What can machine learning reveal about the solid Earth?

March 22 2019, by Josie Garthwaite

Geoscientists have harnessed a technique commonly used for speech recognition to detect events ranging from alpine rockslides to volcanic warning signs that would otherwise go unnoticed. Credit: USGS Hawaiian Volcano Observatory

Scientists seeking to understand Earth's inner clockwork have deployed

armies of sensors listening for signs of slips, rumbles, exhales and other disturbances emanating from the planet's deepest faults to its tallest volcanoes. "We measure the motion of the ground continuously, typically collecting 100 samples per second at hundreds to thousands of instruments," said Stanford geophysicist Gregory Beroza. "It's just a huge flux of data."

Yet scientists' ability to extract meaning from this information has not kept pace.

The solid Earth, the oceans and the atmosphere together form a geosystem in which physical, biological and chemical processes interact on scales ranging from milliseconds to billions of years, and from the size of a single atom to that of an entire planet. "All these things are coupled at some level," explained Beroza, the Wayne Loel Professor in the School of Earth, Energy & Environmental Sciences (Stanford Earth). "We don't understand the individual systems, and we don't understand their relationships with one another."

Now, as Beroza and co-authors outline in a paper published March 21 in the journal *Science*, machine-learning algorithms trained to explore the structure of ever expanding geologic data streams, build upon observations as they go and make sense of increasingly complex, sprawling simulations are helping scientists answer persistent questions about how the Earth works.

From automation to discovery

"When I started collaborating with geoscientists five years ago, there was interest and curiosity around machine learning and data science," recalled Karianne Bergen, lead author on the paper and a researcher at the Harvard Data Science Initiative who earned her doctorate in computational and mathematical engineering from Stanford. "But the

community of researchers using machine learning for geoscience applications was relatively small."

That's changing rapidly. The most straightforward applications of machine learning in Earth science automate repetitive tasks like categorizing volcanic ash particles and identifying the spike in a set of seismic wiggles that indicates the start of an earthquake. This type of machine learning is similar to applications in other fields that might train an algorithm to detect cancer in medical images based on a set of examples labeled by a physician. More advanced algorithms unlocking new discoveries in Earth science and beyond can begin to recognize patterns without working from known examples.

"Suppose we develop an earthquake detector based on known earthquakes. It's going to find earthquakes that look like known earthquakes," Beroza explained. "It would be much more exciting to find earthquakes that don't look like known earthquakes." Beroza and colleagues at Stanford have been able to do just that by using an algorithm that flags any repeating signature in the sets of wiggles picked up by seismographs – the instruments that record shaking from earthquakes – rather than hunting for only the patterns created by earthquakes that scientists have previously catalogued.
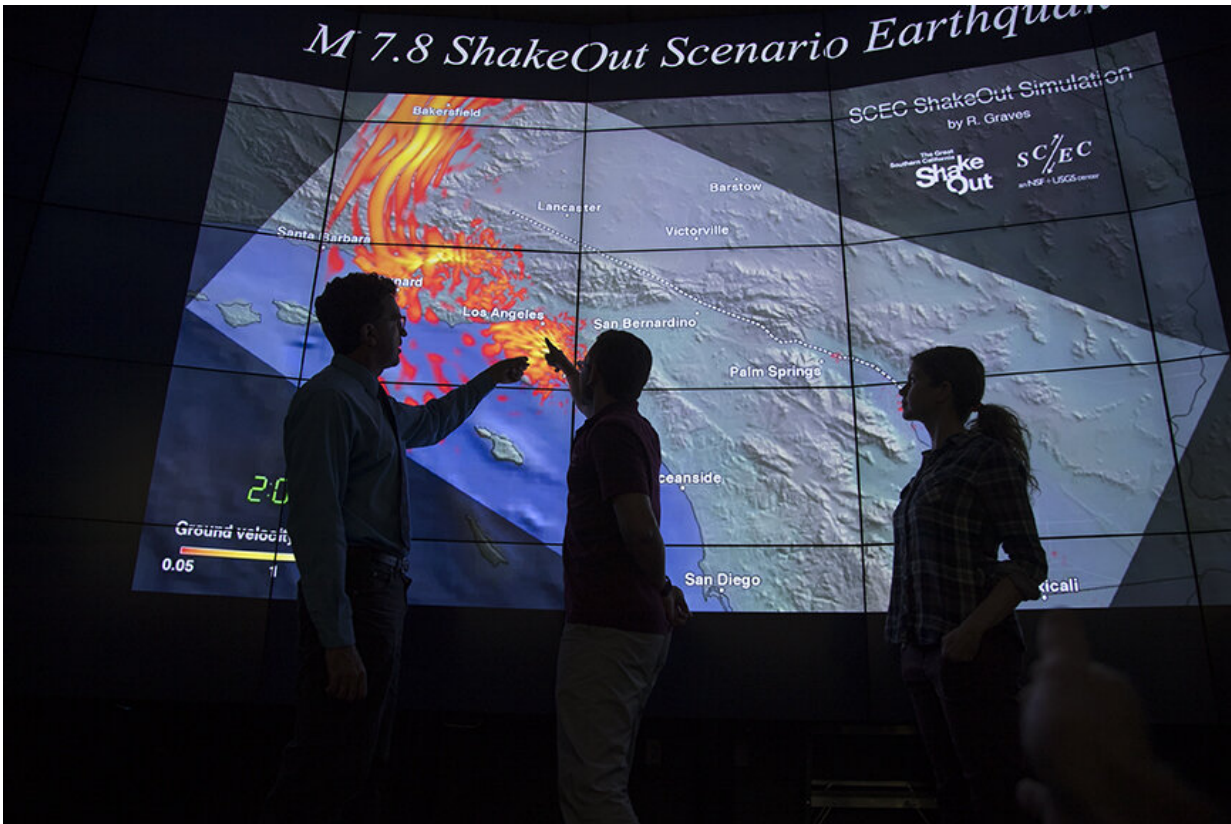
Geophysicist Gregory Beroza is among a growing number of scientists who are training machine learning algorithms to make sense of increasingly complex simulations of Earth's geosystems. Credit: Stacy Geiken

