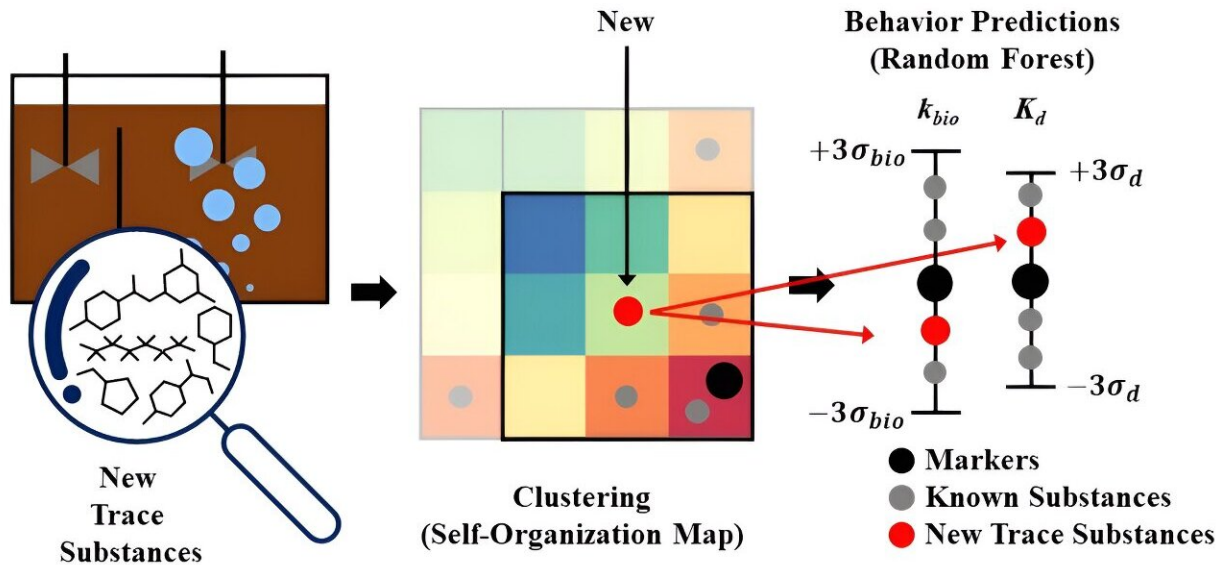


Researchers quickly and easily predict emerging contaminant concentrations in wastewater with AI

January 23 2024



Machine learning approaches for predicting the behavior of new trace substances. Credit: Korea Institute of Science and Technology (KIST)

The global consumption of pharmaceuticals is growing rapidly every

year, reaching 4 billion doses in 2020. As more and more pharmaceuticals are metabolized by the human body and enter sewage and wastewater treatment plants, the amount and types of trace substances found in them are also increasing.

When these trace substances enter rivers and oceans and are used as [water sources](#), they can have [harmful effects](#) on the environment and [human health](#), including carcinogenesis and endocrine disruption. Therefore, technologies are needed to quickly and accurately predict the properties and behavior of these trace substances, but analyzing unknown trace substances requires expensive equipment, skilled experts, and a long time.

A Korea Institute of Science and Technology (KIST) team led by Hong Seok-won, head of the Water Resources and Cycle Research Center, and Son Moon, a senior researcher, has developed a technology to classify emerging trace substances according to their physicochemical properties and predict their concentrations using clustering and prediction-based artificial intelligence technology.

The researchers used self-organizing maps, an AI technique that clusters data into maps based on their similarities, to classify 29 known trace substances, including medicinal compounds and caffeine, based on information such as physicochemical properties, functional groups, and biological reaction mechanisms. The research is [published](#) in the journal *npj Clean Water*.

Random forests, a machine learning technique that classifies data into subsets, were then further built to predict the properties and concentration changes of new trace substances.

If a new trace substance belongs to a cluster in the self-organizing map, the properties of other substances in that cluster can be used to predict

how the properties and concentration of the new trace substance will change.

As a result of applying this clustering and prediction AI model (self-organizing map and random forest) to 13 new trace substances, the prediction accuracy of about 0.75 was excellent, far exceeding the prediction accuracy of 0.40 of existing AI techniques using biological information.

Compared to traditional prediction methods based on formulas, the KIST research team's data-driven analysis model has the advantage of only inputting the physicochemical properties of trace substances and efficiently identifying how the concentration of new trace substances will change in the sewage treatment process through clustering with substances with similar data.

In addition, the data-driven AI model can be used in the future to predict the concentration of new substances such as drugs that are of social concern.

"It can be applied not only to actual [wastewater treatment plants](#), but also to most water treatment-related facilities where new trace substances exist, and can provide quick and [accurate data](#) in the policy-making process for related regulations," said Dr. Seokwon Hong and Dr. Moon Son of KIST.

"Since it utilizes machine learning technology, the accuracy of the prediction will improve as relevant data is accumulated."

More information: Seung Ji Lim et al, Clustering micropollutants and estimating rate constants of sorption and biodegradation using machine learning approaches, *npj Clean Water* (2023). [DOI: 10.1038/s41545-023-00282-6](https://doi.org/10.1038/s41545-023-00282-6)

Provided by National Research Council of Science and Technology

Citation: Researchers quickly and easily predict emerging contaminant concentrations in wastewater with AI (2024, January 23) retrieved 29 April 2024 from <https://phys.org/news/2024-01-quickly-easily-emerging-contaminant-wastewater.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.