G41B-0683

AGU100 ADVANCING EARTH AND SPACE SCIENCE
FALL MEETING
Washington, D.C. | 10–14 Dec 2018

Scalable InSAR processing and analysis on the Cloud

with applications to geohazard monitoring in the Pacific Northwest

Scott Henderson (scottyh@uw.edu), David Schmidt, Department of Earth and Space Sciences, University of Washington

UNIVERSITY of WASHINGTON eScience Institute

Motivation

1) NASA is planning to move Petabyte-scale archives of satellite imagery to commercial Cloud providers (Amazon AWS)

2) Researchers are inhibited by the need to download, analyze, and share large data sets

3) ESA's Sentinel-1 SAR archive is growing at 1Pb/yr, has many research use cases, and serves as a proxy for the upcoming NASA NISAR mission

4) Geohazard monitoring benefits from fast processing and dissemination of products that allow for rapid quantitative investigation

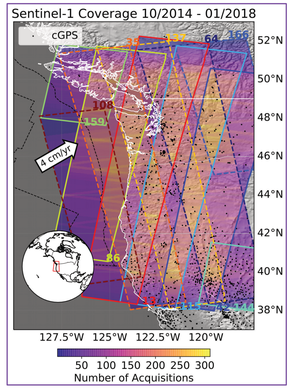

Figure 1: Colored polygons depict Sentinel 1 orbital track coverage of Cascadia region. Arrow on the the Juan de Fuca plate (dashed black line) represents subduction at approximately 4cm/yr with respect to stable North America. Black dots are continuous GPS stations.

Scalable InSAR Processing

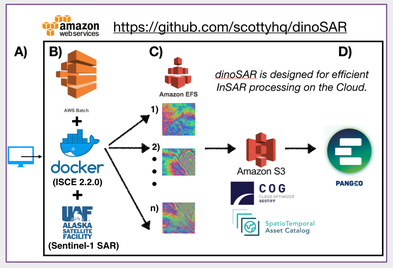

Figure 2: dinoSAR workflow from left-to-right. **A)** Sentinel-1 archive is queried via ASF and a list of interferometric pairs is selected for processing. **B)** 'n' processing directories are mapped onto network storage and 'n' instances of ISCE software beginning processing via Docker and AWS Batch. **C)** ISCE outputs are converted to Cloud-Optimized Geotiffs (COGs) with accompanying STAC metadata. **D)** Quantitative post-processing analysis (e.g. stacks, profiles, time series) can be in parallel with Pangeo.

See Also:

ISCE - https://winsar.unavco.org/software/isce

ARIA - https://disasters.nasa.gov/programs/aria

COMET LiCS - https://comet.nerc.ac.uk/COMET-LiCS-portal

SARVIEWS - http://sarviews-hazards.alaska.edu

***dinoSAR* Key Features:**

- Avoid downloading data and egress cost by running in same region as data (us-east-1)

- Customize compute resources (managed or unmanaged, large or small ephemeral clusters)

- Customize processing parameters (anything exposed via ISCE software, e.g. DEM, filter strength).

- Lower costs with spot-market instances

- GPU enabled via *nvidia-docker*

- Easy installation as a Python package (pip install dinosar)

- Versioned documentation

- Cloud-optimized outputs for post-processing with Pangeo (see next panel)

Efficient Post-processing

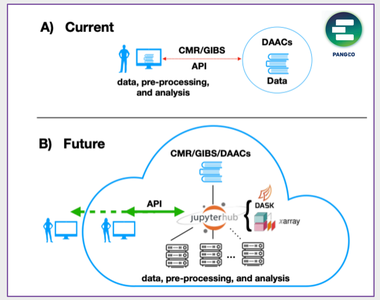

Figure 3: Schematic of interactive, scalable Cloud processing with Pangeo. **A)** The current paradigm where scientific computations are preformed locally on duplicated datasets. **B)** Representation of the Pangeo toolset which enables scientists to work entirely in the cloud via an API and scalable interactive computing resources. The dashed line indicates the ability to access the API and server-side processing via the Internet, for example to request large computations and then download only results to a local workstation.

