Demonstrating cloud-based remote sensing data workflows with **xarray**

Emma Marshall
University of Utah, National Center for Atmospheric Research

MOTIVATION

Recent advances in **satellite imagery availability**, **cloud-computing resources** and **open-source software** necessitate the need for **detailed, accessible educational resources** in order to fully realize the **public benefit** and **scientific potential** of these resources.

Mentors: **Deepak Cherian** (NCAR), **Scott Henderson** (University of Washington), **Jessica Scheick** (University of New Hampshire), **Kevin Paul** (NCAR)

BACKGROUND

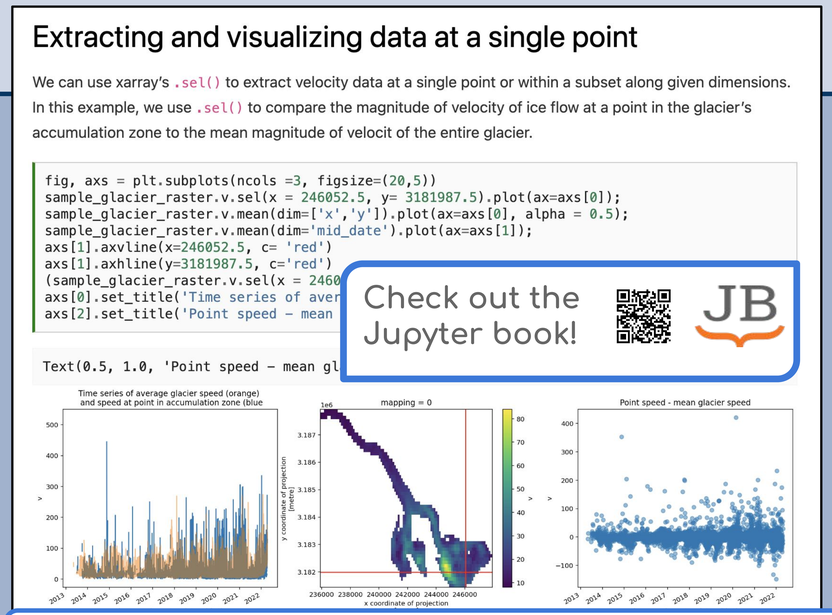
- Cloud computing resources can democratize scientific participation, reduce computational barriers to entry
- Accessing, manipulating data is a common bottleneck in remote sensing research workflows
- Xarray has functionality for organizing, analyzing raster data, and backend integration with cloud-optimized data types

OBJECTIVES:

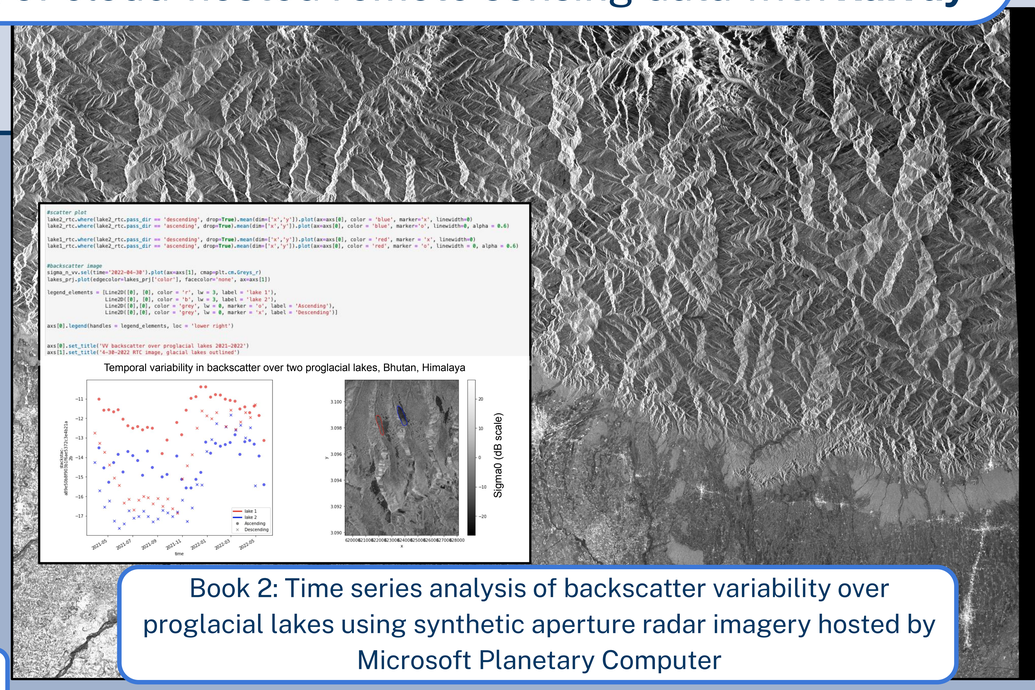
- Gain skills and experience working with cloud computing resources and parallelizing workflows
- Develop **educational resources** and make **open-source contributions** to support the processing of cloud-hosted remote sensing data with **xarray**

JUPYTER BOOK TUTORIALS

- Show data access steps for multiple cloud-hosting platforms
- Use accessible, explanatory text
- Include errors and solutions (not just showing what works!)
- Focus on manipulating data to form data cubes
- Demonstrate parallelized workflows
- Emphasizing and supporting open, reproducible science



Book 1: Accessing and analyzing ITS_LIVE ice velocity dataset stored in S3 buckets on Amazon Web Services



Book 2: Time series analysis of backscatter variability over proglacial lakes using synthetic aperture radar imagery hosted by Microsoft Planetary Computer

OPEN SOURCE CONTRIBUTIONS & INVOLVEMENT

Netcdf data cleaning example for xarray-tutorial, documentation contributions to xarray-contrib, dataset contribution to xarray-pydata, co-presented @ SciPy 2022 xarray tutorial, Austin, TX July 2022

WHAT DID I LEARN?

- Collaborative workflows
- Vectorized and xarray-native programming
- Working with cloud-hosted data and cloud computing resources
- Better data sharing practices
  - Jupyter book, github gists, notebook{sharing}.space

ITS_LIVE book github repo:

THE UNIVERSITY OF UTAH
NCAR
NASA
SIParCS
Research Poster Template