Microsoft

# AI for Earth Grantee Profile

## Columbia University

## Large-scale forest ecological surveys


## Summary

Professors Tian Zheng and Maria Uriarte at Columbia University are using ground observations of forest plots to create a machine learning pipeline that's capable of correctly classifying the species of individual trees using aerial photographs and LiDAR data collected by NASA using remote sensing technologies, in order to better understand how storms affect a forest's ability to store carbon and aid in climate change mitigation, and how damaged forests recover over time.


## Keeping a close watch on our forests, for our future

Trees play an important part in the global carbon cycle, drawing a large amount of carbon dioxide from the air and using the carbon to build their tissues. Every year, tropical forests worldwide remove up to 2.8 billion tons of carbon from the atmosphere—a great amount, but only about a fifth of the carbon that human activity is producing.

> Every year, forests worldwide remove up to 2.8 billion tons of carbon from the atmosphere.

Powerful storms such as hurricanes can seriously disrupt a forest's ability to absorb carbon. When trees are knocked down by storms and decompose, most of their carbon gets released back into the atmosphere. New trees that spring up in their place grow more quickly and could potentially make up the difference. The mix of tree species also often changes after a storm, as trees that thrive in full sunlight are more likely to fill in the gaps where the larger old growth trees fell down.

Meanwhile, the frequency of large and powerful storms is expected to increase, due to climate change. Extreme weather previously considered to be hundred-year events could happen as often as every ten or twenty years. In such a case, many forests might never be able to fully recover from the effects of storms, permanently reducing their ability to store carbon. However, more data and analysis of how forests currently recover from these storms is required for scientists to understand the impacts.

## Combining ground work and sky surveys to understand the forests better

Professors Tian Zheng and Maria Uriarte, members of the Data Science Institute at Columbia University, are using their AI for Earth grant to apply an innovative technological approach, based on machine learning, to study those storm effects. Specifically, they are using ground observations of forest plots to create a machine learning pipeline that's capable of correctly classifying the species of individual trees in high-resolution aerial images, to better understand how damaged forests recover over time and how storms affect the forests' ability to store carbon and aid in climate change mitigation.

> Zheng and Uriarte are using their AI for Earth grant to apply machine learning to study the effects of powerful storms on forests.

Collecting sufficient data about a forest's condition is very challenging, even without dealing with storm damage. Between the sheer number of individual trees of various species and the often-difficult terrain and undergrowth to traverse, it's a time-consuming and laborious process simply to stake out and detail small plots to use as samples, and a practical impossibility to map and measure an entire forest. Furthermore, the mix of tree species in any one plot may not reflect the overall diversity and distribution in the forest as a whole, so multiple plots are needed, adding to the manual labor.

To better understand the forests, the effects of major storms, and the impact of climate change, scientists need a bigger picture—literally as well as figuratively. Aircraft surveys and satellite imagery can provide a more comprehensive view of forests and show in broad terms the effects of a major storm, but recent advances in remote sensing and machine learning offer more than that. New techniques such as Light Detection and Ranging (LiDAR) and stereo photos provide large-scale high-resolution data that would allow researchers to measure key factors mediating storm impacts, such as the overall structure of forests and the local topography. Unlike plot-based approaches, remote sensing allows monitoring and assessment of storm damage to forests across large areas. However, the remote sensing alone would be insufficient without some way to ground that data, which is where the machine learning comes in. In combination with samples of high-accuracy georeferenced tree data from ground observations, a data science workflow based on machine learning models will enable ecological surveys of forests at an unprecedented scale.

## Enabling ecological surveys at an unprecedented scale

Professor Uriarte has been studying the forest ecology of Puerto Rico for the past 15 years, mapping and measuring thousands of trees in multiple plots around the island. Hurricane Maria, which hit Puerto Rico and

the U.S. Virgin Islands in September and October of 2017, caused widespread damage to forests across the islands. With years of pre-hurricane data to work with, Professors Uriarte and Zheng and their team are in a unique position to assess the damage of Hurricane Maria by more efficiently processing images and LiDAR data collected by NASA before and after the storm—and use this as a test case to develop a better methodology empowered by AI.

The professors are building a data science workflow based on Microsoft Azure to carry out a virtual ecological survey of tree species over large areas, using imaging and remote sensing data as well as ground observations from field plots in Puerto Rico. By applying machine learning tools to that imaging data and correlating it to the ground observation data, they can infer the ecological distribution of various tree species during different time points—before and after Hurricane Maria and through the recovery period. The workflow constitutes feature engineering for high resolution images and LiDAR data using deep representational learning tools (that is, convolutional neural networks); training of a deep learning model for species classification at the pixel level; and species segmentation of the surveyed area based on pixel-level classification.

The team is starting by constructing high-quality labelled training data and developing the computational tools required to process the large amount of imaging and remote sensing data. This includes image processing and LiDAR data processing; applying existing image semantic segmentation tools to establish a performance baseline; creating image snippets near root locations of trees with known species information; using LiDAR data to improve tree species labels; and developing the proposed workflow based on labeled data. Up to 5 terabytes of storage on Microsoft Azure will be used to host the high-resolution imaging and sensor data, as well as the computational results. Microsoft Azure Data Science Virtual Machine and Deep Learning Virtual Machine will provide both computational resources and the state-of-the-art data science building blocks to accelerate the research.

### Going forward

Once the data science workflow is functional, the team can begin working on answering the following questions: 1) How do landscape characteristics (topography, slope, aspect, geology, and forest age and fragmentation) influence wind damage and biomass loss after a severe hurricane? 2) What is the selective pressure of hurricane damage on forest composition? 3) How does landscape heterogeneity influence recovery of species?

## About Tian Zheng and Maria Uriarte

Tian Zheng is a Professor of Statistics and Associate Director for Education at Data Science Institute, Columbia University. She develops novel methods for exploring and understanding patterns in complex data from different application domains such as biology, psychology, climatology, etc. Her current projects are in the

fields of statistical machine learning, spatiotemporal modeling and social network analysis. Her work was recognized with the 2008 Outstanding Statistical Application Award from the American Statistical Association (ASA), the Mitchell Prize from ISBA and a Google research award. She became a Fellow of American Statistical Association in 2014.

Maria Uriarte is a Professor of Ecology, Evolution, and Environmental Biology at Columbia University. Through an integrated program of empirical and quantitative approaches, research in her lab examines forest ecological dynamics in response to natural disturbance and human land use. Research projects focus on disturbance ecology, forest succession, and community assembly. Field sites span geographic regions where forests have been subject to different forms of anthropogenic disturbance, including fire, hurricanes, fragmentation, and expansion of tree plantations. Her current projects include studies of the dynamics of post-agriculture and post-hurricane forest recovery in landscapes of Puerto Rico and the effects of fires in agriculture-forest mosaics in the Peruvian Amazon.

# Resources

### Websites

Maria Uriarte's page on the Columbia University's Data Science Institute site
Tian Zheng's page on the Columbia University's Data Science Institute site

### Press

Lei, Bonnie, and Jennifer Marsman. "Species around the world are in trouble, but we can help." *Microsoft Green Blog*. Microsoft. April 18, 2018.

Krajick, Kevin. "Climate Change and the Re-Greening of Puerto Rico." *State of the Planet*. Earth Institute, Columbia University. March 22, 2018.

Fountain, Henry. "Forests Protect the Climate. A Future With More Storms Would Mean Trouble." *New York Times*. March 7, 2018.

Florida, Robert. "DSI Researchers Win Microsoft AI for Earth Grant." Data Science Institute, Columbia University. February 22, 2018.