STV Precursor Coincident Datasets

Version 1.0
September 6, 2024

https://science.nasa.gov/earth-science/decadal-surveys/decadal-stv/coincident-datasets/

Prepared by:
- David Shean, University of Washington, dshean@uw.edu
- Jack Hayes, University of Washington
- Scott Henderson, University of Washington

Table of Contents

Abstract

Over the past two decades, large archives of 3D surface elevation measurements have been acquired from a range of airborne and satellite platforms. Many of these datasets are now publicly available from NASA Earthdata Cloud and commercial cloud storage, enabling efficient, on-demand search, subsetting, and data access.

Many "coincident" or near-contemporaneous measurements in these archives were fortuitously acquired over the same location within a short temporal window (e.g., <1-14 days). These archives remain largely untapped for a broader, systematic identification and analysis of spatiotemporally coincident 3D measurements.

We are developing efficient, open-source tools to identify, curate, process and distribute coincident 3D datasets contained within these archives across a range of representative terrain and landcover types. These datasets can be used by the larger STV community for calibration/validation, fusion algorithm development, and discipline-specific scientific analysis.

Change History Log

Version	Description of Change	Date
1.0	Initial release	2024-09-06

Introduction

The core Surface Topography and Vegetation (STV) science priorities and applications involve a range of features and processes that span broad spatial (sub-meter to 100s of km) and temporal (seconds to millions of years) scales, over a wide array of landcover classes and terrain characteristics [*Donnellan et al.*, 2021]. Candidate STV architectures involve multiple instrument modalities (lidar, stereo, radar) and multiple platforms (satellite, airborne, HAPS), which inherently involves variable time offsets between different measurements for any given location.

Primary STV requirements and objectives involve the creation of global reference basemap and change products with high resolution, accuracy, and precision. Meeting these objectives will require the development of data fusion algorithms to combine multi-modal measurements from multiple platforms over variable time periods (several months to years). This approach is much more challenging than past global mapping efforts like the Shuttle Radar Topography Mission (SRTM, [*Farr et al.*, 2007]), which involved single-pass InSAR from two instruments with fixed baseline over a relatively short, 11-day period in February 2001, effectively offering a "snapshot" view of the Earth's surface topography.

Motivation

The following factors can introduce differences in measurements of surface elevation and vegetation structure:

1. Different instrument configuration (e.g., wavelength), acquisition geometry, platform, etc. for a single measurement approach
2. Different measurement approaches (e.g., lidar and radar), each with unique physical principles and sensitivity to surface and vegetation properties
3. Real change over time in surface elevation (e.g., ice melt), surface properties (e.g., soil moisture) and vegetation structure (e.g., leaf-out, growth, disturbance)

In order to understand and develop solutions to account for the first two variables, we must eliminate (or reduce) the third. This can be accomplished by **ensuring that all measurements are acquired contemporaneously or near-contemporaneously – as close as possible to the same instantaneous acquisition time**. As a corollary, to clearly define and meet STV requirements, we must also understand acceptable maximum time offsets for successful data fusion across a range of processes, features and landcover types.

Coordinated airborne campaigns can offer near-contemporaneous, multi-modal observations for a subset of priority sites, often with complementary ground measurements (e.g., ASCENT, AfriSAR, SnowEx, Delta-X). These airborne campaigns often involve time offsets of hours to weeks between measurements from different aircraft or repeat passes with the same aircraft, and sometimes involve coordinated satellite observations.

Coincident datasets in existing archives

Over the past two decades, large archives of 3D surface elevation measurements have been acquired from a range of airborne and satellite platforms. Many of these datasets are now publicly available from NASA Earthdata Cloud and commercial cloud storage, enabling efficient, on-demand search, subsetting, and data access.

Many "coincident" or near-contemporaneous measurements in these archives were fortuitously acquired over the same location within a short temporal window (e.g., <1-14 days). These archives remain largely untapped for a broader, systematic identification and analysis of spatiotemporally coincident 3D measurements.

We are developing efficient, open-source tools to identify, curate, process and distribute coincident 3D datasets contained within these archives across a range of representative terrain and landcover types. These datasets can be used by the larger STV community for calibration/validation, fusion algorithm development, and discipline-specific scientific analysis.

