



OPEN

DATA DESCRIPTOR

High-resolution repeat topography of drifting ice floes in the Arctic Ocean from terrestrial laser scanning

David Clemens-Sewall^{1,2,✉}, Chris Polashenski^{1,3}, Ian A. Raphael¹, Matthew Parno⁴, Don Perovich¹, Polona Itkin⁵, Matthias Jaggi⁶, Arttu Jutila^{7,8}, Amy R. Macfarlane⁶, Ilkka S. O. Matero^{7,9}, Marc Oggier¹⁰, Ronald J. W. Visser¹¹ & David N. Wagner^{6,12}

Snow and ice topography impact and are impacted by fluxes of mass, energy, and momentum in Arctic sea ice. We measured the topography on approximately a 0.5 km² drifting parcel of Arctic sea ice on 42 separate days from 18 October 2019 to 9 May 2020 via Terrestrial Laser Scanning (TLS). These data are aligned into an ice-fixed, lagrangian reference frame such that topographic changes (e.g., snow accumulation) can be observed for time periods of up to six months. Using *in-situ* measurements, we have validated the vertical accuracy of the alignment to ± 0.011 m. This data collection and processing workflow is the culmination of several prior measurement campaigns and may be generally applied for repeat TLS measurements on drifting sea ice. We present a description of the data, a software package written to process and align these data, and the philosophy of the data processing. These data can be used to investigate snow accumulation and redistribution, ice dynamics, surface roughness, and they can provide valuable context for co-located measurements.

Background & Summary

Repeat Terrestrial Laser Scanning (TLS) topography measurements were part of the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAIC) expedition, in which researchers aboard *R/V Polarstern*¹ drifted with and studied the same collection of ice floes in the Central Arctic from October 2019 to May 2020^{2–4}. Arctic sea ice has grown dramatically younger and thinner in recent decades⁵ and the overall objectives of MOSAIC were to understand the causes and consequences of this ‘new Arctic’. To do so, researchers were divided into teams studying the snow and ice (including on-ice and satellite remote sensing)², atmosphere³, ocean⁴, ecosystem, and biogeochemistry. We conducted the TLS measurements as part of the ice team, for the primary purpose of quantifying snow accumulation and redistribution. Other applications of these data include observations of ice dynamics; surface roughness for ice-atmosphere interactions; and providing context for atmospheric observations, remote sensing instruments (e.g., on-ice radars⁶), autonomous buoys, snow pit measurements, and more.

Snow substantially affects the Arctic sea ice mass balance due to its opposing impacts of insulating in the winter (restraining ice growth) and reflecting shortwave radiation in the summer (protecting against ice melt⁷). Three of the four most important uncertainties impacting September sea ice volume in the CICE sea ice model⁸ are related to the thermal and optical properties of the snow⁹. Snow spatial variability due to wind driven snow redistribution impacts these properties^{10,11}. However, redistribution is challenging to measure due to this spatial

¹Thayer School of Engineering at Dartmouth College, Hanover, NH, USA. ²NSF National Center for Atmospheric Research, Boulder, CO, USA. ³USACE-CRREL Fort Wainwright, Fairbanks, AK, USA. ⁴Solea Energy, Thetford, VT, USA. ⁵UiT The Arctic University of Norway, Tromsø, Norway. ⁶WSL Snow and Avalanche Research Institute SLF, Davos, Switzerland. ⁷Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, 27570, Germany. ⁸Finnish Meteorological Institute, Helsinki, Finland. ⁹Svalbard Integrated Arctic Earth Observing System Knowledge Centre, Longyearbyen, Norway. ¹⁰International Arctic Research Center, University of Alaska Fairbanks, Fairbanks, AK, USA. ¹¹GroenDoen, Westerwijtwerd, The Netherlands. ¹²CRYOS, School of Architecture, Civil and Environmental Engineering, EPFL, Lausanne, Switzerland. ✉e-mail: dcsewall@ucar.edu

variability. Repeat observations of snow changes on a substantial area of the same piece of ice are needed to observe these redistribution processes and understand their mechanisms. Furthermore, the magnitude of snow accumulation can be very small. For example, the largest snowfall event on MOSAiC precipitated just 1.6 cm water equivalent¹². Thus, our observations of changes must be highly accurate. Finally, to observe snow accumulation and redistribution throughout the winter, measurements must be feasible in polar night.

