

Stanford students deploy machine learning to aid environmental monitoring

October 1 2018, by Rob Jordan



Caption: Satellite images of river outflows to the Atlantic Ocean in the wake of Hurricane Florence show water discolored by debris and pollutants. Credit:

NASA

As Hurricane Florence ground its way through North Carolina, it released what might politely be called an excrement storm. Massive hog farm manure pools washed a stew of dangerous bacteria and heavy metals into nearby waterways.

More efficient oversight might have prevented some of the worst effects, but even in the best of times, state and federal environmental regulators are overextended and underfunded. Help is at hand, however, in the form of machine learning—training computers to automatically detect patterns in data—according to Stanford researchers.

Their study, published in *Nature Sustainability*, finds that machine learning techniques could catch two to seven times as many infractions as current approaches, and suggests far-reaching applications for public investments.

"Especially in an era of decreasing budgets, identifying cost-effective ways to protect public health and the environment is critical," said study coauthor Elinor Benami, a graduate student in the Emmett Interdisciplinary Program on Environment and Resources (E-IPER) in Stanford's School of Earth, Energy & Environmental Sciences.

Optimizing resources

Just as the IRS can't audit every taxpayer, most government agencies must constantly make decisions about how to allocate resources. Machine learning methods can help optimize that process by predicting where funds can yield the most benefit. The researchers focused on the Clean Water Act, under which the U.S. Environmental Protection

Agency and state governments are responsible for regulating more than 300,000 facilities but are able to inspect less than 10 percent of those in a given year.

Using data from past inspections, the researchers deployed a series of models to predict the likelihood of failing an inspection, based on facility characteristics, such as location, industry and inspection history. Then, they ran their models on all facilities, including ones that had yet to be inspected.

This technique generated a risk score for every facility, indicating how likely it was to fail an inspection. The group then created four inspection scenarios reflecting different institutional constraints—varying inspection budgets and inspection frequencies, for example—and used the score to prioritize inspections and predict violations.

Under the scenario with the fewest constraints—unlikely in the real world—the researchers predicted catching up to seven times the number of violations compared to the status quo. When they accounted for more constraints, the number of violations detected was still double the status quo.

Limits of algorithms

Despite its potential, machine learning has flaws to guard against, the researchers warn. "Algorithms are imperfect, they can perpetuate bias at times and they can be gamed," said study lead author Miyuki Hino, also a graduate student in E-IPER.

For example, agents, such hog farm owners, may manipulate their reported data to influence the likelihood of receiving benefits or avoiding penalties. Others may alter their behavior—relaxing standards when the risk of being caught is low—if they know their likelihood of

being selected by the algorithm. Institutional, political and financial constraints could limit machine learning's ability to improve upon existing practices. The approach could potentially exacerbate environmental justice concerns if it systematically directs oversight away from facilities located in low-income or minority areas. Also, the machine learning approach does not account for potential changes over time, such as in public policy priorities and pollution control technologies.

The researchers suggest remedies to some of these challenges. Selecting some facilities at random, regardless of their risk scores, and occasionally re-training the model to reflect up-to-date risk factors could help keep low-risk facilities on their toes about compliance. Environmental justice concerns could be built into inspection targeting practices. Examining the value and trade-offs of using self-reported data could help manage concerns about strategic behavior and manipulation by facilities.

The researchers suggest future work could examine additional complexities of integrating a machine learning approach into the EPA's broader enforcement efforts, such as incorporating specific enforcement priorities or identifying technical, financial and human resource limitations. In addition, these methods could be applied in other contexts within the U.S. and beyond where regulators are seeking to make efficient use of limited resources.

"This model is a starting point that could be augmented with greater detail on the costs and benefits of different inspections, violations and enforcement responses," said co-author and fellow E-IPER graduate student Nina Brooks.

More information: M. Hino et al. Machine learning for environmental monitoring, *Nature Sustainability* (2018). [DOI:](#)

[10.1038/s41893-018-0142-9](https://doi.org/10.1038/s41893-018-0142-9)

Provided by Stanford University

Citation: Stanford students deploy machine learning to aid environmental monitoring (2018, October 1) retrieved 26 April 2024 from <https://phys.org/news/2018-10-stanford-students-deploy-machine-aid.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.