

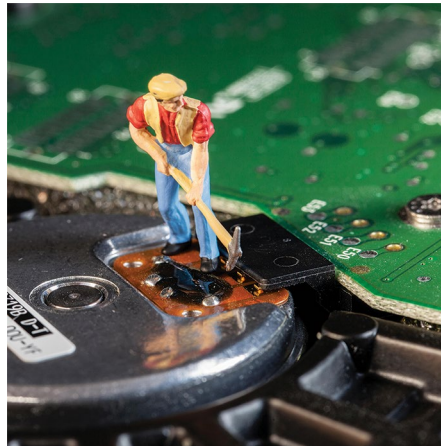
# Use machines to tame big data

Machine learning allows geoscientists to embrace data at scales greater than ever before. We are excited to see what this innovative tool can teach us.

Geoscientists need a comprehension of scale and often an acceptance of data limitation: their object of study — the Earth — is big, old and complicated<sup>1</sup>. For decades — even centuries — geoscientists were necessarily experts filling in the gaps of limited data sets. However, with the advent of automation in all areas of life, including observational techniques, the diversity and wealth of data types has vastly increased. As such, a field that was once characterized by data scarcity now has some geoscientists learning how to tackle a different challenge: big data. A pair of papers in this issue of *Nature Geoscience* takes on that challenge, using machine learning to pull important information on fault physics out from a cacophony of background noise generated by laboratory earthquakes (Hulbert et al.) and the Cascadia subduction zone (Rouet-Leduc et al.).

Whilst there is still plenty of scope for a toolkit of hammer and hand-lens, in this New Year we are excited to see how our authors embark towards new frontiers, such as machine learning<sup>2</sup>. Not all disciplines in the geosciences have been equally exposed to big data as the world and science have become increasingly digital. In the planetary sciences, the need to process large image databases resulted in the relatively early adoption of trainable software and machine-learning techniques. Over two decades ago, the pioneering JARtool was developed in an effort to automate the recognition of volcanoes from images of Venus<sup>3</sup>. It was easier for geologists to identify — and to train an algorithm to identify — a volcano than it was to distinguish at a pixel level what is and is not a volcanic edifice.

Similar techniques can be used when satellite technology is turned to look at Earth. Beyond simple automatic feature recognition, the InSAR (interferometric synthetic aperture radar) technique compares two or more images. Ground deformation can be detected from the difference between them. The launch of



**Fig. 1 |** Modern-day (data) mining in the geosciences. Credit: Duncan Astbury / Alamy Stock Photo

Sentinel-1, a pair of polar-orbiting InSAR imaging satellites, allows global monitoring of surface elevation — but there is too much incoming data for even the most dedicated PhD student to process manually. Machine-learning algorithms can handle these data better and faster than humans. They allow global ground deformation to be detected automatically so that, for example, volcanoes in remote and inaccessible regions may be routinely monitored for hazard assessment<sup>4</sup>.

Not only global data benefit from machine-learning techniques. Large detector arrays and continuous monitoring in seismology produce overwhelming but often local datasets. On the north Cascadia subduction zone, signal was separated from noise using machine learning (Rouet-Leduc et al.), which meant there was no need to input a theoretical model with all of the assumptions that would have required. As opposed to discrete episodic bursts of tremor, a continuous tremor-like signal was detected. And, reduced down to a laboratory scale, machine learning can predict the timing, duration and shear displacement

of a laboratory earthquake based on the attributes of the preceding micro-failure events (Hulbert et al.). Machine learning is a promising tool in seismology but real-world earthquake forecasting remains a goal rather than reality.

In some disciplines such as remote sensing, meteorology and others, analysing large datasets is run-of-the-mill. Other fields — say those that require manual sample collection or observation — do not naturally generate massive datasets. However, in some instances, many smaller datasets can be combined to a repository worth mining. For example, global biocrust coverage was quantitatively shown to be vulnerable to changes in land use and climate, not through inherently big data but through the careful collation of over 500 publications and analysis by machine learning<sup>5</sup>. It seems likely that even more insights are available from existing data for those that are willing to trawl the literature and apply these methods (Fig. 1).

Machine-learning techniques can help to recognize patterns; identify features; separate signal from noise; and build predictive models. Rather than discipline specific, it is a tool that allows computers to learn by example — to sift through data, learn the rules and build a model. With data at its heart, a great algorithm is useless without high-quality training data but the data are hard to make sense of without the algorithm. And make sense they must: it is what machine learning can teach us about the Earth and other planets that is so exciting. □

Published online: 20 December 2018  
<https://doi.org/10.1038/s41561-018-0290-6>

## References

1. Kastens, K. A. et al. *Eos Trans. AGU* **90**, 265–266 (2009).
2. Marone, C. *Nat. Geosci.* **11**, 301–302 (2018).
3. Burl, M. C. et al. *Machine Learning* **30**, 165–194 (1998).
4. Anantrasirichai, N., Biggs, J., Albino, F., Hill, P. & Bull, J. *JGR Solid Earth* **123**, 6592–6606 (2018).
5. Rodriguez-Caballero, E. et al. *Nat. Geosci.* **11**, 185–189 (2018).