TLS is routinely used to make highly-accurate measurements of snow accumulation on areas of 40 m² to 0.5 km²^{13–17}. The instrument is a laser scanner that is mounted on a ~2.2 m tall tripod. Due to the generally flat topography of sea ice and shadowing, the instrument collects topographic information up to 100–200 m from itself. To observe a larger area, we relocate the tripod and collect measurements from different locations, which we then co-register into a common reference frame. TLS measurements on Arctic sea ice face unusual challenges and our procedures for the MOSAiC Expedition were informed by prior experiments, including: the Seasonal Ice Zone Observing Network project¹⁸; the Snow, Wind, and Time project¹⁷; and the Sea Ice Dynamics Experiment. First, the typical temperatures of –15 to –35 °C are below the operating range of commercially-available TLS instruments. We addressed this issue with a custom-designed heater case. Second, TLS measurements are typically aligned in a geodetic reference frame via highly-accurate GNSS measurements. However, on drifting sea ice, the relevant reference frame for snow processes is a lagrangian reference frame fixed to the surface of the ice. We developed a custom software package—pydar¹⁹—to align repeat TLS measurements into this lagrangian, ice-fixed reference frame. To verify that our alignment achieved the necessary vertical accuracy for snow accumulation, we statistically validated it through comparison with *in-situ* measurements²⁰. The quantitative results from our alignment validation are specific to this dataset. Finally, the TLS data contain potentially-useful information that are irrelevant to our needs (e.g., sub-cm snow surface roughness in areas near the scanner, backscatter reflectance, roughness relevant to aerodynamic drag, etc...). We hope that future researchers will use these data for purposes that we have not imagined. To facilitate this future usage, we have designed pydar to preserve the full scope of the data and to make our data processing decisions transparent to future researchers.

The primary purpose of this manuscript is to describe the TLS data collected at MOSAiC. However, given the unique challenges of using TLS on drifting sea ice, some discussion of lessons learned and future methodological developments is warranted. Prior experience on sea ice near Utqiagvik, AK^{17,18} found that spacing scan positions between 150 and 200 m apart generally produced acceptable data. However, the ice at MOSAiC was rougher than those experiments, and the placement of scan positions was also restricted due to not trespassing in sensitive measurement sites. Under these constraints, we found that spacing scan positions around 100 m apart produced better data, although we sometimes prioritized having measurements co-located with complementary measurements sites (e.g., the snow and ice thickness transects) over achieving the best TLS coverage. Additionally, the alignment procedures and validation were developed after the expedition, thus the alignment validation relies on what coincident *in-situ* measurements were available and had not experienced blowing snow events between the *in-situ* measurement and the TLS acquisition. The statistical validation methods presented herein are applicable for future campaigns, but the specific quantitative results depend on the measurements. Therefore, validation should be conducted for any TLS campaign on sea ice. For future campaigns, we strongly recommend including more *in-situ* point measurements of snow surface changes coincident with the TLS measurements. One expeditious approach for this would be to include a small array of snow thickness stakes near each reflector post. This would ensure that the *in-situ* measurement sites were visible from multiple scan positions, and the snow surface measurements could be made quickly while distributing the reflectors at the start of a TLS measurement day.

This manuscript describes the repeat TLS data collected on the MOSAiC expedition from October 2019 to May 2020²¹. We present the philosophy of the data processing, and its implementation in pydar¹⁹. Finally, we validate the vertical alignment and discuss considerations for reuse of these data.

Methods

Terminology. We use the following terms throughout this manuscript. They are mostly drawn from their usage in RiSCAN (Riegl's software for acquiring and processing TLS data):

- Scan Position: set-up the tripod at a given location and measure the topography within the scanner's line of sight.
- SingleScan: the data collected from a single scan position. We use this term to refer to both the point cloud of topographic measurements from this scan position and ancillary data such as the locations of TLS reflectors within the scanner's reference frame at this scan position and the rigid transformations that register and align this SingleScan with others (see below).
- Project: a collection of SingleScans covering a contiguous area that were collected during a sufficiently short time interval such that no topographic change occurred between scan positions (sometimes ice deformation occurred during a Project, these exceptions are described in the Usage Notes). Typically, the set of SingleScans in a Project were collected in a single day of measurements although on some occasions measurements were collected over two days.
- Registration: the act of computing the rigid transformations that describe the spatial relationships between the different SingleScans in a Project. Registration places all SingleScans in a Project into a common reference frame (whose origin and unit vectors are typically defined by an arbitrary SingleScan).
- Scan Area: a region of ice whose topography we measured over time with a succession of Projects.
- Alignment: the act of computing the rigid transformations such that SingleScans from different Projects in the same Scan Area are in a common reference frame. Alignment is necessary to precisely locate topographic changes (e.g., how much snow accumulation occurred at specific location on the ice from 4 January to 11 March).



