

# Big data approach to water quality applied at shale drilling sites

December 7 2016

---

A computer program is diving deep into water quality data from Pennsylvania, helping scientists detect potential environmental impacts of Marcellus Shale gas drilling.

The work, supported by a new \$1 million grant from the National Science Foundation, pairs a cross-disciplinary team of Penn State computer scientists and geoscientists studying methane concentrations in the state's streams, rivers and private water wells.

"We want to take a data-driven approach to assess the impact of shale gas development," said Zhenhui "Jessie" Li, assistant professor of information sciences and technology, Penn State, and co-principal investigator on the project. "We are using data-mining techniques and learning computer models to look at how methane concentration correlates with other factors, like distance from unconventional shale gas wells and geological features such as faults."

Methane occurs naturally in waterways, but may also be released by unconventional drilling, or fracking, associated with natural gas development. While environmental impacts from shale drilling appear to be rare compared to the number of drilled wells, the testing will give scientists a better idea of how and why they do occur.

Penn State geoscientists have studied methane concentrations around shale gas drilling for years and have collected large datasets from samples taken by researchers, environmental groups and government

agencies. However, sifting through big data is labor intensive, and complex patterns that might reveal potential problems can be difficult for humans to spot.

"It's always been frustrating to me because we get these water datasets that are sort of here, there and everywhere, and you can't put them together into a scientific explanation," said Susan Brantley, distinguished professor of geosciences and director of the Earth and Environmental Systems Institute, Penn State, and principal investigator on the project. "It's like a jigsaw puzzle with a lot of pieces missing. The story is still there, but you can't see it as a human being. I think it's possible the computer techniques that Jessie has can help us pull that story out and fill in some of the missing pieces." Li uses computer models and data mining techniques to analyze the data for hot spots, areas where methane concentrations are higher than expected and may not be readily explained by natural causes. Geoscientists can then focus on those narrow areas for further study.

Using the techniques, Li and her team found methane levels tend to be higher around fault lines. But the models can go further, analyzing the impacts of different [shale gas](#) wells and older conventional gas and oil wells around the faults. "The combination of these two features may cause the methane to be slightly higher in some areas," Li said. "It could be a very complicated rule involving multiple factors together. So the way the machine-learning model works is from this massive data, we can learn these kinds of complicated rules."

The work could lead to a better understanding of how fracking, orphaned and abandoned oil and gas wells, and other factors occasionally impact the environment. As part of the project, researchers will train citizen scientists to collect additional water samples and will host workshops aimed to share the data with the public and foster dialogue among diverse stakeholders.

"I think it's going to be helpful for people in Pennsylvania," Brantley said. "I think what we've found is if there are problems, they are relatively infrequent. But we are also starting to look at other issues like lead or arsenic in drinking water. We can see the impact of coal mine- and acid mine-drainage. We can look at all different kinds of water resources, and we can work with people in Pennsylvania to teach them about it."

The researchers said the project also provides important, cross-disciplinary opportunities for students from the Colleges of Earth and Mineral Sciences and IST. Computer science students have had the rare chance to go into the field, while geosciences students have had the opportunity to look at their data in way that was previously unavailable to them. "I'm just a person who really likes interdisciplinary work, because I want to make real-world impacts," Li said. "I feel like data-mining researchers should go in this direction. There are a lot of interesting real-world problems we can help solve."

Provided by Pennsylvania State University

Citation: Big data approach to water quality applied at shale drilling sites (2016, December 7) retrieved 26 April 2024 from

<https://phys.org/news/2016-12-big-approach-quality-shale-drilling.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.