

Shane P. Grigsby, Ph.D.

Emeryville, CA • 720.837.0809 • refuge@rocktalus.com • espg.github.io • github.com/espg

SUMMARY

Geospatial research scientist and data-infrastructure engineer with over a decade and a half building distributed pipelines that turn petabyte-scale geospatial data archives into science-grade outputs. PhD in cryospheric remote sensing with doctoral training in geostatistics, machine learning, and laser remote sensing; masters in atmospheric correction and hyperspectral/thermal remote sensing. Persistent open-source software contributor. Passionate about spatial indexing, serverless cloud compute, and structuring massive geospatial data for efficient client-side access.

EXPERIENCE

Site Reliability Engineer / Open Science Infrastructure Engineer *Astera Institute* July 2025 – May 2026

- Cofounded [englacial](#); primary author of multiple libraries in the [englacial github](#) organization including xopr, icechunk-js, zagg, and others.
- Designed and led the cloud-native data-infrastructure stack for the englacial open-science cryosphere program: STAC catalogs, virtualized Zarr stores, and client-side data delivery on S3.
- Own GitHub Actions CI/CD across the program's repos: unit + integration tests, code-coverage gating, multi-platform wheel/sdist releases, reproducible pinned environments, and deploy gates for data-publication pipelines.
- Evaluated and selected proposals for five active research awards; managed engineering-support contracts for grant awardees.

Senior Research Engineer *Byrd Polar & Climate Research Center, Ohio State University* Dec 2023 – Dec 2024

- Built a custom Xarray data-accessor for PyTorch that handled tile-level access and dataset shuffling against chunked, cloud-resident Sentinel-2 raster archives.
- Stood up the unit-test, code-coverage, and pip-release CI/CD for the ParallelGAMIT ([pgamit](#) / [Geode](#)) distributed GNSS-processing library.
- Developed new bespoke clustering algorithm for GNSS data processing (Grigsby & Gómez 2026, GPS Solutions).
- Lead multi-institute science proposals to federal agencies.

R&D Scientist / Program Manager *National Geospatial-Intelligence Agency (NGA)* Nov 2021 – Jan 2023

- Technical lead and SME for ML and remote-sensing systems at petabyte scale across the NGA Research Directorate; led architecture review and source-selection for cloud-native analytic capability development.
- Program Manager for the AMOB program: cross-team portfolio spanning cloud infrastructure, internal integration, and external FFRDC evaluation. Authored agency-wide open-source software policy (adopted).
- Authored test-and-evaluation strategy for ML systems on classified geospatial data.
- Intelligence officer, Payband 4 civil servant in the excepted service (GS 13/14 equivalent); Tier 5 investigation (TS/SCI).

Visiting Assistant Scientist, ICESat-2 Project Office *NASA Goddard / UMD ESSIC* Jan 2021 – Nov 2021

- Built cloud-native distributed pipelines (Dask on object storage) converting ICESat-2 photon archives into Parquet/Hive partitioned tables for fast range-scan access — [early Lakehouse-pattern work](#) in the geosciences.
- Designed HEALPix-based spatial-indexing strategy enabling fast spatial joins between altimetry point data and global raster archives; coordinated with NSF "Jupyter meets the Earth" (PI: Fernando Pérez, UC Berkeley).

Postdoctoral Researcher *Colorado School of Mines / CU Boulder CIRES* Aug 2019 – Jan 2021

- Ported stochastic-simulation (covariance-method) and image-processing pipelines to GPU via CUDA / CuPy; upstreamed CuPy speed enhancements (PR #3018) shipped in v8.0.0b+.
- Co-Investigator on three funded NASA grants (total ~\$2M) spanning ICESat-2 validation, polar-snow assimilation, and subglacial-lake science.
- External reviewer, ICESat-2 ATL11 Algorithm Theoretical Basis Document.

Earlier roles 2009 – 2019

- Team Lead & Data Architect, Orbital Micro Systems (2018–2019): led 8–10-person team designing operational satellite-weather data-ingest systems.
- Research Analyst, Intel "BigData" Science & Technology Center (2013–2014): schema design for petabyte-scale remote-sensing array databases on SciDB; spatio-temporal indexing prototypes for fast tile/subset retrieval against MODIS archives.
- Linux SysAdmin, CU Boulder Research Computing (2009–2011): built and maintained production HPC clusters and custom software builds.

