

TOOLBOX

ECOLOGY'S REMOTE-SENSING REVOLUTION

Satellite data, and the tools that ecologists use to analyse them, are more accessible and plentiful than ever.

ILLUSTRATION BY THE PROJECT TWINS



BY ROBERTA KWOK

When ecologist Nicholas Murray started digging into remote-sensing data for his PhD project, he had no idea how hard his task would be. Murray wanted to know why shorebirds that migrate through Asia were declining in number. Because the birds stopped in places that were difficult for Murray to access, such as North Korea and China, he turned to satellite data to evaluate

their habitat.

When Murray started the project in 2010 he guessed it would take a few months, but it ended up taking about a year. Murray first had to download metadata for about 5,500 publicly available US government satellite images to identify those of tidal wetlands taken during low tide along the Yellow Sea, which borders China and the Korean peninsula. He then wrote custom software code to classify land cover in a final set of 80 images. An algorithm to

distinguish water from land already existed, but he needed to make manual adjustments for each image. More than one-quarter of the wetland area had vanished between the 1980s and 2000s, Murray discovered. But the analysis wasn't easy. "Throughout that whole process, I was thinking, 'This is so difficult, it's unbelievable,'" recalls Murray, now at the University of New South Wales in Kensington, Australia.

Today, Murray's task would be much simpler. Numerous tools have been developed to ►

► access and analyse remote-sensing data, allowing ecologists to tackle large-scale conservation problems more easily. Government agencies, open-source developers and commercial firms are offering everything from point-and-click interfaces to command-line-driven software. “I think we’re at the best time possible to be doing it,” Murray says of analysing satellite data for ecology research. “It’s getting so accessible.”

‘Remote sensing’ encompasses a suite of techniques for observing something without touching it. The term usually refers to collecting data about Earth from space or from airborne platforms by measuring energy reflected or emitted at various wavelengths. Researchers can use these data to infer, for example, the level of deforestation. “We’ve seen a real explosion in the use of satellite data,” says Allison Leidner, a contract senior support scientist at NASA’s Biological Diversity research programme in Washington DC.

Landsat data, gathered by NASA and the US Geological Survey (USGS), extend back to the 1970s and enable the study of planetary change over many decades. NASA’s Moderate Resolution Imaging Spectroradiometer (MODIS) instruments, launched in 1999 and 2002, measure reflected solar radiation and emitted radiation, and the data are automatically converted into ecologist-friendly parameters such as vegetation greenness. And Europe’s Sentinel satellites, which monitor the land, ocean and atmosphere, have been providing data since 2014.

Users can browse free government data sets at online portals such as NASA’s Earthdata Search, EarthExplorer from the USGS and the European Space Agency’s Copernicus Open Access Hub. Earth data are typically divided into sections called ‘scenes’ or ‘tiles’ — snapshots of energy of varying wavelengths reflected from that area. But to obtain higher spatial and temporal resolution, researchers might want to consider commercial options.

The Dove satellites operated by Planet in San Francisco, California, for example, gather global data at a resolution of 3.7 metres — sharp enough to distinguish individual large trees — about once a day. The Sentinel-2 satellites, by contrast, which are among the highest-resolution government satellites with free and open data, have 10-metre pixels and sample each spot every 5 days. University researchers can apply for free access to 10,000 square kilometres of Planet’s satellite data per month through the firm’s Education and Research programme. Similarly, academic environmental researchers can apply for free access to data from sub-metre-resolution satellites operated by DigitalGlobe in Westminster, Colorado, through the non-profit organization DigitalGlobe Foundation.

USER-FRIENDLY ACCESS

Data sets can be unwieldy, however. Kyla Dahlin, an ecogeographer at Michigan State University

in East Lansing, notes that 30 years of data collected for one Landsat scene could exceed 1.5 terabytes “for an area that’s smaller than Michigan”. Visualization software for remote-sensing data might not function well with certain file formats, and although these files can be converted into an easier-to-use format, that step adds another obstacle, says Cindy Schmidt, associate programme manager of Ecological Forecasting Applications at the NASA Ames Research Center in Moffett Field, California. Inexperienced users “just want to throw in the towel sometimes”, she says. “They don’t have time to deal with that kind of stuff.”

