

EOS

VOL. 101 | NO. 8
AUGUST 2020

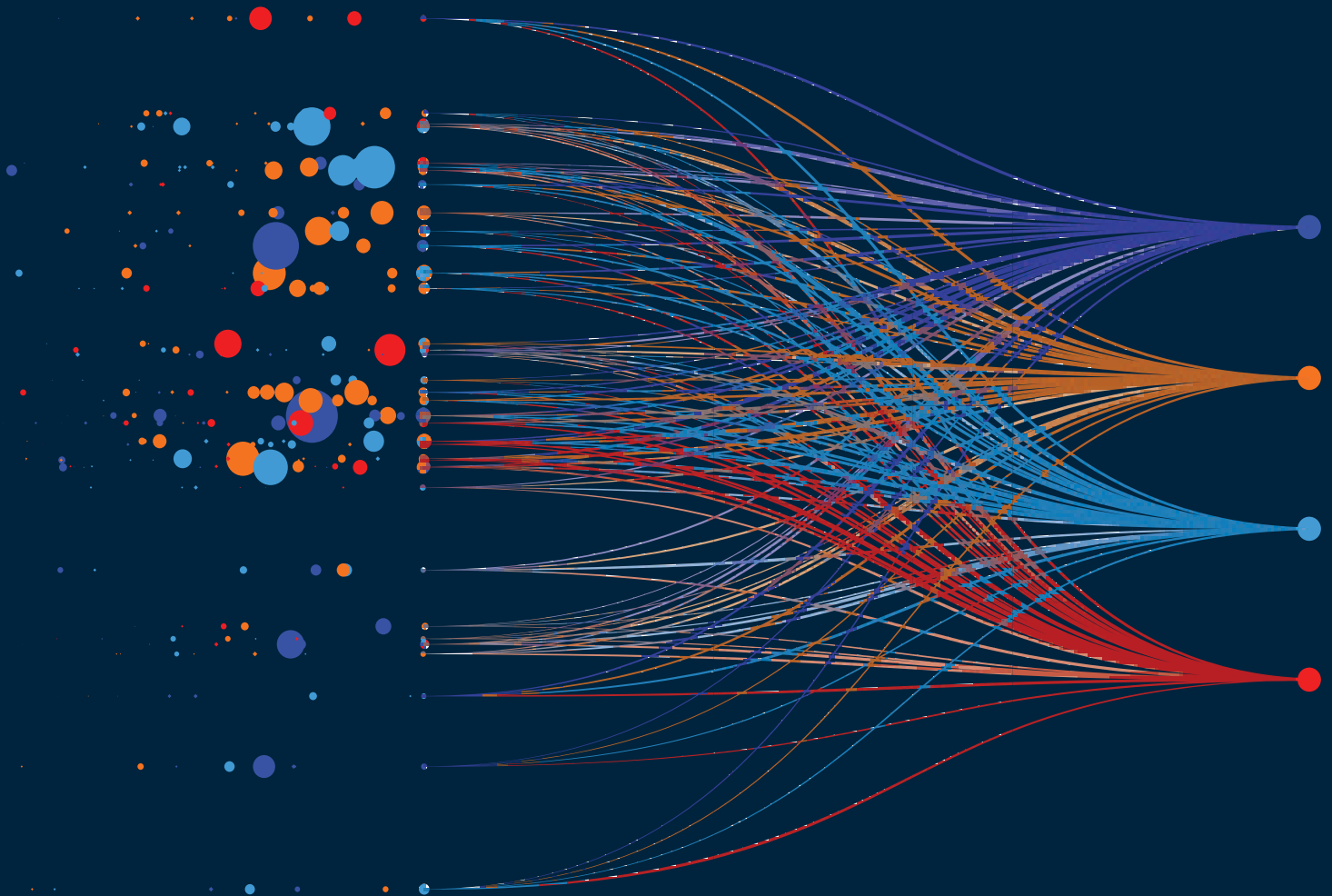
SCIENCE NEWS BY AGU

Europe's Biodiversity Strategy

A Virtual Hackathon
Fights Locusts

MH370's Search
Reveals New Science

INNOVATIONS IN TECHNOLOGY GOT US INTO
THE DATA PROBLEM.
WE NEED AN EVOLUTION IN TECHNOLOGY TO GET US OUT.