SELECTED OPEN-SOURCE & PLATFORM WORK

Maintainer / primary author

- **Cal-Adapt Analytics Engine JupyterHub** (github.com/Eagle-Rock-Analytics/jupyterhub-tofu) — Designed the production JupyterHub for California’s Cal-Adapt climate-services platform (130+ users). Full Infrastructure-as-Code (OpenTofu / Terraform, EKS, GitOps) for multi-user, GPU-capable compute against cloud-optimized climate archives.
- **xopr** (github.com/englacial/xopr) — Client-side Xarray-native data loader and STAC-catalog tooling for the Open Polar Radar archive. Catalog runs as DuckDB against cloud-hosted GeoParquet on S3: zero-server design where the client pulls a few MB of S3-hosted catalog, then does full spatial/temporal search in memory, client-side.
- **icechunk-js** (github.com/EarthyScience/icechunk-js) — TypeScript client-side reader for Icechunk and other virtualized Zarr v3 stores; browser-native, server-less access to chunked raster archives on S3 via range requests.
- **zagg** (github.com/englacial/zagg) — Cloud-native spatial-aggregation engine. Performs tiled / windowed reductions over massive Earth-data archives (NASA CMR, STAC-indexed data, etc.) & writes results incrementally to Zarr arrays.
- **mortie** (github.com/espg/mortie) — Morton-order indexing for HEALPix discrete global grids. Locality-preserving tiling scheme for spatial joins, sharding, and range-scan access patterns on global data.
- **ISMIP6 indexing + viewer** (ismip-indexing, ismip-viewer) — Ingested and virtualized ~2 TB of ISMIP6 ice-sheet model output (virtualized dataset at source.coop/englacial/ismip6); the companion [browser viewer](https://browser-viewer) renders model fields client-side via icechunk-js — no app server, no pre-rendered tiles.

Contributor

- **scikit-learn OPTICS** — Primary author of the OPTICS density-based clustering module ([PR #1984](https://doi.org/10.26434/chemrxiv-2019-08-01), shipped in v0.21+). Multi-year contribution covering algorithm, matching of sklearn API, unit tests, benchmarks, and numpdoc.
- **CuPy** — Contributor to CuPy (NumPy-compatible CUDA array library); multivariate-normal sampling speed enhancements shipped in v8.0.0b+ ([PR #3018](https://doi.org/10.26434/chemrxiv-2019-08-01)).

TECH STACK

Languages: Python (expert), TypeScript, C/CUDA, Rust, myst-md, nix (working), SQL, Bash, Linux.

Distributed compute & ML: Dask, AWS Lambda (Dask → serverless scale-out), PyTorch, scikit-learn, CuPy, NumPy, Xarray, PyArrow, Pandas, pymc, tensorflow, vaex. GPU / CUDA-aware pipelines.

Data formats & storage: Zarr v3, icechunk, Cloud-Optimized GeoTIFF (COG), Parquet / GeoParquet, STAC, NetCDF/HDF5, S3 (range requests, sharding, Apache Hive partitioning), geojson.

Geospatial: waveform, point-cloud, and photon lidar; raster + vector pipelines, projections (PROJ), tiling schemes.

Cloud & platform: AWS (EKS, EC2, S3, IAM), Kubernetes, Terraform / OpenTofu, Docker, JupyterHub, GitHub Actions CI/CD (unit + integration tests, coverage gating, multi-platform wheel/sdist builds, reproducible-env pinning), observability and structured logging (grafana, prometheus, etc).

EDUCATION

Ph.D., Geography *University of Colorado Boulder / CIRES* 2019

Dissertation: Greenland Surface Roughness Retrieval and Status Adviser: Dr. Waleed Abdalati

M.A., Geography *University of California, Santa Barbara* 2014

Thesis: Improved Surface Temperature Estimates with MASTER/AVIRIS Sensor Fusion Adviser: Dr. Dar Roberts

B.A., Geography & Philosophy, magna cum laude *University of Colorado Boulder* 2011

SELECTED PUBLICATIONS & RECOGNITION

h-index 7 • 600+ citations (400+ since 2021). Full list: <https://espg.github.io/publications>

- Grigsby & Gómez. Efficient clustering of GNSS stations for processing using double differences. *GPS Solutions* 30(1):60, 2026.
- Snow, Straneo, Holte, Grigsby, et al. More than skin deep: SST as a means of inferring Atlantic Water variability near Helheim Glacier. *J. Geophys. Res. Oceans*, 2021.
- Christoffersen, Bougamont, Hubbard, Doyle, Grigsby & Pettersson. Cascading lake drainage on the Greenland Ice Sheet triggered by tensile shock and fracture. *Nature Communications*, 9(1), 1–12, 2018.
- Grigsby, Hulley, Roberts, et al. Improved surface temperature estimates with MASTER/AVIRIS sensor fusion. *Remote Sensing of Environment* 167:53–63, 2015.
- NASA Group Achievement Award (2019), NASA Student Airborne Research Program (SARP) — "For sustained exceptional achievement in training and recruiting the next generation of Earth system scientists and engineers."
- NGA CORE Award (2022) — "For sustained excellence and leadership of the AMOB program."