Developing educational resources for cloud -based remote sensing data with xarray

Emma Marshall
SIParCS intern

July 26, 2022

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsor ed by the National Science Foundation under Cooperative Agreement No. 1852977.

Who am I?

PhD student
Snow & Ice Research Lab
Geography Department
University of Utah

Research interests:

- Mountain glaciers in High Mountain Asia
- Lake-terminating glaciers
- Remote sensing -derived glacier surface velocity data

Background

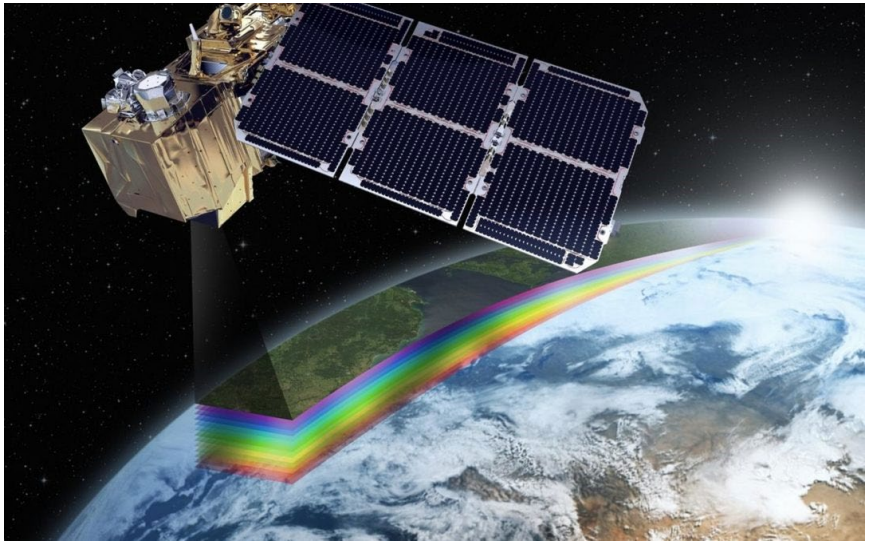
I. Satellite remote sensing data

II. Xarray for multi -dimensional gridded datasets

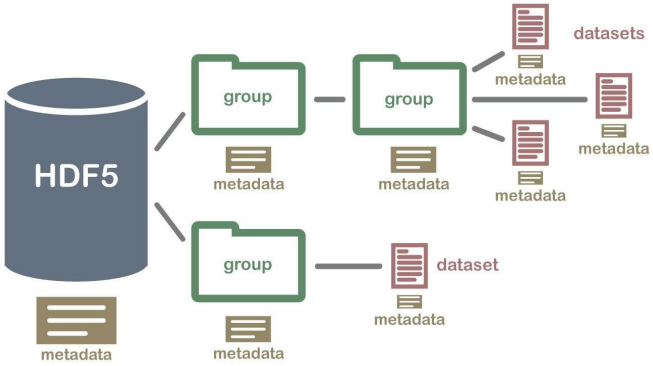
III. Transition to open, cloud -based science

Satellite remote sensing data

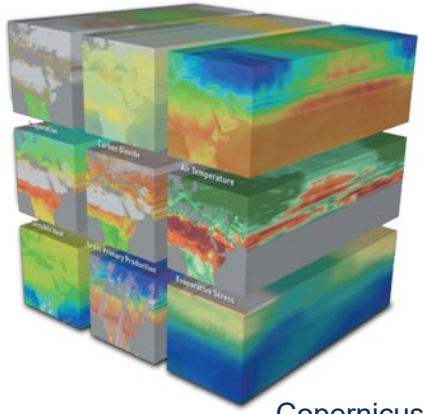
- Remote sensing datasets are large, complex; increasing volume of available data
- Multi -variate, multi - dimensional, metadata