Fig. 1 Photo showing data acquisition of a typical scan position. The Riegl VZ1000 terrestrial laser scanner is inside the custom heater case (metal cylinder) on top of the tripod. The researcher controls the scanner from the laptop via WiFi. To the right of the tripod is a Riegl 10 cm cylinder reflector mounted on a post frozen into the ice. The Polarstern can be seen in the background along with other installations of the ice camp including Balloon Town³ (orange tents) and Ocean City⁴ (blue and white tent). Photo credit: S. Svavarsdóttir.

Data collection. We used a Riegl VZ1000, which has an eye-safe, near infrared laser (1550 nm). For each scan position, we mounted the scanner on a tripod, and the scanner rotated on vertical and horizontal axes to create a point cloud of its surroundings. The scanner was controlled via WiFi from a field laptop. The angular step-widths in the azimuthal and vertical directions were each 0.025° and it took approximately 8 minutes to acquire a point cloud with a Laser Pulse Repetition Rate of 300 kHz. The origin and unit vectors of the point cloud are defined relative to the scanner and this reference frame is named the Scanner's Own Coordinate System (SOCS). Air temperature was typically below the VZ1000's minimum operating temperature, so we placed the scanner in a custom-designed heater case (Fig. 1) to maintain its temperature within acceptable bounds. Due to the overall flat topography, occlusions, and low reflectivity of snow and ice at 1550 nm, the scanner collects useful data up to 100–200 m from its location. To map a larger area, on each measurement day we relocated the scanner to additional scan positions and acquired subsequent measurements (i.e., we collected a SingleScan at each scan position), which we then linked together into a Project. Choosing the locations for scan positions was a trade-off between maximizing the area covered, co-locating with other measurements, minimizing shadows within the measured area, and not trespassing in sensitive measurement sites. In order to make a complete map, we registered the SingleScans measured from each scan position into a common reference frame. We placed Riegl 10 cm cylinder reflectors on posts frozen into the ice (Figs. 1, 2), and used them to locate and orient each SingleScan in the Project's common reference frame (described below in Data Processing in RiSCAN). Typically between 4 and 10 reflectors were visible from each scan position. These reflectors also served as the starting point for aligning Projects collected in the same Scan Area on different days into a lagrangian, ice-fixed reference frame.

Scan areas. Repeat TLS observations on MOSAiC were focused on three primary Scan Areas: Snow1, Snow2, and ROV (Fig. 2 and Table 1). These Scan Areas were selected to observe a variety of ice topography, to co-locate with other measurements, and to be logistically accessible. In March, ice dynamics caused the Met City atmospheric measurement site³ and the on-ice Remote Sensing measurement site² to be removed from the Snow1 scan area (where they had been located from October through February). Additional TLS measurements were made at these sites (which we labelled 'RS' for Remote Sensing and 'MET' for Met City; Table 1). Below are descriptions of each Scan Area.

Snow1 Scan Area. The Snow1 Scan Area (hereafter we use 'Snow1' to refer specifically to the TLS measurement area) was generally off the bow and port side of Polarstern and composed of residual ice of which only the upper 30 cm was solid when we arrived in October²², refrozen melt ponds, first year ice in refrozen leads, and first year ridges. We first measured Snow1 on 18 October and our final measurement was on 3 May. Ice dynamics caused frequent (multiple times per month) crack and ridge formation in the Scan Area. These dynamics resulted in substantial variation in the area of ice measured and the number of SingleScans collected in each Project. Snow1 was co-located with the SLoop snow and ice thickness transect²³, the Snow1 snow sampling area², the BGC3 ice coring area, the Met City atmospheric measurement site³ (18 October to 28 February), the first on-ice Remote Sensing site² (18 October to 15 November), the second on-ice Remote Sensing site² (6 December to 28 February), the Ocean City oceanic measurement site⁴ (18 October to 6 December), the Bow Stakes mass balance site²⁰ (6 December to 28 February), the Met Stakes mass balance site²⁰ (18 January to 3 May), the Stakes3 mass balance site²⁰ (6 December to 26 April), the second Remotely Operated Vehicle site² (1 November to 6 December), and a number of vibrating wire ice stress gauges². The given dates are the first and last dates that the installations were present in the TLS data, not necessarily the installation or decommission dates of the installations.

