The Journal of Open Source Software

SlideRule: Enabling rapid, scalable, open science for the NASA ICESat-2 mission and beyond

David Shean[1], **J. P. Swinski**[*2], **Ben Smith**[3], **Tyler Sutterley**[3], **Scott Henderson**[4,5], **Carlos Ugarte**[2], **Eric Lidwa**[2], and **Thomas Neumann**[2]

1 University of Washington, Department of Civil and Environmental Engineering **2** NASA Goddard Space Flight Center **3** University of Washington, Applied Physics Laboratory **4** University of Washington, Department of Earth and Space Sciences **5** University of Washington, eScience Institute

DOI: 10.21105/joss.04982

Software

- Review
- Repository
- Archive

Editor: Kristen Thyng

Reviewers:

- @jhkennedy
- @betolink

Submitted: 22 August 2022
Published: 18 January 2023

Summary

`SlideRule` is an open source server-side framework for on-demand processing of science data in the cloud. The `SlideRule` project offers a new paradigm for NASA archival data management – rapid delivery of customizable on-demand data products, rather than hosting large volumes of standard derivative products, which will inevitably be insufficient for some science applications.

The scalable server-side components of `SlideRule` run in the AWS cloud with optimized functions to read HDF5 data hosted by NASA in S3 cloud object storage. While `SlideRule` can be accessed by any HTTP client (e.g., curl) through GET and POST requests, the `sliderule-python` client provides a user-friendly API for synchronous interaction with the `SlideRule` service. The client library returns standard Python data containers (i.e., Pandas DataFrame) and facilitates serialization with provenance metadata for reproducible science.

`SlideRule` uses a plugin framework to support different NASA missions and data products. The ICESat-2 `SlideRule` plugin offers customizable algorithms to process the archive of low-level data products from the NASA Ice Cloud and land Elevation Satellite-2 (ICESat-2) laser altimetry mission. The user defines a geographic area of interest and key processing parameters via an interactive web interface or the API, and SlideRule returns high-level surface elevation point cloud products in seconds to minutes, enabling rapid algorithm development, visualization and scientific interpretation.

Statement of need

ICESat-2 launched in September 2018 as the follow-on mission for the original ICESat mission (2003-2010), which pioneered the application of space-based laser altimetry to precisely measure the height of the Earth's changing surface (Markus et al., 2017). These missions are optimized to measure elevation change of the cryosphere, including ice sheets and sea ice, allowing scientists to quantify and understand changes in a warming climate.

ICESat-2 carries the Advanced Topographic Laser Altimeter System (ATLAS), which records the two-way travel time of transmitted photons from six beams (Neumann et al., 2019). The Level 2A Geolocated Photon Data Product (ATL03) provides unique latitude, longitude, height and timing of each photon, along with many other attributes, including basic classification as a return from the Earth's surface vs. background photons from sunlight or instrument noise.

The ATL03 data granules are stored as ~1-2 GB HDF5-format files containing ~20-100 million records. This data volume and complexity impedes new users and slows the speed of algorithm

*co-first author

development. The ICESat-2 project team generates many higher-level products that reduce the ATL03 data using established processing algorithms. For example, the ATL06 product (Smith et al., 2019) provides high-precision estimates of surface height at 40-meter resolution, using parameters appropriate for flat, highly reflective polar snow surfaces. Similar to ATL03, the ATL06 products contain a large number of parameters describing each height measurement that may not be relevant to all users. Furthermore, standard ATL06 products are only produced over glaciers and ice sheets, and the default algorithm parameters (such as 20-meter posting) may not be optimal for complex land surfaces or vegetation.

`SlideRule` offers a solution to these issues, allowing users to create on-demand products with the vetted ATL06 algorithm, but with adjustable parameters and photon classification strategies tailored to the characteristics of their application and study sites.