Cervest

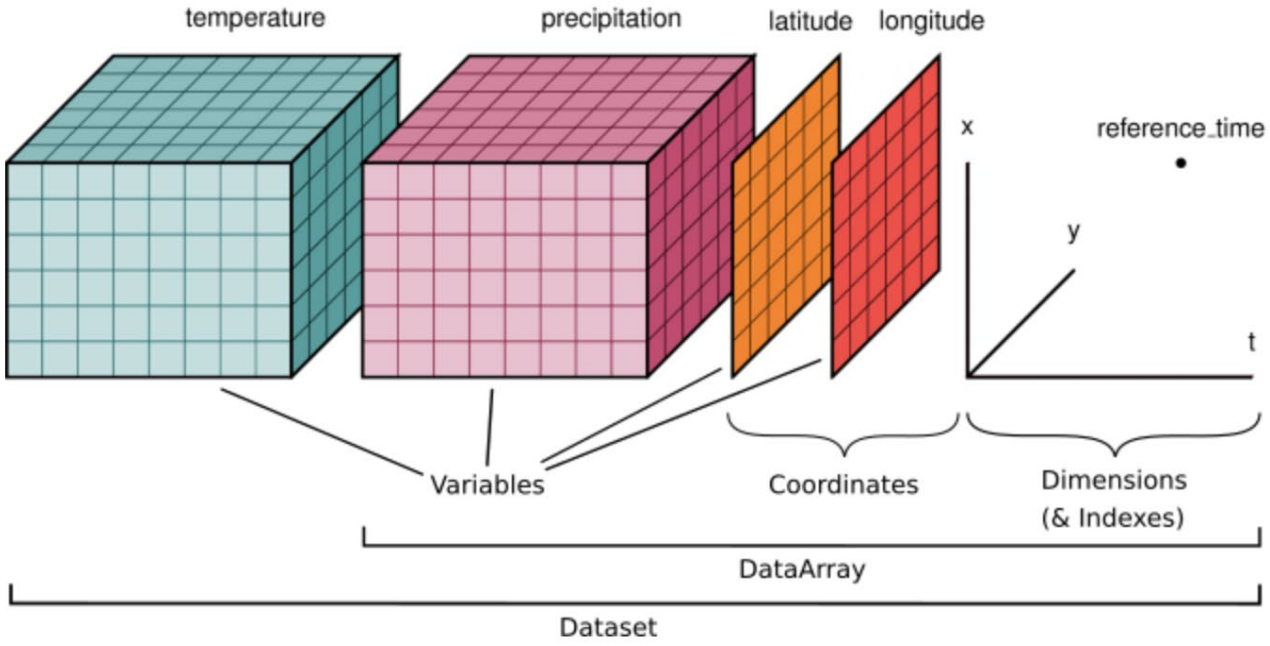


NEON



Copernicus

Xarray



"Xarray is an open source project and Python package that makes working with labelled multi -dimensional arrays simple, efficient, and fun"

Xarray has functionality for organizing, analyzing raster data, and backend integration with cloud -optimized data types

Transition to cloud -based, open science

Accessing, manipulating data is a common bottleneck in remote sensing research workflows

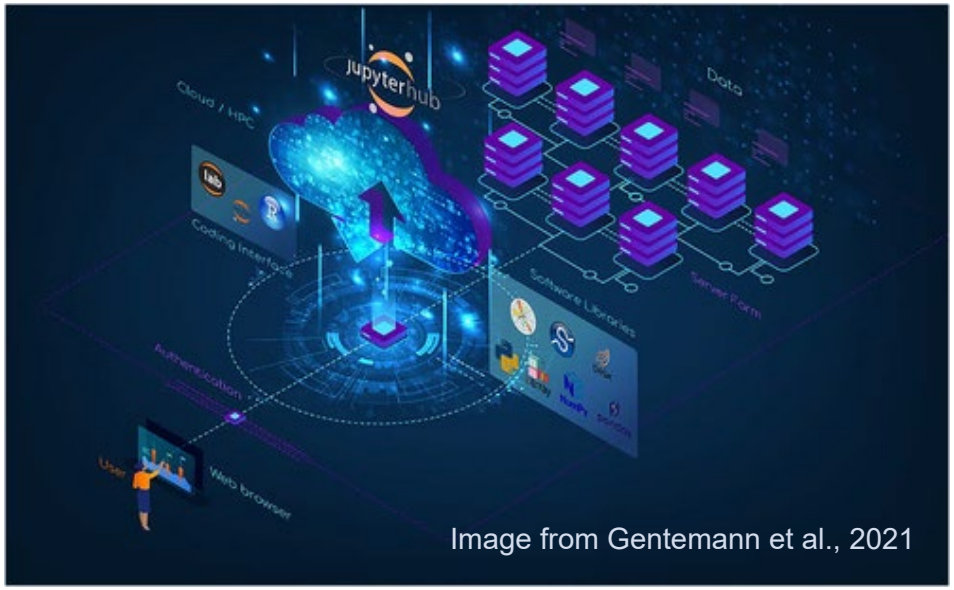


Image from Gentemann et al., 2021

"... [the transition to open science] ... **is a product of practices, norms, and community behavior around that technology… it is important to deliberately design a new community infrastructure** ..." Gentemann et al., 2021