Both types of algorithms – those with explicit labeling in the training data and those without – can be structured as deep neural networks, which act like a many-layered system in which the results of some transformation of data in one layer serves as the input for a new computation in the next layer. Among other efforts noted in the paper, these types of networks have allowed geoscientists to quickly compute the speed of seismic waves – a critical calculation for estimating earthquake arrival times – and to distinguish between shaking caused by Earth's natural motion as opposed to explosions.

An imperfect mimic

In addition to spotting overlooked patterns, machine learning can also help to tame overwhelming data sets. Modeling how an earthquake affects the viscous part of the layer in Earth's interior that extends hundreds of miles below the planet's outermost crust, for example, requires an insurmountably large amount of computing power. But machine learning algorithms can find shortcuts, essentially mimicking solutions to more detailed equations with less computing.

"We can get a pretty good approximation to reality, which we'll be able to apply to data sets that are so big or simulations that are so extensive that the most powerful computers available would not be able to process them," Beroza said.

What's more, any shortfalls in the precision of artificial intelligence-based solutions to these equations often pale in significance compared to the influence of scientists' own decisions about how to set up calculations in the first place. "Our largest source of error comes not from our inability to solve the equations," Beroza said. "It comes from knowing what the interior structure of the Earth is really like and the parameters that should go into those equations."

Open science

To be sure, machine learning is far from a perfect tool for answering the thorniest questions in Earth science. "The most powerful machine-learning algorithms typically require large labeled data sets, which are not available for many geoscience applications," Bergen said. If scientists train an algorithm on insufficient or improperly labeled data, she warned, it can cause models to reproduce biases that don't necessarily reflect reality.

This type of error can be combated in part through greater transparency and creation of "benchmark" data sets, which the researchers argue can spur competition and allow for apples-to-apples comparisons of algorithm performance. According to Bergen, "Adoption of open science principles, including sharing of data and code, will help to accelerate research and also allow the community to identify and address limitations or weaknesses of proposed approaches."

Human impatience may be harder to keep in check. "What I'm worried about is that people are going to use AI naively," Beroza said. "You could imagine someone training a many-layer, deep neural network to do earthquake prediction – and then not testing the method in a way that properly validates its predictive value."

Provided by Stanford University