Machine Learning Can Help Decode Alien Skies—Up to a Point

Future telescopes like the James Webb Space Telescope (JWST) and the Atmospheric Remote-sensing Infrared Exoplanet Large-survey (ARIEL) are designed to sample the chemistry of exoplanet atmospheres. Ten years from now, spectra of alien skies will be coming in by the hundreds, and the data will be of a higher quality than is currently possible.

Astronomers agree that new analysis techniques, including machine learning algorithms, will be needed to keep up with the flow of data and have been testing options in advance. A study that has been accepted by *Monthly Notices of the Royal Astronomical Society* trialed one such algorithm against the current gold standard method for decoding exoplanet atmospheres to see whether the algorithm could tackle this future big-data problem (bit.ly/machine-learning-exoplanets).

“We got really good agreement between [the answers from] our machine learning method and the traditional Bayesian method that most people are using,” said Matthew Nixon. Nixon is the lead researcher on the project and an astronomy doctoral student at the University of Cambridge in the United Kingdom.

However, he said, “as we increased the parameter space, the computational efficiency of our method dropped.... As we started to add more parameters, we started to get hit by the curse of dimensionality.”

A Random Forest Breathing in Exotic Air

Astronomers measure the spectrum of an exoplanet’s atmosphere when starlight shines through it or when heat from the planet lights up the atmosphere from within. In either scenario, the atmosphere imprints its chemical signature on the light and is detected by our telescopes.

The current front-runner for best deciphering a planet’s spectrum is called atmospheric retrieval. It uses statistical inference to calculate the likelihood that given an observed spectrum, an exoplanet’s atmosphere has a certain composition, temperature, level of cloud cover, and heat flow. The technique has so far proven very reliable but can be computationally expensive.

“The more detailed the data, the more detailed the model needs to be,” said Ingo Waldmann, an astrophysicist at University



College London in the United Kingdom who was not involved with this study. “Perhaps unsurprisingly, the more detailed the model, the longer it takes to compute its results. Today we are rapidly reaching a stage where our traditional techniques become too slow to compute these increasingly complex models.”

“As we started to add more parameters, we started to get hit by the curse of dimensionality.”

Nixon and his adviser and coauthor, Nikku Madhusudhan, also at the University of Cambridge, tested a type of supervised machine learning algorithm called a random forest, which is made up of thousands of decision trees. Each decision tree makes its prediction for a likely combination of atmospheric properties, and then the algorithm generates an artificial spectrum that has those properties.

The algorithm compares each artificial spectrum with the real one and chooses the closest match. This is a new and modified application of the random forest algorithm for exoplanetary atmospheres that was originally developed by astronomers at the University of Bern in Switzerland.

The researchers tested their algorithm on two exoplanets with exceptionally well studied atmospheres and found that the random forest’s solution matched the one from atmospheric retrieval. Moreover, “the authors achieve a much faster interpretation of the data than otherwise possible with traditional techniques,” Waldmann said.

However, the two exoplanets in question, WASP-12b and HD 209458b, are both very hot Jupiter-sized planets. The algorithm could easily simplify its decision because each planet’s atmosphere consists mostly of hydrogen and helium, Nixon said.

“Generally speaking,” Madhusudhan explained, “it is going to be slightly harder to retrieve atmospheric properties of cooler and smaller planets,” for example, super-Earths or Earths. “This is because the spectral signatures are expected to be smaller for such planets, which makes it harder to extract the same amount of information as we have for

hot Jupiters currently.” For planets with faint signals and those whose base constituents are unknown—an ocean world, super-Earth, or temperate-zone planet—the random forest would lose its computational edge.