State of the field

The current paradigm for ICESat-2 data access involves downloading large volumes of standard data products from a NASA Distributed Active Archive Center (DAAC), then writing custom routines to prepare those products for analysis. The National Snow and Ice Data Center (NSIDC) offers data discovery and limited subsetting services, allowing users to request and download products for a user-specified geographic area with a user-defined subset of returned variables. Even with these subsetting services, the full workflow to request, stage, and download hundreds of products can take several minutes to hours, especially for larger areas, and these services do not currently support custom server-side data processing.

On-demand science data processing

Several projects are exploring on-demand, cloud-based processing for satellite and/or point cloud data. For example, the Alaska Satellite Facility's Hybrid Pluggable Processing Pipeline (ASF HyP3) enables custom processing of satellite SAR images from multiple missions (Hogenson et al., 2020). The OpenTopography project offers "Web service-based data access, processing, and analysis capabilities that are scalable, extensible, and innovative" with emphasis on "high-resolution (meter to sub-meter scale), Earth science-oriented, topography data acquired with LiDAR and other technologies." The current processing options and data products are focused on airborne LiDAR point clouds, with no plans to support the more complex ICESat-2 data products.

Currently available ICESat-2 processing packages

Several existing projects offer software and/or APIs to process, analyze and visualize ICESat-2 data products. We briefly summarize these efforts to provide context and justification for the ICESat-2 SlideRule project.

`icepyx` is a python library supporting programmatic access to ICESat-2 data through the NASA Common Metadata Repository (CMR) and NSIDC services (Scheick & others, 2019). `icepyx` allows for queries based on spatial and temporal parameters, as well as ICESat-2 orbital cycle and Reference Ground Track (RGT). Accessing NSIDC API services through `icepyx` allows users to subset ICESat-2 data and convert data to other file formats. At present, `icepyx` facilitates reading and visualizing data available from NSIDC, but it is not a data processing service.

`OpenAltimetry` offers discovery, access, and visualization of data from NASA's ICESat and ICESat-2 missions. This service includes an API that can provide access to either photon-level ATL03 data or derived height variables (e.g., ATL06) for a single reference ground track, but does not offer processing from photons to higher-level products (Khalsa et al., 2020).

`phoREAL` (Photon Research and Engineering Analysis Library) from the University of Texas is an open source application allowing for subsetting and conversion of the ATL03 and derived

The Journal of Open Source Software

ATL08 Land and Vegetation Height data products (Neuenschwander & Pitts, 2019). `phoREAL` allows users to spatially subset locally-stored data based on reference ground track. `phoREAL` allows users to visualize data from both ATL03 and ATL08 and extract statistics for heights based on time or geospatial location.

`captoolkit` (Cryosphere Altimetry Processing Toolkit) from the NASA Jet Propulsion Library (Paolo et al., 2020) allows users to estimate elevation change using altimetry data from multiple airborne and satellite missions. `captoolkit` has functions to apply geophysical corrections, calculate elevation change, and interpolate points into gridded fields. `captoolkit` uses parallelized functions that operate on local granule files, with optimization for High Performance Computing (HPC) clusters.

SlideRule project

As of version 1.5.8, the SlideRule project includes the following repositories:

- `sliderule` server framework with core functionality, including plugins for different missions and the `H5Coro` driver
- `sliderule-python` client, with language-specific API, example Jupyter notebooks, and source for the interactive web interface.
- `sliderule-docs` project documentation and website

SlideRule server framework

`SlideRule` is a C++/Lua framework for on-demand data processing (Figure 1). It is a science data processing service that runs in the cloud and responds to REST API calls to process and return science results. The SlideRule service was designed for synchronous processing - the client connection remains open after a request is submitted, and results are streamed back to the engaged user in near-real-time. This model is preferable over asynchronous processing, where requests are queued and users are notified to retrieve results at a later time.

