Microsoft

# AI for Earth Grantee Profile

## Heather Lynch

## Tracking Antarctic penguin populations

## Summary

Dr. Heather Lynch is working at the intersection of geography, remote sensing, statistical modelling, and ecology to map the distribution and abundance of Antarctic penguins. She is using computer vision and deep learning to help train computers to identify penguin colonies and estimate penguin populations based on guano stains visible in satellite imagery.

## Tracking Antarctic penguin populations in real time

Antarctica and its surrounding oceans are home to more than 8,000 marine species, more than half of which are seen nowhere else in the world. However, this remote and pristine wilderness is under threat from the impacts of climate change, with melting ice and warming temperatures changing the habitat and feeding grounds for local wildlife populations, including penguins.

> Models of Adélie penguin populations that account for climate change predict 60 percent of colonies will be declining by 2099.

Adélie penguins are one of the most well-studied of Antarctic species, and yet most of their nesting areas are so remote and inaccessible that scientists cannot monitor their populations. Tracking their abundance and distribution is particularly critical now. There is tremendous concern that climate change may reduce the quality and availability of their food, and that increases in precipitation may increase mortality rates among vulnerable chicks. In fact, while Adélie populations have risen over the last 40 years, models that incorporate forecasts of climate change in Antarctica predict population declines on the horizon affecting approximately 30 percent of colonies by 2060 and approximately 60 percent by 2099.

While much remains unknown about how Adélie penguins might fare in the future, new technologies are radically changing the ability of scientists to track their populations. Satellite imagery has quickly become an essential tool for finding penguin colonies and estimating abundance, but imagery interpretation has, up to now, required extensive manual annotation by trained experts. The exceptionally time-consuming nature of imagery interpretation has made it difficult to do regular surveys across all of Antarctica's hundreds of Adélie colonies. New methods in deep learning are showing tremendous potential for automating the interpretation of high-resolution commercial satellite imagery, but in the case of penguin colonies, the contrast between guano stains (the visual signature for a colony) and bare rock is often subtle. Furthermore, the size of the training set is unavoidably limited by the existing number of penguin colonies.

### Using computer vision to identify penguin colonies in satellite imagery

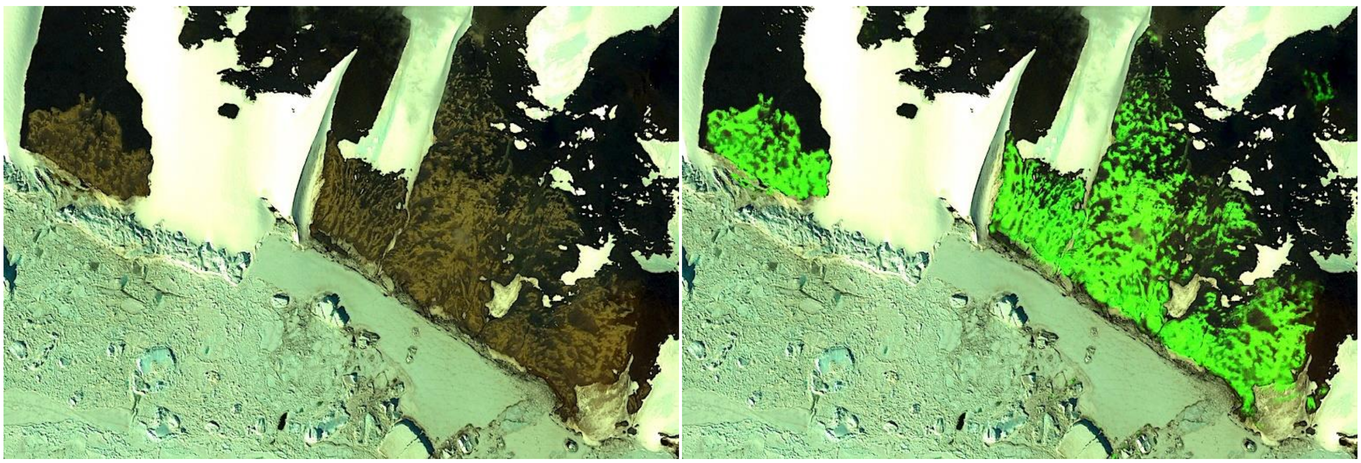
Dr. Heather Lynch, a quantitative ecologist and IACS Endowed Chair of Ecology & Evolution at Stony Brook University, is working to overcome these challenges. Dr. Lynch has spent over a decade modeling every aspect of penguin population dynamics and developing the underlying methods required to use satellite imagery for tracking penguin populations. The next big step is to use computer vision to speed up imagery annotation, since that is a necessary component of any real-time monitoring system. "Right now, there's so much imagery and there are so many colonies that there's just no way that we can manually annotate all this imagery," says Dr. Lynch. "Even if we had the theoretical capacity to track all these populations, we don't have the human capacity to do it."