A Balanced Approach for the Way Forward

This study adds to a growing effort by exoplanet scientists to find an efficient way to handle the upcoming deluge of atmospheric data. “It is great to see a growing group in the community using machine learning methods and cross-checking each other’s results and claims,” said Daniel Angerhausen, an astrophysicist at ETH Zürich in Switzerland who was not involved with this research.

Missions like JWST and ARIEL are first at bat, but Angerhausen is also thinking about missions that will come after those. Astronomers will need to strategize the most efficient ways to observe interesting targets. “This problem is predestined for a [machine learning] approach,” Angerhausen said. A random forest approach is just the “tip of the iceberg” for algorithms to try.

Nixon agreed and said that “going forward, looking at different machine learning algorithms is definitely a positive [step], and also looking at how we can combine these machine learning approaches into hybrid methods to really boost these retrievals to the next level.”

As exoplanet atmosphere research moves into the big-data era, machine learning will become an increasingly important research tool scientists should be trained to use, Madhusudhan said. Some graduate programs are already integrating more data science learning into students’ training. (Nixon’s doctorate work is supported by one such program in the United Kingdom.)

“On the other hand,” Madhusudhan added, “it also needs to be recognized that while machine learning is a great research tool in various areas, there are also important areas of research where other numerical, statistical, and analytic approaches are more suitable for some important problems. Therefore, I believe the right balance needs to be met while integrating machine learning into graduate programs in the right research areas.”

“Machine learning may never replace an atmospheric expert,” Waldmann said, “but I’m certain that artificial intelligence will certainly play a role as a helping hand.”

By **Kimberly M. S. Cartier** (@AstroKimCartier), Staff Writer

Search for MH370 Revealed Ocean Crust Waves



The M/V Fugro Equator, seen here, was one of three ships that collected sonar data from June 2014 to June 2016 while searching for Malaysia Airlines Flight 370 (MH370). Those data led to a new high-resolution bathymetry map of an area near the Southeast Indian Ridge. Credit: Justin Baulch, Australian Transport Safety Bureau, CC BY 4.0 (bit.ly/ccby4-0)

Malaysia Airlines Flight 370 (MH370) disappeared on 8 March 2014 somewhere in the southern Indian Ocean. The tragic loss of all 239 people aboard led to a multinational, multiyear search to find and recover the plane. Despite this effort, it remains missing.

As part of that search effort, three survey vessels collected multibeam bathymetry data using hull-mounted sonars. These data were used to create maps of the seafloor that span roughly 88,000 square kilometers near the Southeast Indian Ridge west of Australia.

The bathymetry, at a resolution 15 times sharper than previous maps of the area, did not reveal MH370’s final resting place. However, a recent analysis of those seafloor maps showed that the ocean crust surrounding the ridge grew at a rate that ebbed and flowed on a timescale of hundreds of thousands of years. The results, published in *Geophysical Research Letters* (see bit.ly/oceanic-crustal-formation), will help scientists better understand the

processes that control ocean crust formation over millions of years.

Waves on the Seafloor

Most of the ocean floor is only very roughly mapped. “Earth is about 70% ocean floor,” said Joyce Sim, “but we really don’t know the bathymetry of the ocean floor, and it’s really hard to have this sort of high-resolution data on all of it.” Sim is a coauthor on the study and a geodynamics postdoctoral researcher at the Earth and Planets Laboratory at the Carnegie Institution for Science in Washington, D.C. “It takes quite a while just to cover a pretty small patch of the ocean” with mapping vessels and costs millions of dollars per expedition.

Moreover, the seafloor accumulates layers of sediment that can prevent sonar from seeing the ocean crust in detail. The area that was searched for MH370, however, has a thinner-than-average sediment buildup that makes the crust underneath visible to sonar.