Our initial search will identify a subset of precursor coincident datasets over the Contiguous United States (CONUS), Alaska and Hawaii in the following archives:

- USGS 3DEP airborne lidar (~2003 to present)
- NASA ICESat-2 photon-counting satellite lidar (October 2018 to present)
- NASA GEDI full-waveform satellite lidar (December 2018 to present)
- Maxar in-track very-high-resolution (VHR) satellite stereo (2007 to present)

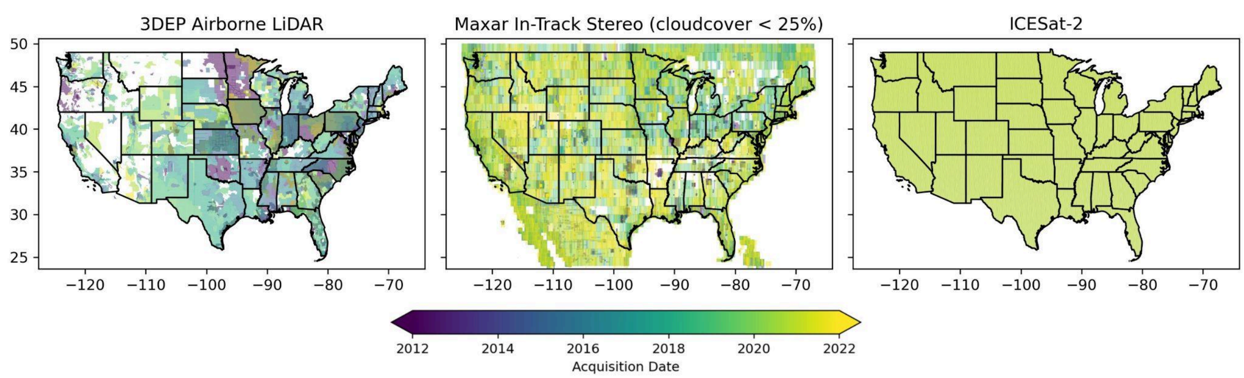


Figure 1: Preliminary evaluation of archive coverage for CONUS as of December 2023, with polygon color representing acquisition date. Left) 3DEP airborne lidar data (1-m DTM), center) Maxar in-track very-high-resolution (VHR) stereo pairs from Worldview-1/2/3 (30-50 cm GSD) with cloud cover < 25%, and right) ICESat-2 tracks.

Initial archive search efforts will focus on sites with all four source datasets acquired within a fixed temporal window, initially set for +/-14 days around the source lidar acquisition date range (<60 days). The flexible, open-source archive search tools will also support custom parameters, including user-defined temporal windows, as well as searches that return all possible records for a given area of interest with no temporal filter.

We will leverage public-facing APIs and STAC metadata for archive search whenever possible. This approach enables on-demand searches of the latest archives, as new records are added daily to some archives (Maxar, 3DEP), and with regular data releases for NASA missions.

Airborne lidar

Airborne laser scanning (ALS; airborne lidar) data is the “keystone” or primary "truth" dataset for the STV preliminary coincident datasets. The archives of publicly available airborne lidar data continue to grow, with increasing data quality (i.e., coverage, accuracy/precision, point density).

Initial archive search efforts will focus on the United States Geological Survey (USGS) 3D Elevation Program (3DEP, [*Stoker and Miller*, 2022]) archive for the Contiguous United States (CONUS), Hawaii and Alaska. The 3DEP team performs QA/QC and systematically processes airborne lidar data products from a range of vendors. As of September 2024, 2906 projects were available in the 3DEP archive, acquired since late 2000.

Most of the 3DEP collections are intentionally flown during leaf-off conditions in early spring (March-May) or late fall (September-November) for improved ground returns and terrain mapping applications. The acquisition periods vary across individual projects, from a few hours for smaller sites to multiple years for large-area mapping surveys (often involving strategic planning to avoid variable surface conditions, like seasonal snowcover).

Following initial support for the 3DEP archive, we will research and attempt to support other publicly available airborne lidar archives, prioritizing those with cloud-hosted data products, including:

- NEON Elevation - LiDAR https://data.neonscience.org/data-products/DP3.30024.001
- NASA G-LiHT https://gliht.gsfc.nasa.gov/index.php?section=34
- NASA LVIS https://lvis.gsfc.nasa.gov/Data/GE.html
- NASA Operation IceBridge (OIB)
- NCALM https://ncalm.cive.uh.edu/data/data-distribution
- OpenTopography https://portal.opentopography.org/dataCatalog
- International airborne lidar archives (Europe, Canada, New Zealand)

Airborne lidar processing

The 3DEP archive hosts both original vendor products (OPR), and standardized products. These include tiled point clouds (laz format) and tiled 1-m digital terrain model (DTM) raster products (Cloud-optimized GeoTiff, COG). In some cases, these also include tiled digital surface model (DSM) raster products (COG) and intensity raster products.