> "There's so much imagery and there are so many colonies that there's just no way that we can manually annotate all this imagery."—Dr. Heather Lynch

Thanks to an AI for Earth Innovation Grant from Microsoft and the National Geographic Society, Dr. Lynch and her colleagues have been building on their initial work (supported by the National Science Foundation) to facilitate imagery annotation with machine learning models. The team has been developing more sophisticated classification algorithms that improve accuracy by accounting for our existing knowledge about penguin biology, population growth rates, habitats, and colony fluctuations through time. For example, Dr. Lynch explains, "When I'm looking at a satellite image, I know that very weathered rocks at the top of a ridge look very similar to guano, and as a human annotator I just dismiss that out of hand, because I know there's not guano on the top of that mountain. But the computer vision model doesn't have a lot of intuition on where penguins should be or shouldn't be, so it trips up the computer in a way that it doesn't trip up a human."

Building that level of sophistication into the algorithms took some groundbreaking work. Dr. Lynch says of the project, "The challenge of identifying guano in satellite imagery turned out to be much harder—and correspondingly more interesting—than we thought. It's required us to develop entirely new computer vision approaches, and we've published some of those methods in a computer-vision conference this year."

These next-generation algorithms are capable of using a wider range of training data required for imagery annotation, including relatively low-resolution images from the NASA Landsat satellite system (which was responsible for the recent discovery by Lynch and collaborators of a major penguin "hotspot" on the Antarctic Peninsula). "One of the reasons we had to develop new methods is that we just were never going to have a training dataset that was as large as we wanted," explains Dr. Lynch. "One of the things we did was use Landsat imagery, which is much coarser resolution imagery, but we developed a method of using this really coarse resolution imagery to train the model." The algorithms will also have applications for a variety of imagery sensors from commercial vendors such as DigitalGlobe and Planet.



*Lynch developed an API demonstrating model prediction of penguin guano, highlighted in green, in an image. [Images from Google Earth]*

This project represents a compelling application of machine learning for Earth observation that will enhance an existing framework for modeling and decision support within the Antarctic community. The Microsoft Azure cloud computing resources made available through the AI for Earth Innovation Grant have provided vital support for this computationally intensive work. Azure also provides a platform to host a version of the classification tool for public use. Dr. Lynch says, "We want to give the public an opportunity to try out our algorithm for finding penguins in satellite imagery, so we created a gallery of publicly available images to show how our model finds the penguin guano, and then allow people to upload and annotate their own images and see whether the model agrees with them."

### Streaming population data in real time to support conservation decision making

This project is a critical step toward fully integrating dynamic time-series modeling, population forecasting, image classification, and decision support for conservation and natural resources management. Dr. Lynch's ultimate goal is to provide real-time information to scientists and the policy community about changing penguin population numbers and habitats. The Mapping Application for Penguin Populations and Projected

Dynamics ([MAPPPD]) she has created is a first step, but its reliance on "slow" data such as field counts and manually annotated satellite imagery prevents the kind of real-time data feed that she envisions. "We're working towards that full automation where people will undoubtedly continue to work on the algorithms and improve them, but the whole cycle can happen without a human having to do anything," says Dr. Lynch. "We're very close to it literally going from satellite through to population estimate and onto the web where policymakers can find it, without having to pass through the hands of people."