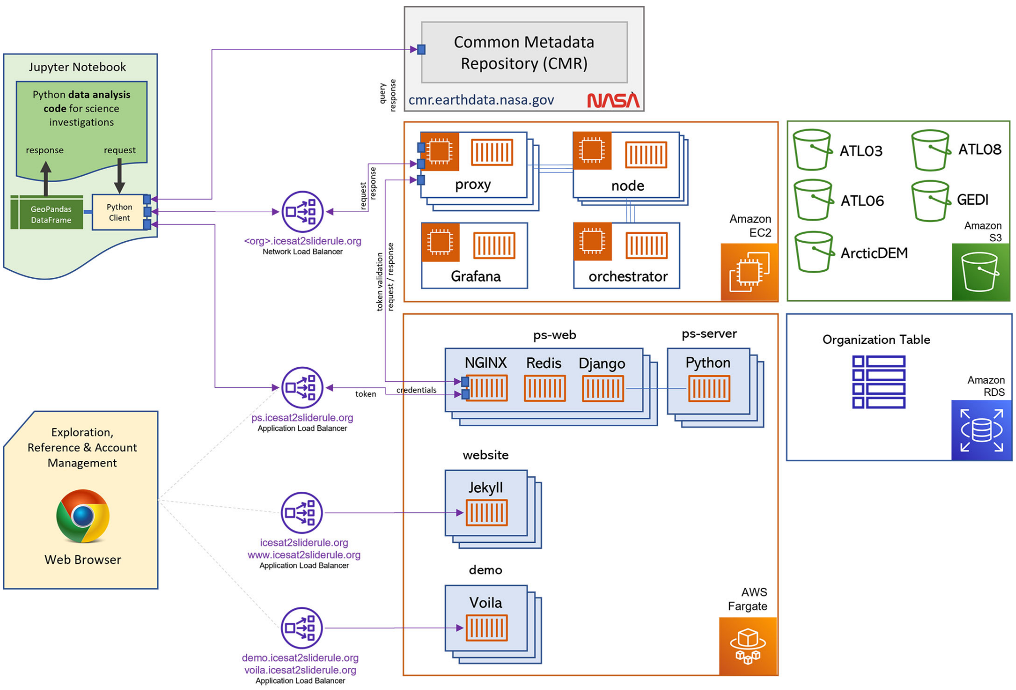

Figure 1: SlideRule architecture schematic.

ICESat-2 SlideRule plugin

The ICESat-2 SlideRule plugin provides two main types of services: 1) efficient access to archived ICESat-2 products, and 2) custom derived products (ATL06-SR) generated from the low-level ATL03 cloud assets.

Standard data product access

`SlideRule` provides services for parallel access to original mission data products without the need to download or restructure in cloud-friendly formats (e.g., Zarr (Miles et al., 2022)). This is accomplished using the HDF5 Cloud-Optimized Read-Only (H5Coro) library for efficient reading of HDF5 files. The returned points preserve a subset of the many original ATL03 parameters, including those needed to interpret surface heights from the photon-height distribution. Additional functions provide similar access to the standard ATL06 cloud assets.

Custom data products: ATL06-SR algorithm

The standard ATL06 (land-ice height) product (Smith et al., 2019) (hereafter "ATL06-legacy") provides estimates of surface height and slope with along-track spacing of 20 m, derived from 40-m segments of ATL03 photons classified as likely surface returns. The algorithm iteratively improves the initial photon classification, and calculates a set of corrections based on the residuals between the segment heights and the selected photon heights to correct for sub-centimeter biases in the resulting height estimates.

The ATL06-SlideRule ("ATL06-SR") algorithm implements much of the ATL06-legacy algorithm in a framework that allows for customization of the initial photon classification, the segment length and spacing, and the filtering strategy used to identify valid segments in the product. For the sake of simplicity and speed, ATL06-SR omits some of the corrections required for sub-centimeter accuracy over polar surfaces, including the first-photon bias and pulse-shape corrections (Smith et al., 2019). Omitting these corrections should result in biases that are no more than a few centimeters, which, over complex or rapidly changing non-polar surfaces, are unlikely to be the dominant source of errors.