Our core data search tools will identify subsets of the relevant tiles with metadata containing urls for download or on-demand direct access to desired data products. Additional data access and processing tools will leverage custom workflows to process and deliver the source point clouds using Point Data Abstraction Library (PDAL) pipelines.

Satellite laser altimetry

Satellite laser altimetry (satellite lidar) offers global or near-global coverage with regular repeat intervals. Current NASA multi-beam profiling altimetry missions include the Ice Cloud and Elevation Satellite 2 (ICESat-2, [*Abdalati et al.*, 2010; *Markus et al.*, 2017]) and Global Ecosystems Dynamics Investigation (GEDI, [*Dubayah et al.*, 2020]).

Both data archives are hosted by NASA in AWS S3 (us-west-2). Initial spatiotemporal search will be performed using the NASA EarthData Common Metadata Repository (CMR) system, which contains records for each ICESat-2 and GEDI granule. Each granule contains a simplified polygon that approximates spatial extent, though actual measurement spatial coverage is smaller (e.g., CMR polygons for ICESat-2 granules are ~12 km wide compared to the actual 6.6 km separation between the outermost beams).

Satellite laser altimetry processing

Following initial search, we will use the SlideRule Earth service [*Shean et al.*, 2023] to efficiently subset, process in parallel, and deliver the ICESat-2 and GEDI products in user-friendly GeoDataFrame or GeoParquet formats. A second round of spatiotemporal intersection will refine coincident datasets for actual return locations and any data gaps due to clouds or other quality filters.

SlideRule Earth supports multiple, customizable ICESat-2 photon classification and aggregation approaches. Initial efforts will prepare ATL06-SR products derived from subsets of classified ATL03 photons for top of canopy and ground returns. Initial GEDI data will include Level-2A records, with canopy top elevation and lowest mode ground elevation, among other height metrics derived from the Level-1B geolocated waveforms.

Satellite stereo

At the time of writing, there are over 40 electro-optical very-high-resolution (VHR, <1 m ground sample distance [GSD]) imaging satellites in orbit, operated by several commercial companies. Since the launch of WorldView-1 in September, 2007, Maxar (formerly DigitalGlobe) has systematically acquired in-track VHR stereo satellite images with global coverage and regular refresh. In-track stereo collections involve systematic off-nadir acquisitions of two (or more) images by a single instrument on the same orbital pass, separated by ~60-90 seconds. The parallax and relief displacement of surface features in the images acquired from different perspectives provides depth information. Dense correspondences for all pixels in the two images and ephemeris/attitude information enable triangulation to determine the 3D positions on the ground.

The current Maxar archive includes hundreds of thousands of individual in-track stereo acquisitions from WorldView-1, Worldview-2, WorldView-3, and GeoEye-1, all with GSD of 0.3-0.5 m, swath widths of ~13-17 km, and lengths of ~15-110 km).

We prepared preliminary notebooks to query and quickly identify in-track stereo data using Maxar's new image archive API, Maxar Geospatial Platform (MGP).

The commercial stereo image ordering and processing represents the current bottleneck for direct data analysis and intercomparison. We will order a subset of the Maxar in-track stereo images identified during the archive search process leveraging existing federal licensing agreements, and obtain permission to redistribute derived 3D products.

Satellite stereo processing

We will leverage the latest processing workflows for the NASA Ames Stereo Pipeline [*Shean et al.*, 2016, 2021] to perform bundle adjustment using the near-contemporaneous airborne lidar DSM products as a reference to ensure product self-consistency. We will generate dense point clouds (laz) and gridded DEMs (COG) for each acquisition, with per-pixel error and uncertainty products.