Dr. Lynch believes that this will open new doors to understanding the biggest threats to the species and how we can best respond. As she puts it, "I would not get in my car and drive to New York City without looking at my phone and knowing where the accidents are and approximately how long it will take to get there. Today, we take this kind of streaming data service for granted. But when it comes to ecological problems, we don't yet have anything like that. With this penguin project, we have one of the first real examples where we can automatically get and process satellite imagery, generate population estimates, and put that information in decision support tools—making it as readily available as a weather forecast."

## About Dr. Heather Lynch

Dr. Heather Lynch is a quantitative ecologist and IACS Endowed Chair of Ecology & Evolution at Stony Brook University with a joint appointment in the Department of Ecology & Evolution and the Institute for Advanced Computational Science. Her research is dedicated to understanding the population dynamics of Antarctic

*Dr. Lynch's research is dedicated to understanding population dynamics of Antarctic wildlife.  Photo courtesy of Heather Lynch.]*

wildlife, with a particular focus on Antarctic penguins, and she has more than a decade of field experience in Antarctica as co-principal investigator (PI) of the Antarctic Site Inventory project.

Dr. Lynch helped pioneer the use of satellite imagery for studying the distribution and abundance of Antarctic seabirds and published the first Antarctic-wide satellite-based surveys of both Adélie penguins and Antarctic petrels. She serves as PI for a large, multi-institution National Science Foundation (NSF) award tasked with building the cyberinfrastructure required to unite high-resolution commercial imagery and high-performance computing for imagery-enabled science in the polar regions. Dr. Lynch has served two terms on the Science & Operations Committee of the University of Minnesota's Polar Geospatial Center and is currently that committee's chair. She received an NSF CAREER Award for her work on the spatial dynamics of Antarctic penguins and was elected an early career fellow of the Ecological Society of America. Dr. Lynch has an AB in physics from Princeton University, an MA in physics from Harvard University, and a PhD in organismic and evolutionary biology from Harvard University.

# Resources

### Press

Microsoft and National Geographic Society announce AI for Earth Innovation grantees

### Websites

Dr. Lynch's home page with Stony Brook University
Lynch Lab for Quantitative Ecology site
MAPPPD (Mapping Application for Penguin Populations and Projected Dynamics) site
AI for Earth penguin guano classification API demo

### Publications

Lynch, H.J., White, R., Black, A.D. et al. "Detection, differentiation, and abundance estimation of penguin species by high-resolution satellite imagery." *Polar Biology* (2012) 35: 963. https://doi.org/10.1007/s00300-011-1138-3

H. J. Lynch, M. A. LaRue. "First global census of the Adélie Penguin." *The Auk: Ornithological Advances*, Volume 131, Issue 4, pages 457–466. October 1, 2014. https://doi.org/10.1642/AUK-14-31.1

Le, Hieu & Gonçalves, Bento & Samaras, Dimitris & Lynch, Heather. (2019). "Weakly Labeling the Antarctic: The Penguin Colony Case."

### Documentation

"Tracking Antarctica." National Geographic. October 18, 2016.
https://blog.nationalgeographic.org/2016/10/18/tracking-antarctica/

Christian Che-Castaldo et al. “Pan-Antarctic analysis aggregating spatial estimates of Adélie penguin abundance reveals robust dynamics despite stochastic noise.” *Nature Communications* 8, Article number: 832 (2017). October 10, 2017. https://www.nature.com/articles/s41467-017-00890-0

“Climate Change May Shrink Adélie Penguin Range By End of Century.” NASA. July 8, 2016. https://www.nasa.gov/feature/goddard/2016/climate-change-may-shrink-adelie-penguin-range-by-end-of-century