One important difference between the ATL06-legacy and ATL06-SR algorithms is the iterative photon-selection strategy. Both algorithms begin with photons selected based on the input photon classification. The selection is then refined to reject photons that have large residual values (i.e., larger than three times a robust estimate of the standard deviation of the previously selected residuals) and the fit is recalculated. This process continues until a specified number of iterations has occurred, or until subsequent iterations leave the photon selection unchanged. In ATL06-legacy, at each iteration the photons are selected from the complete set of input photons, regardless of how they were initially flagged or selected, while in ATL06-SR, each iteration selects photons exclusively from those selected in the previous iteration. This means that in ATL06-legacy, the number of photons can increase from one iteration to the next, while in ATL06-SR it can only remain constant or decrease. This choice lets ATL06-SR more faithfully reflect the height distribution of the initial selection, at the expense of sometimes missing photons close to the surface that might add more information to the fitting algorithm, possibly improving the accuracy of the fit. Under most circumstances, the two algorithms produce very similar results.

Photon classification

The ATL06-SR algorithm supports three sources of photon classification data (Figure 2): 1) ATL03 photon confidence values (from respective classification algorithms for land, ocean, sea-ice, land-ice, or inland water) (Neumann et al., 2019), 2) ATL08 photon classification (Neuenschwander & Pitts, 2019), and 3) YAPC (Yet Another Photon Classification) photon-density-based classification (Sutterley, 2022).

The Journal of Open Source Software

The ATL08 Land and Vegetation Height product (Neuenschwander & Pitts, 2019) includes improved photon-level classification for vegetation, including returns from the canopy, ground surface, and noise. The ICESat-2 plugin can efficiently retrieve the photon classifications from ATL08 cloud assets and apply to the desired subset of ATL03 photons.

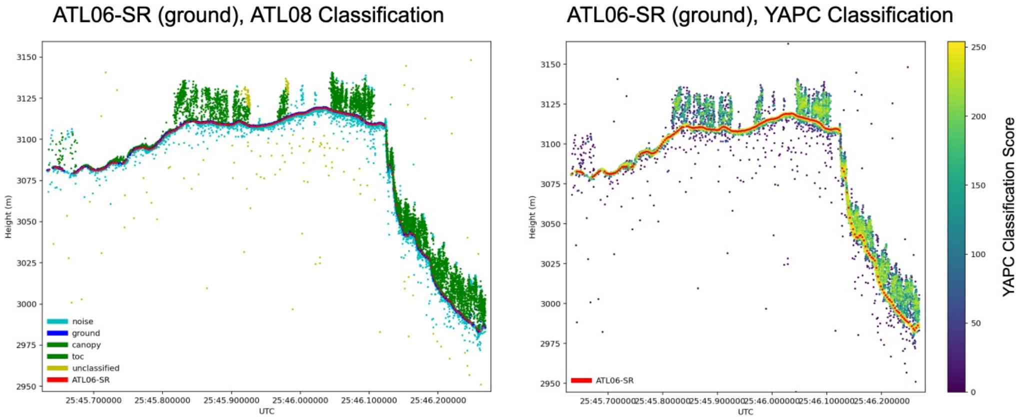

Figure 2: Example SlideRule output of classified ATL03 photons and corresponding ATL06-SR points.

Users can specify the classification values to be included during initial photon selection for the ATL06-SR algorithm. For example, a user might wish to generate ATL06-SR products using only photons with ATL03 classification values indicating medium or better confidence, or photons flagged as ground returns in ATL08 (removing any photons from canopy returns). This powerful approach enables precise surface elevation measurement with the ATL06-SR algorithm over all land surfaces, not just polar ice sheets and glaciers.

Future Work

The rapid processing offered by the `SlideRule` service allows for interactive testing and evaluation of various combinations of ATL06-SR parameter and classification schemes for a broad range of scientific and engineering applications. In the coming years, we plan to extend SlideRule to support altimetry data from other NASA missions (e.g., GEDI (Dubayah et al., 2020)), raster data, and efficient server-side fusion of the two data types.