Following initial support for the Maxar archive, we will research and attempt to support other publicly available satellite stereo archives, prioritizing those with cloud-hosted data products, including:

- Planet SkySat-C
- BlackSky Global
- Airbus - Pleiades-HR/NEO, SPOT

Airborne and Satellite InSAR

Following initial efforts to support the above primary datasets, we will research and attempt to support other publicly available InSAR product archives, prioritizing those with cloud-hosted data products, including:

- NASA UAVSAR
- 3DEP Airborne IfSAR (Alaska)
- Airbus - TerraSAR-X/TanDEM-X

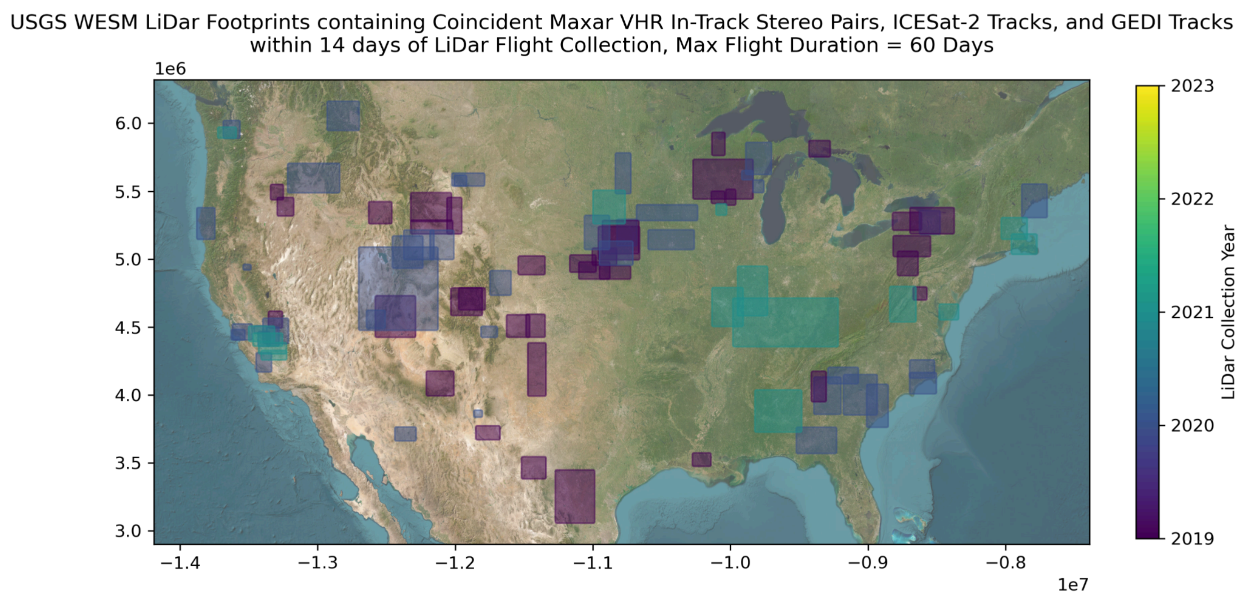
Data Processing

After compiling the 3D point cloud and raster products from the various archives for each precursor coincident dataset site, we will perform the necessary steps to enable direct integration, comparison and analysis.
This requires epoch-aware 3D coordinate reference system (CRS) transformations to ensure that all datasets share a common global CRS and epoch. For example, independent of any real change, there is a ~2.1 m 3D offset between 3DEP products distributed using the NAD83(2011)

2010.0 epoch, and contemporaneous ICESat-2 and Maxar products, which use ITRF2014/2020 and modern realizations of the WGS84 ellipsoid (e.g., G1762, G2319). Beyond this, we will identify and correct any vertical datum offsets (e.g., height above NAVD88 vertical datum vs. ellipsoid height).
We will then perform robust 3D co-registration of all coincident data products using the airborne lidar data as a reference, minimizing 3D residuals over high-confidence static control surfaces. The co-registration workflows will be streamlined, tested, and QA/QC'd for a range of terrain and landcover types.
Final co-registered datasets will be delivered to NASA DAAC for archiving and distribution, for users who would prefer to download samples, rather than run on-demand data preparation notebooks.

Expected coverage

Initial archive search efforts identified many near-contemporaneous collections for 3DEP airborne lidar, Maxar in-track stereo images, ICESat-2 satellite laser altimetry and GEDI satellite laser altimetry (Figure 2).



	LiDar Project Name	LiDar Collect Start Date	LiDar Collect End Date	Maxar VHR In-Track Stereo Pair Count	ICESat-2 ATL06 Granule Count	GEDI02A Granule Count
0	NE_PostSpringFlood_OrthoLidar_2019_D19	2019-05-25	2019-07-13	54	7	16
1	CA_YosemiteNP_2019_D19	2019-10-07	2019-10-23	20	2	1
2	UT_StatewideKane_2020_A20	2019-10-11	2019-11-16	14	7	18
3	CO_NESEColorado_2019_C20	2019-07-31	2019-09-08	12	5	8

Figure 2: Map showing preliminary subset of coincident datasets for CONUS, identified using 3DEP airborne lidar collections with short acquisition periods (<60 days) and a +/-14 day window for cloud-free WorldView in-track stereo images, ICESat-2, and GEDI acquisitions. There are 86 total sites that meet these criteria, and the table shows the first four records.