Cloud computing resources can democratize scientific participation, reduce computational barriers to entry

Objectives

I. Gain experience working with cloud -hosted data, parallelized workflows and xarray
II. Contribute educational resources related to xarray, remote sensing data and cloud - computing resources to the open -source scientific community

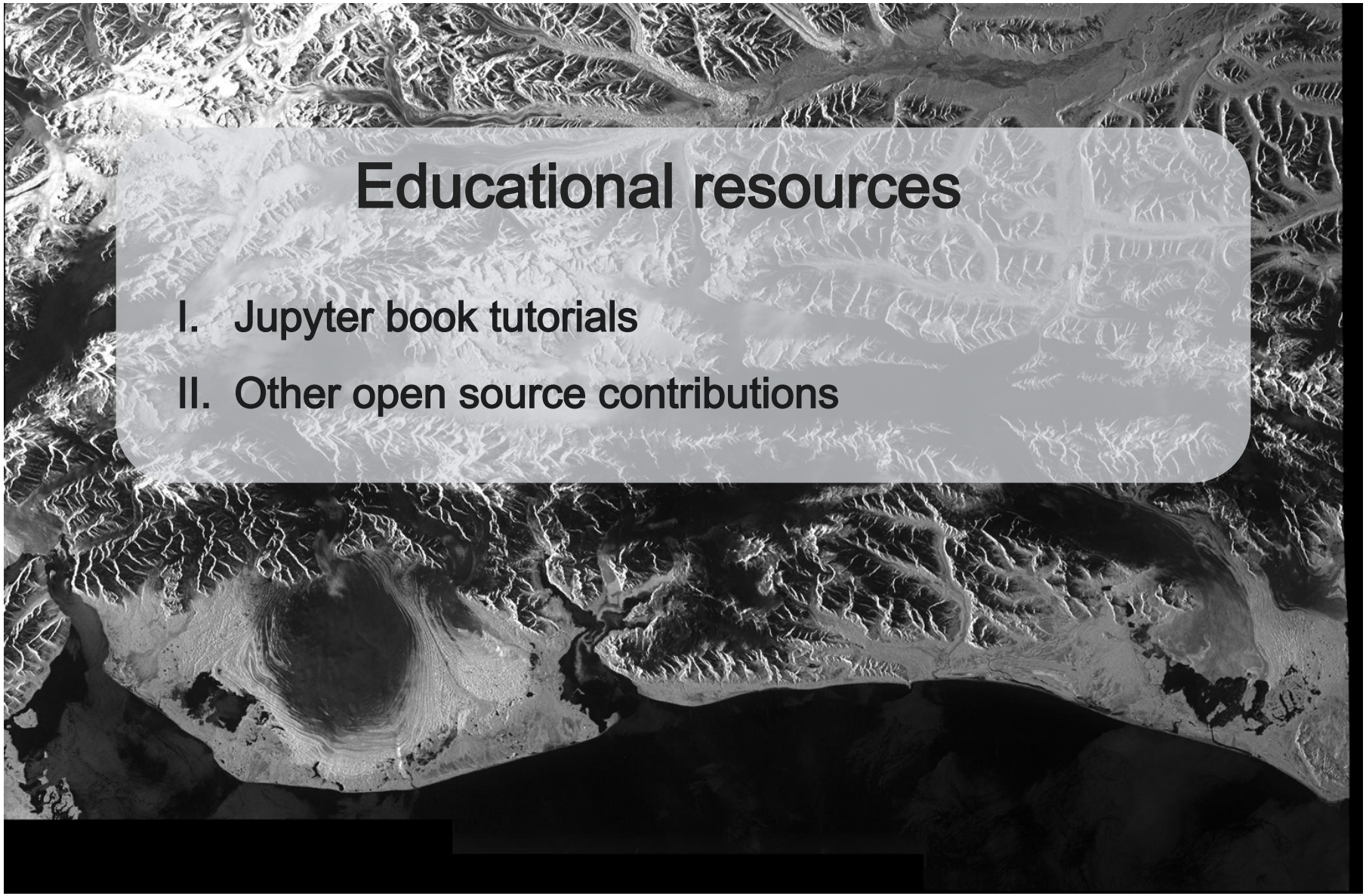
Cloud computing and software resources

Cloud platforms

- Amazon Web Services
- Microsoft Planetary Computer
- Pangeo Jupyter hub
- Alaska Satellite Facility On Demand processing -