Free and commercial resources are available, however. In 2017, Murray’s team released a free online tool called Remap, which enables users to generate maps from remote-sensing data. Users train the software to classify land-cover types, such as forest or wetlands, by uploading geo-referenced data or identifying pixels on the basis of fieldwork or their knowledge. Remap then uses machine learning to classify the remaining pixels. As of March 2018, about 4,300 people from more than 100 countries have used Remap, Murray says. Another online tool, called Global Forest Watch, creates maps of deforestation patterns.

Dahlin recommends the online tool AppEEARS (Application for Extracting and Exploring Analysis-Ready Samples), which allows users to grab data specific to their study site, instead of an entire tile or scene. “Imagine the archive is this big lake of data,” explains Tom Maier-Sperger, project scientist at the NASA Land Processes Distributed Active Archive Center (DAAC) in Sioux Falls, South Dakota, which led the tool’s development. “We’re allowing people to come in with a syringe and suck up that little sample that they want.” Users can provide geographical coordinates, a time span and variables of interest — such as tree cover — and the software returns the data as a comma-separated-values (CSV) file.

Similarly, the US Oak Ridge National Laboratory DAAC, in Tennessee, has tools to provide for example a time series of greenness for a study site as a spreadsheet and graph, or processed data, such as inferred forest disturbance. Ecologists can then analyse links between vegetation and other variables such as animal populations. One team, for example, studied the Andaman Islands off the Indian coast and found that vegetation degraded more quickly in areas where elephants and spotted deer had been introduced.

For ecologists who want to write their own analysis software, but avoid the hassle of downloading satellite data, Google Earth Engine is a popular choice. Google has already downloaded satellite data sets onto its servers, and researchers can access them in the cloud for free through Google’s JavaScript and Python programming interfaces. This service allows researchers to perform large-scale analyses much faster than they could on their local computers.

Murray, for instance, leveraged that processing power to map global intertidal zones over time. Because it used more than 700,000 satellite images, the analysis would have taken years on a single computer — but it took less than a week on Google Earth Engine. The tool has “revolutionized the sorts of remote sensing questions I can ask”, Murray says.

Google says that users need not worry that it will claim ownership over their intellectual property (IP), such as code and scientific results. “Our terms of service make it clear that your IP is your IP and we make no claims on it,” says Noel Gorelick, an engineer at Google in Zürich, Switzerland, who co-developed Google Earth Engine. Still, Martin Wegmann, a remote-sensing researcher at the University of Würzburg in Germany, prefers to download satellite data and run his code locally. Because his analyses are relatively small in scale or coarse in resolution, performance is not an issue, he says.

Other cloud-computing options include the Centre for Environmental Data Analysis, run by the Science and Technology Facilities Council in Harwell, UK; Copernicus Data and Information Access Services, funded by the European Commission and scheduled to go live in June; and DigitalGlobe’s GBDX platform.

OPEN-SOURCE OPTIONS

Whichever platform they choose, researchers typically write custom code to drive data analysis, often in the programming language R. Wegmann and his colleagues are developing an R package called *getSpatialData*, which will allow users to download satellite data without using a browser interface. His team also developed the *RStoolbox* package, which includes different algorithms for computing vegetation measures so that users do not have to calculate specific formulae individually.

Researchers can also use commercial desktop analysis and visualization packages such as ENVI from Harris in Melbourne, Florida; ERDAS IMAGINE from Hexagon Geospatial in Madison, Alabama; ArcGIS from Esri in Redlands, California, as well as free, open-source alternatives such as QGIS.

Because using these tools can involve steep learning curves, Anita Graser, a geographic information scientist at the Austrian Institute of Technology in Vienna and a member of the QGIS project steering committee, advises beginners to take online classes. NASA’s Applied Remote Sensing Training programme offers webinars, and the agency gives workshops at ecology and conservation conferences.

The possibilities are enticing, but researchers must remember to stay grounded. “If you wanted to see how often a butterfly visits a nectar plant, you’re not going to pick that up on a satellite,” Leidner says. But for larger-scale problems, “it’s an incredibly powerful tool”. ■

Robert Kwok is a freelance writer in Kirkland, Washington.