***Pangeo* Key Features:**

- Foster collaboration around the open source scientific python ecosystem for ocean / atmosphere / land / climate science

- Improve scalability of these tools to handle petabyte-scale datasets on HPC and Cloud platforms

- Contribute to general scientific python packages such as ***rasterio, xarray, dask*** to accelerate discovery in geoscience.

- Use Jupyter as a user interface to analyze data wherever it is stored, completely avoiding the need to download.

- Enable interactive data exploration with the same interface whether we have 10Mb of imagery on a laptop or 10Tb of imagery on the Cloud.

Read more!

http://pangeo.io

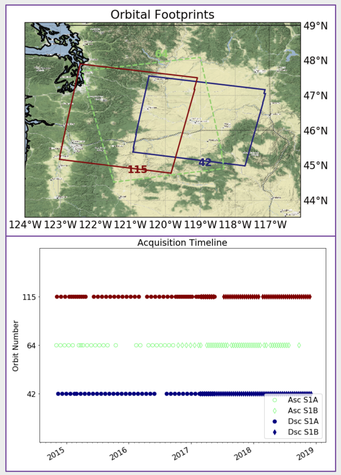

Figure 4: Typical Sentinel-1 coverage for several frames in Washington State. For dense InSAR time series studies, thousands of pairwise combinations can be generated. Raw data for orbit number 115 is 385 Gb, all 12-day, 24-day, and 36-day pairs can be processed in parallel in several hours with dinoSAR.

dinoSAR/Pangeo example 1:

Analyze 12-day, 24-day, and 36-day pairs processed at 15-m resolution from relative orbit 115 covering Rattlesnake Ridge Landslide:

GRFN/Pangeo example 2:

Visualize and analyze interferograms from the Getting Ready for NiSAR project covering volcanic activity in Hawaii:

Rattlesnake Ridge Landslide

Key Points:

- Fissure noticed in October 2017 spurred installation of ground monitoring network

- Sentinel-1 Archive goes back to October, 2014 allowing for investigation of onset and continued monitoring

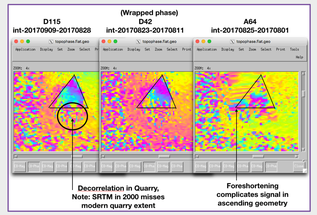

Figure 6: Small footprint of slide requires processing with high resolution DEM, even with upsampled 15-m STRM radar foreshortening in ascending data is limiting.

Figure 5: Areal view of Rattlesnake Ridge landslide near Union Gap, Washington State. Key features labeled on image.

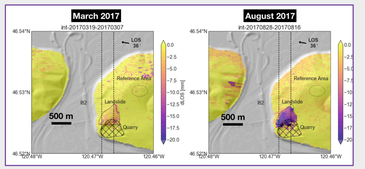

Figure 7: 2mm/hr southward motion of slide detectable in 12-day pairs as far back as March 2017.

Conclusions

1) *dinoSAR* can process large collections of SAR data in parallel on AWS. The workflow is dockerized and scripted to be portable and reproducible. Outputs stored as Cloud-Optimized Geotiffs with STAC metadata enable easy dissemination and analysis.

2) The Pangeo platform offers a solution to eliminate downloading data and perform scalable analysis wherever large datasets (such as SAR) are stored.

3) Options exist to balance computational efficiency and cost. Current per-interferogram generation costs are $0.1 to $1.

4) The Rattlesnake Ridge Landslide monitored by regular Sentinel-1 observations suggests onset of motion before spring 2017.

Acknowledgements

Funding for this project comes from the eScience Institute and Washington Research Foundation, a NASA NNH17ZDA001N-ACCESS grant and an Earth on AWS grant. Sentinel-1 data is generously provided by ESA and distributed by NASA's ASF DAAC.