G23C-0487

Practical cloud computing for InSAR

Scott Henderson (scottyh@uw.edu)[1], Eric Gagliano[1], Emma Marshall[2], George Brencher[1], David Shean[1]
[1]University of Washington Earth & Space Sciences & Civil Engineering
[2]University of Utah Geography

AGU23 WIDE. OPEN. SCIENCE.

UNIVERSITY of WASHINGTON eScience Institute

Motivation

* NASA is moving Petabyte-scale archives of satellite data (including Sentinel-1 & NISAR) to commercial cloud

* Government & commercial SAR operators are increasingly adopting cloud-hosted archives & standard data formats - COG+STAC

* The transition to the cloud makes sharing data, code, and even computational infrastructure easier

* This poster highlights examples and practical steps for individual researchers to take advantage of cloud-hosted SAR data and computing resources

Start here: standard products

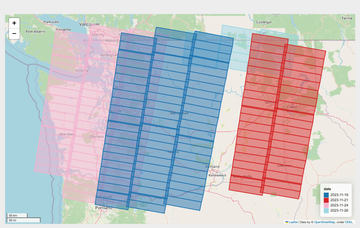
Gone are the days of downloading 4+GB SLCs! Geocoded SAR products are becoming more widely available - such as radiometric terrain-corrected backscatter (RTC), and co-registered single single look complex data (CSLC).

Example: OPERA-ISCE3 RTC & CSLC v1

OPERA

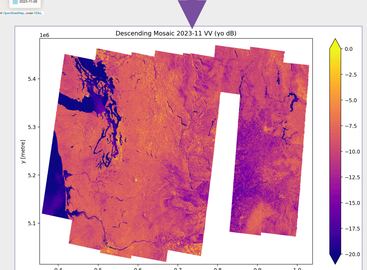
Newly released co-registered SLCs for fast and flexible interferogram formation!

- Bursts ~20x80km
- North America only
- Sentinel-1 from October 2023 onwards
- Processed with Copernicus 30m DEM
- CSLC 5x10m posting, H5 format
- RTC 30m posting, COG format



ASF OpenSARlab

xarray dask



-182 bursts to 90m monthly mean mosaic in ~1min

- Can use ASF OpenSARLab server (8vCPU, 16GB) next to data in AWS us-west-2

https://github.com/scottyhq/opera

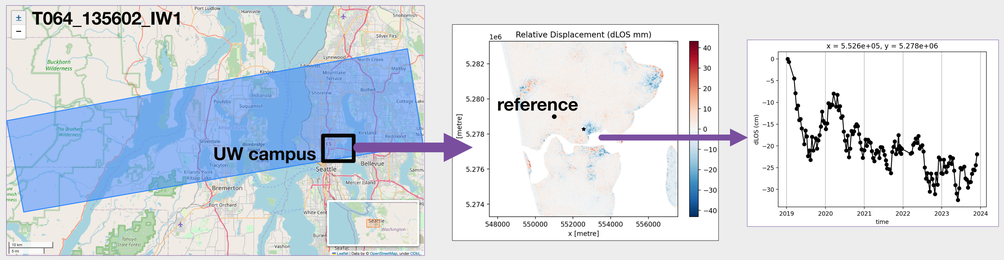
Next: on-demand products

If standard products do not meet your scientific needs, there are options for generating your own products on scalable cloud-computing resources

hyp3 Example: Hyp3-ISCE2 displacement time series

- ASF now offers global Sentinel-1 burst SLCs from June 2023 onwards
- Burst interferograms are created with custom open source ISCE2 workflow
- Processing uses 30m Copernicus DEM with 20m final product pixel posting
- ASF Hyp3 generously offers quota 1000 jobs per month per person
- Hyp3 offers parallel scaling (20 min for 1 burst or 20 minutes for 100 bursts!)
- Outputs available publicly for 2 weeks

Minimal network of 142 interferograms, 138 dates to investigate urban areas of artificial fill:

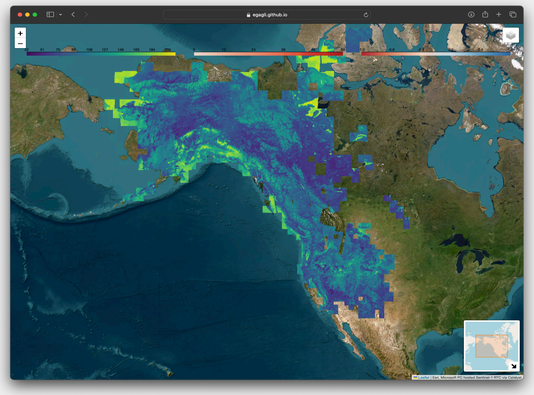


STAC **Tip: A STAC catalog on GitHub for public hyp3 outputs allows for easy collaboration!**

https://radiantearth.github.io/stac-browser/#/external/raw.githubusercontent.com/relativeorbit/agu2023/main/catalog.json

Example: S1 RTC time series mapping

Scientific goal: quantify changes in changes in annual snowmelt runoff onset at regional and global scales at 80m resolution from 2014 to present



- Sentinel-1 RTC Global Dataset (~100TB) hosted by MS Planetary Computer
- Microsoft Planetary Computer JupyterHub (Azure)
- Dask-gateway scaling (50-400 vCPU cluster)
- Output cloud-optimized geotiff (COG) to Azure Blob storage (az://) for public

***See also:** Gagliano et. al. 2023 GRL &
C41B-05: Eight years of global snowmelt runoff onset from Sentinel-1 SAR

https://github.com/egagli/sar_snowmelt_timing

Cloud-computing infrastructure

jupyter

Pre-configured JupyterHubs are a convenient way experience some of the benefits of cloud computing. They are often deployed in the same data center as large archives of data to enable efficient data-proximate computing

- Deployed in range of major cloud data centers
- Curated software environments
- Scalable compute resource selection
- Community forums for technical support

Service	Max vCPU	Max RAM (GB)	Storage (GB)	Datacenter
CryoCloud	4	32	10	AWS us-west-2
Pangeo JupyterHub	16	32	10	GCP us-central-1b
ASF Open Science Lab	8	16	500	AWS us-west-2

- Dedicated object storage (S3://)
- Dask distributed computing clusters
- GPUs

Service	vCPU	RAM (GB)	GPU	GPU RAM (GB)	Storage (GB)	Max Session (hr)	Datacenter
Google Colab	2	12	T4	16	40	12	random!
AWS Sagemaker Studio Lab	4	12	T4	16	15	4	us-east-2
Microsoft Planetary Computer	4	32	T4	16	150	12	eu-west-2

*If existing platforms don't suffice, see this summary for fee-based options and recommendations for Cloud account credits for academic research:

https://geosmart.hackweek.io/tutorials/computing

General recommendations

* Abstract away files and structure workflows so that data are always referenced as URLs (http://, s3://)

* URLs work best pointing at a cloud native formats like COG, ZARR, GeoParquet (read more https://guide.cloudnativegeo.org)

* Move algorithms to the data for efficiency and reproducibility

* Engage in public discourse early and regularly to shape the future of data and software design (e.g. https://discourse.pangeo.io)

Acknowledgements

Financial support from University of Washington eScience Institute Winter Incubator Program and NASA Support for Open-Source Tools, Frameworks, and Libraries Grant. Workflows highlighted here depend on open data products from NASA & ESA and open source software and cloud infrastructure made available via ASF DAAC and the OPERA project.

poster link