Open source software

- xarray
- dask



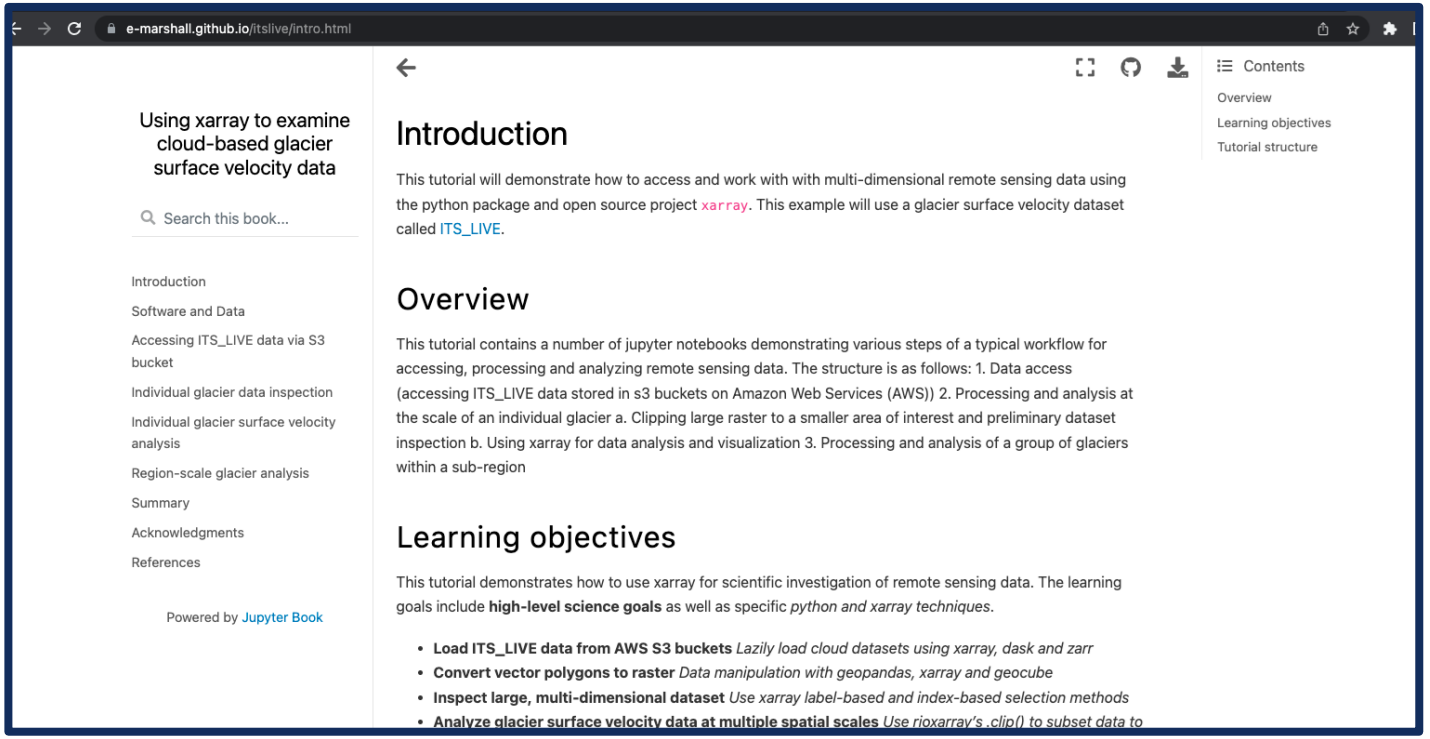
Educational resources

I. Jupyter book tutorials

II. Other open source contributions

Jupyter book 1:

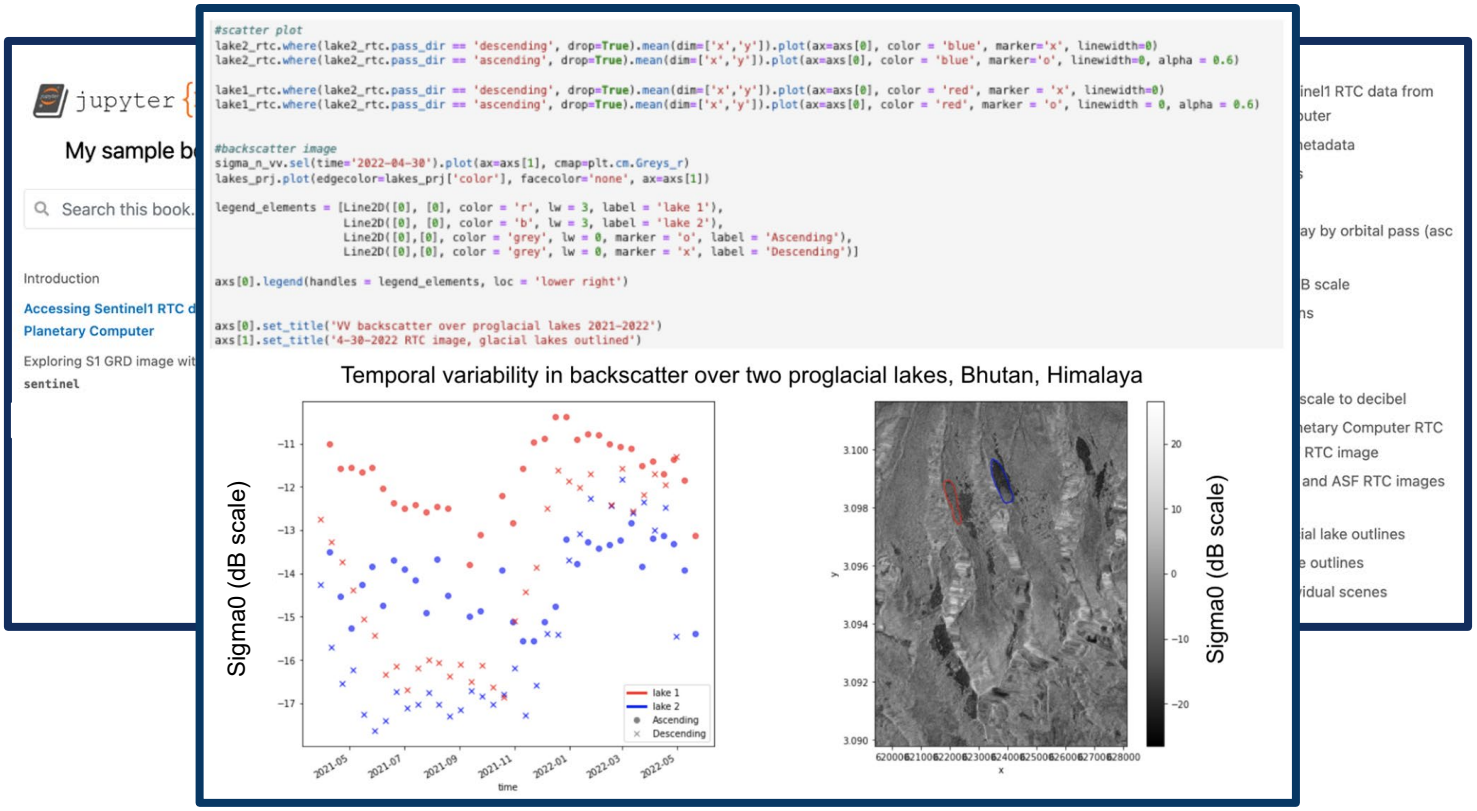
ITS_LIVE glacier velocity data hosted in AWS S3 buckets



https://e -marshall.github.io/itslive/intro.html

Jupyter book 2:

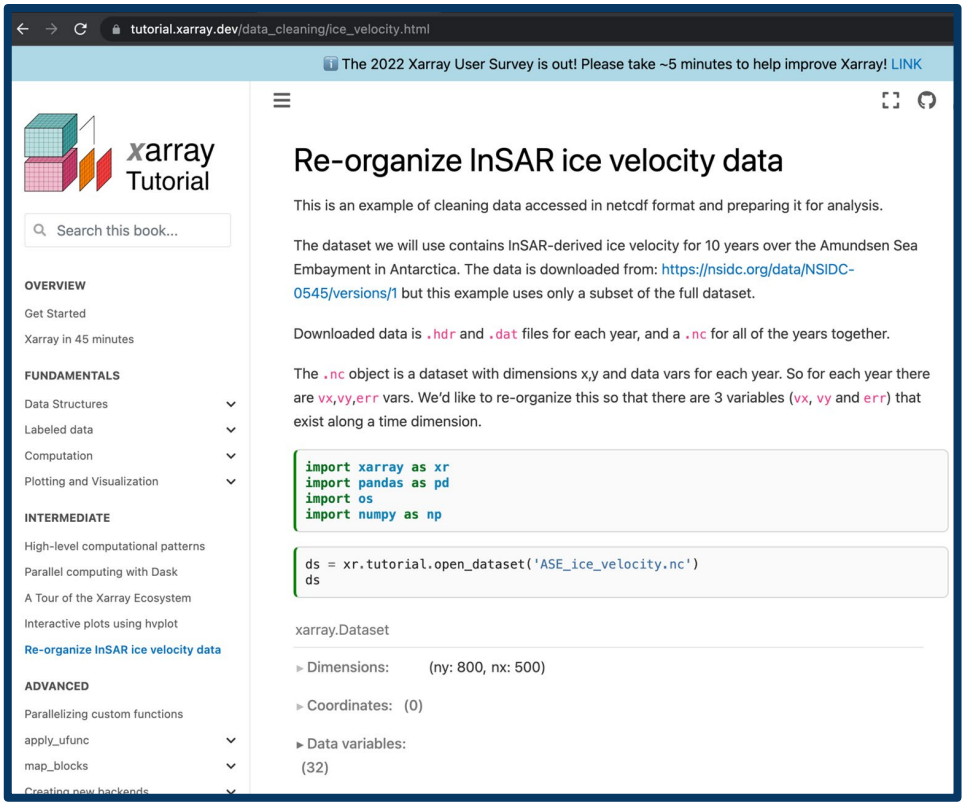
Sentinel1 Radiometric Terrain Corrected backscatter data



https://e -marshall.github.io/sentinel1_rtc/intro.html

Other open source contributions

- Data cleaning example for xarray tutorial
- Added examples to xarray codebase
- Co-presented at xarray tutorial during SciPy 2022 in Austin, TX (July 2022)



https://tutorial.xarray.dev/data_cleaning/ice_velocity.html

What did I learn?

- Collaborative workflows
- Vectorized and xarray - native programming
- Parallelized workflows
- Cloud computing
- Tutorial design + construction
- Introduction to the open source scientific software community

What's next?

- Finish incorporating new chapters in the Sentinel1 book
- Share work in open science session at Fall conference
- Pangeo / xarray blog posts
- Continue to contribute to open source packages !

Thank you:

Deepak Cherian (NCAR)
Scott Henderson (University of Washington)
Jessica Scheick (University of New Hampshire)
Kevin Paul (NCAR)
Virginia Do, Jerry Cyccone, everyone at CISL and across NCAR, and especially all of the NCAR/UCAR interns

References

1. Gentemann, C. L., Holdgraf, C., Abernathey, R., Crichton, D., Colliander, J., Kearns, E. J., et al. (2021). Science storms the cloud. *AGU Advances*, 2, e2020AV000354. https://doi.org/10.1029/2020AV000354
2. Stern C, Abernathey R, Hamman J, Wegener R, Lepore C, Harkins S and Merose A (2022) Pangeo Forge: Crowdsourcing Analysis -Ready, Cloud Optimized Data Production. Front. Clim. 3:782909. doi: 10.3389/fclim.2021.782909
3. NASA turns to the cloud for help with next -generation Earth missions. Sea Level News. October 13, 2021. Sea level change: observations from space. NASA. https://sealevel.nasa.gov/news/226/nasa -turns -to -the -cloud -for -help -with -next -generation -earth -missions/

Images

Satellite Imagery

Alaska Satellite Facility Vertex portal. Scenes:

S1B_IW_GRDH_1SSV_20170213T000216_20170213T000241_004274_0076A5_CA26

S1A_IW_GRDH_1SDV_20220720T155707_20220720T155734_044186_054625_C18A

S1A_IW_GRDH_1SSV_20170511T155632_20170511T155659_016536_01B692_3800

S1A_IW_GRDH_1SSV_20170530T154815_20170530T154839_016813_01BF22_D7E9

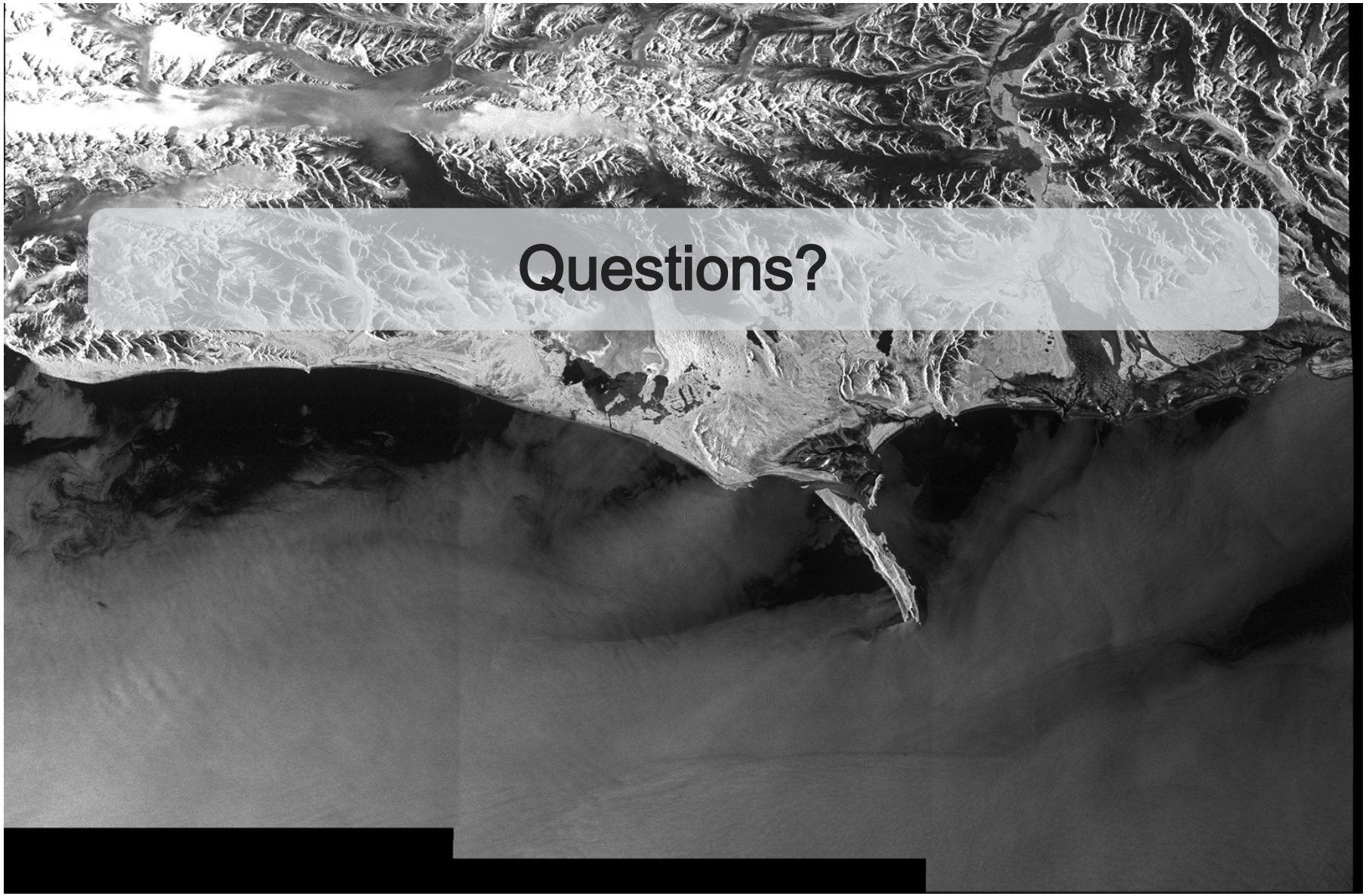
Other images

https://cervest.earth/news/remote-sensing-of-planet-earth-part-1-introduction-to-satellite-imagery

https://www.copernicus.eu/en/news/news/observer-data-cubes-enabling-and-facilitating-earth-observation-applications