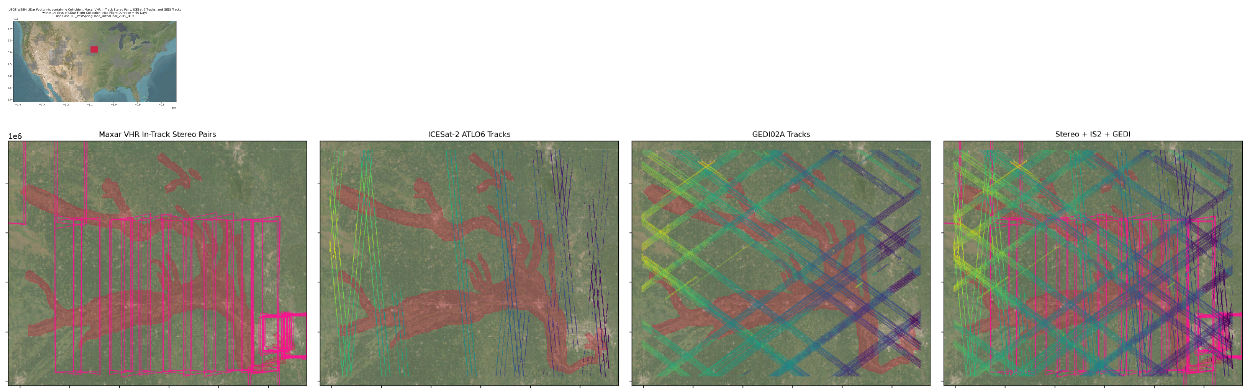


Figure 3: Example coincident dataset coverage for the site corresponding to the NE_PostSpringFlood_OrthoLidar_2019_D19 3DEP airborne lidar project (red box on CONUS map), covering 36 counties in Nebraska between May 25, 2019 and July 13, 2019 ([vendor report](#)). Panels show the 3DEP lidar footprint (red), with cloud-free Maxar VHR in-track stereo acquisitions (pink), ICESat-2 ATL06-SR elevation values (purple to yellow color ramp showing surface elevation values), GEDI02A elevation values (same color ramp), and a combined map of overlapping coverage.

The final number of sites will depend on refined search parameters, specifically the acceptable temporal offset between collections. We expect to deliver final co-registered data products for at least ~10-20 sites across a range of priority STV landcover and terrain characteristics.

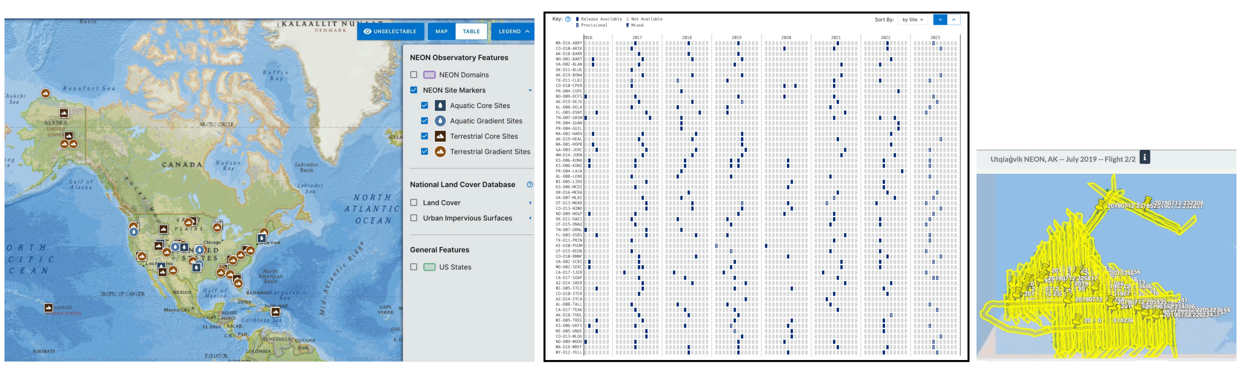


Figure 4: Map of ~60 NEON lidar collections sites with ~annual collection interval, and individual time-tagged flightline polygons for sample site (July 2019 acquisition for Utqiagvik, AK site), which will offer improved spatiotemporal search granularity. Similar swath polygons should be available for 3DEP flights acquired after August 2020 (https://www.usgs.gov/media/files/lidar-base-specification-2020-rev).