Acknowledgements

The `SlideRule` project is funded by the NASA ICESat-2 project and NASA award 80NSSC20K0995. We thank reviewers Luis López (@betolink) and Joe Kennedy (@jhkennedy) for their detailed and insightful comments, which improved the quality of the SlideRule code, documentation and this paper.

References

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P. L., Qi, W., & Silva, C. (2020). The global ecosystem dynamics investigation: High-resolution laser ranging of the earth's forests and topography. *Science of Remote Sensing*, *1*, 100002. https://doi.org/10.1016/j.srs.2020.100002

Hogenson, K., Kristenson, H., Kennedy, J., Johnston, A., Rine, J., Logan, T., Zhu, J., Williams, F., Herrmann, J., Smale, J., & Meyer, F. (2020). *Hybrid Pluggable Processing Pipeline*

(HyP3): A cloud-native infrastructure for generic processing of SAR data (Version v0.3.43) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.6917373

Khalsa, S. J. S., Borsa, A., Nandigam, V., Phan, M., Lin, K., Crosby, C., Fricker, H., Baru, C., & Lopez, L. (2020). OpenAltimetry-rapid analysis and visualization of spaceborne altimeter data. *Earth Science Informatics*, 1–10. https://doi.org/10.1007/s12145-020-00520-2

Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., Farrell, S., Fricker, H., Gardner, A., Harding, D., Jasinski, M., Kwok, R., Magruder, L., Lubin, D., Luthcke, S., Morison, J., Nelson, R., Neuenschwander, A., Palm, S., … Zwally, J. (2017). The ice, cloud, and land elevation satellite-2 (ICESat-2): Science requirements, concept, and implementation. *Remote Sensing of Environment*, *190*, 260–273. https://doi.org/10.1016/j.rse.2016.12.029

Miles, A., jakirkham, Bussonnier, M., Moore, J., Fulton, A., Bourbeau, J., Onalan, T., Hamman, J., Orfanos, D. P., Patel, Z., Lee, G., Rocklin, M., Abernathey, R., Bennett, D., Andrade, E. S. de, Durant, M., Schut, V., dussin, raphael, Chaudhary, S., … Banihirwe, A. (2022). *Zarr-developers/zarr-python: none* (Version v2.12.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.6697361

Neuenschwander, A., & Pitts, K. (2019). The ATL08 land and vegetation product for the ICESat-2 mission. *Remote Sensing of Environment*, *221*, 247–259. https://doi.org/10.1016/j.rse.2018.11.005

Neumann, T. A., Martino, A. J., Markus, T., Bae, S., Bock, M. R., Brenner, A. C., Brunt, K. M., Cavanaugh, J., Fernandes, S. T., Hancock, D. W., Harbeck, K., Lee, J., Kurtz, N. T., Luers, P. J., Luthcke, S. B., Magruder, L., Pennington, T. A., Ramos-Izquierdo, L., Rebold, T., … Thomas, T. C. (2019). The ice, cloud, and land elevation satellite–2 mission: A global geolocated photon product derived from the advanced topographic laser altimeter system. *Remote Sensing of Environment*, *233*, 111325. https://doi.org/10.1016/j.rse.2019.111325

Paolo, F., Nilsson, J., & Gardner, A. (2020). *Fspaolo/captoolkit: First release* (Version 0.1.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.3665785

Scheick, J., & others. (2019). *icepyx: Python tools for obtaining and working with ICESat-2 data*. icesat2py. https://github.com/icesat2py/icepyx

Smith, B., Fricker, H. A., Holschuh, N., Gardner, A. S., Adusumilli, S., Brunt, K. M., Csatho, B., Harbeck, K., Huth, A., Neumann, T., & others. (2019). Land ice height-retrieval algorithm for NASA's ICESat-2 photon-counting laser altimeter. *Remote Sensing of Environment*, *233*, 111352. https://doi.org/10.1016/j.rse.2019.111352

Sutterley, T. (2022). *Tsutterley/yapc: v0.0.0.8* (Version 0.0.0.8) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.6717591