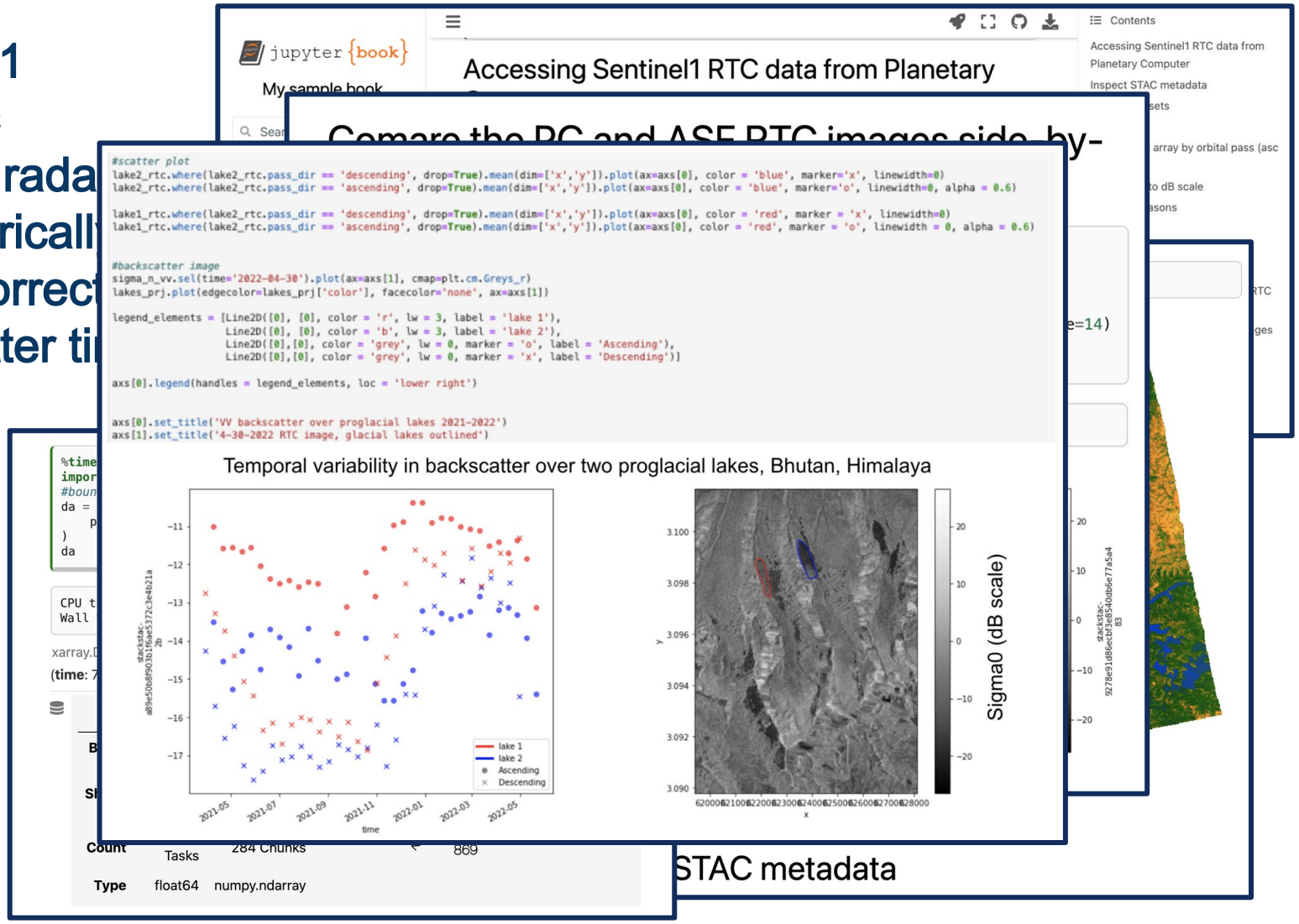
https://www.neonscience.org/resources/learning-hub/tutorials/about-hdf5

https://docs.xarray.dev/en/stable/


Questions?

Jupyter book tutorials

Sentinel 1 synthetic aperture rada radiometricall terrain correct backscatter ti series



What did I learn?

- Collaborative workflows
- Vectorized and 'xarray -native' programming
- Parallelized workflows
- Cloud computing
- Tutorial design + construction
- Introduction to the open source and open science community

Thank you to my internship mentors this summer for their guidance, advice and support:

Deepak Cherian (NCAR)
Scott Henderson (University of Washington)
Jessica Scheick (University of New Hampshire)
Kevin Paul (NCAR)
Thank you to
Virginia Do and the SIParCS and NESSI programs,
all UCAR interns!

What's next?

- Finish incorporating new chapters in the Sentinel1 book
- Share work in open science session at Fall conference
- Pangeo / xarray blog posts
- Continue to contribute to open source packages !

Other open source contributions

- data cleaning example for xarray tutorial
- added examples to xarray codebase
- assisted at xarray tutorial during SciPy 2022 in Austin, TX (July 2022)

What's next?

- Finish incorporating new chapters in the Sentinel1 book
- Share work in an Open Science session at AGU
- Present Sentinel1 jupyter book at NISAR science conference, August 2022, Pasadena, CA
- Publish jupyter books in the journal of open source education?
- Pangeo / xarray blog posts
- Continue to contribute to open source packages !

be more general, trim down

Jupyter book tutorials

Glacier velocity data hosted in AWS S3 buckets

executable book with binder
- cloud hosted data and execution
- open science!