Deliverables (summer 2025)

- Open-source, version-controlled Github repository (with documentation) containing:
  - Jupyter notebooks and Python libraries for generic cross-archive spatiotemporal search and identification of coincident datasets.

        - This includes both broad, full-archive search capabilities, and flexible searches for smaller, user-defined areas of interest.
        - These notebooks will allow users to rerun queries using updated archives, and to customize search parameters (like acceptable temporal offset) based on local site and/or processes requirements.
    - Jupyter notebooks and Python libraries offering efficient data subsetting, processing and direct retrieval of cloud-hosted datasets, with basic functions for inter-comparison and analysis. This will allow users to focus on data fusion and model development
- Samples of co-registered precursor coincident datasets, in cloud-optimized formats, ready to be ingested and hosted by DAAC for the larger STV community

Contact Information

Please reach out with any questions, comments, suggestions. We are eager to develop flexible tools and deliver curated datasets that are most useful for the broader community.

PI: David Shean, dshean@uw.edu
University of Washington

References

Abdalati, W., H. J. Zwally, R. Bindschadler, B. Csatho, S. L. Farrell, H. A. Fricker, D. Harding, R. Kwok, M. Lefsky, T. Markus, A. Marshak, T. Neumann, S. Palm, B. Schutz, B. Smith, J. Spinhirne, and C. Webb (2010), The ICESat-2 Laser Altimetry Mission, *Proc. IEEE*, *98*(5), 735–751, doi:10.1109/JPROC.2009.2034765.

Donnellan, A., D. Harding, P. Lundgren, K. Wessels, A. Gardner, and M. Simard (2021), *Observing Earth's Changing Surface Topography and Vegetation Structure: A Framework for the Decade, NASA Surface Topography and Vegetation Incubation Study*.

Dubayah, R., J. B. Blair, S. Goetz, L. Fatoyinbo, M. Hansen, S. Healey, M. Hofton, G. Hurtt, J. Kellner, S. Luthcke, J. Armston, H. Tang, L. Duncanson, S. Hancock, P. Jantz, S. Marselis, P. L. Patterson, W. Qi, and C. Silva (2020), The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography, *Sci. Remote Sens.*, *1*, 100002, doi:10.1016/j.srs.2020.100002.

Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, and D. Alsdorf (2007), The Shuttle Radar Topography Mission, *Rev. Geophys.*, *45*(2), doi:10.1029/2005RG000183.

Markus, T., T. Neumann, A. Martino, W. Abdalati, K. Brunt, B. Csatho, S. Farrell, H. Fricker, A. Gardner, D. Harding, M. Jasinski, R. Kwok, L. Magruder, D. Lubin, S. Luthcke, J. Morison, R. Nelson, A. Neuenschwander, S. Palm, S. Popescu, C. Shum, B. E. Schutz, B. Smith, Y. Yang, and J. Zwally (2017), The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation, *Remote Sens. Environ.*, *190*, 260–273, doi:10.1016/j.rse.2016.12.029.

Shean, D., S. Bhushan, E. Berthier, C. Deschamps-Berger, S. Gascoin, and F. Knuth (2021), Chamoli Disaster Post-event 2-m DEM Composite (February 10-11, 2021) and Difference Map, , doi:10.5281/zenodo.4558692.

Shean, D., J. p Swinski, B. Smith, T. Sutterley, S. Henderson, C. Ugarte, E. Lidwa, and T. Neumann (2023), SlideRule: Enabling rapid, scalable, open science for the NASA ICESat-2 mission and beyond, *J. Open Source Softw.*, *8*(81), 4982, doi:10.21105/joss.04982.

Shean, D. E., O. Alexandrov, Z. M. Moratto, B. E. Smith, I. R. Joughin, C. Porter, and P. Morin (2016), An automated, open-source pipeline for mass production of digital elevation models (DEMs) from very-high-resolution commercial stereo satellite imagery, *ISPRS J. Photogramm. Remote Sens.*, *116*, 101–117, doi:10.1016/j.isprsjprs.2016.03.012.

Stoker, J., and B. Miller (2022), The Accuracy and Consistency of 3D Elevation Program Data: A Systematic Analysis, *Remote Sens.*, *14*(4), 940, doi:10.3390/rs14040940.