Snow2 Scan Area. The Snow2 Scan Area (hereafter we use 'Snow2' to refer specifically to the TLS measurement area) was generally off the bow and starboard side of Polarstern beyond the logistics area² and was composed of refrozen melt ponds and an approximately 1 m tall (on average) second year ridge. No ice thickness measurements were made in Snow2, but the residual ice was likely thicker than that in Snow1. Similar level ice in the

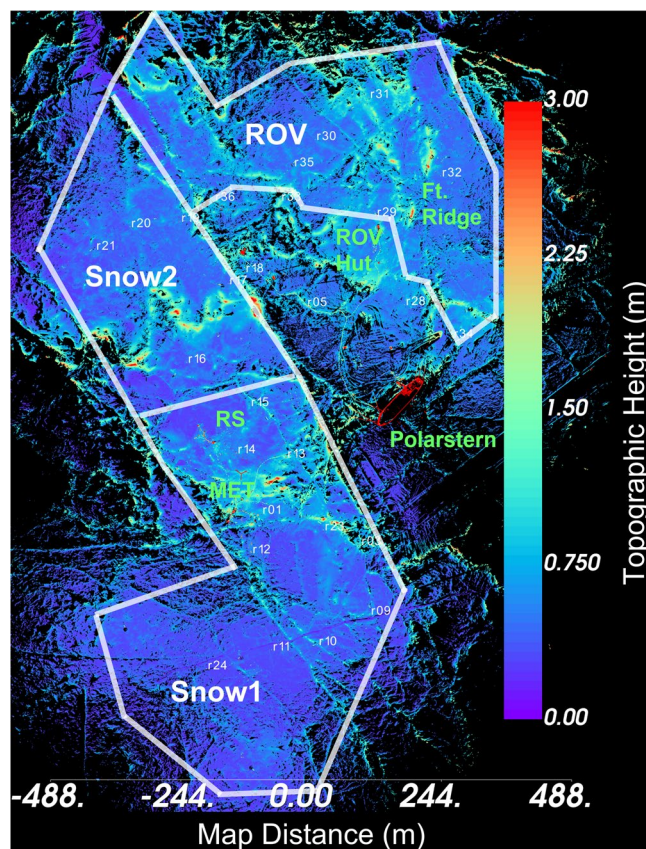


Fig. 2 Map of the three Scan Areas in the MOSAiC Central Observatory² in early February. The Scan Areas are labelled and roughly outlined in white (the exact extent of the measured areas changed due to ice deformation and the number of SingleScans conducted). Notable installations are indicated in green (see MOSAiC snow and ice overview publication² for more installations). Reflector locations are also shown in white. Reflector heights are provided in Table 2. This map is a rendering of the combined point cloud data from Projects mosaic_rov_040220.RiSCAN (collected on 4 February) and mosaic_01_060220.RiSCAN (collected on 6–7 February). The origin for topographic height is the surface of recently frozen lead (so the values approximately correspond to height above the sea surface).

NLoop transect had a modal thickness of approximately 75 cm in early November²³. We first measured Snow2 on 6 November and our final measurement was on 9 May. Snow2 was the most stable region observed. The core of the measurement area was not deformed from 6 November to 9 May with the exception of a 1-m-wide crack that formed between 13 November and 6 December. The measurement area progressively shrank due to ice dynamics removing areas in November, March, April, and May. On two occasions—6 December and 6 February—we made measurements on Snow1 and Snow2 as part of the same Project. Snow2 was co-located with the Snow2 snow sampling area², an atmospheric flux chamber measurement site, a section of the NLoop snow and ice thickness transect²³ (6 November to 6 December), the Stakes1 mass balance site²⁰ (only on 6 December, thereafter this was part of the ROV Scan Area), and the Alli's ridge measurement site² (only on 6 December, thereafter this was part of the ROV Scan Area).

ROV Scan Area. The ROV Scan Area (hereafter referred to as 'ROV', it was named for the Remotely Operated Vehicle installation) was generally off the stern and starboard side of Polarstern beyond the logistics area² and was composed of deformed second year ice, refrozen melt ponds, level first year ice, and first year ridges. We first measured the ROV Scan Area on 4 January and our final measurement was on 9 May. In early January, the modal ice thickness on both level first year ice²⁰ and level second year ice²³ was approximately 1 m. In March, ice dynamics displaced the first year ice region out of the ROV Scan Area and demolished the Ft. Ridge measurement site. These dynamics also produced young ice in a 40-m-wide, refrozen lead. The core region of ROV was connected to Snow2 from 6 December to 9 May. On two occasions—4 April and 9 May—we made measurements on Snow2 and ROV as part of the same project. ROV was co-located with the third Remotely Operated Vehicle site², the NLoop snow and ice thickness transect²³, the Ft. Ridge measurement site² (4 January to 18 March), half of the Alli's Ridge measurement site² (4 January to 22 February), the Ridge Ranch mass balance site²⁰ (19 January to 11 March), the Stakes4 mass balance site²⁰ (4 January to 22 February), the David's Ridge measurement site² (14 March to 9 May), the ROV3 broadband albedo transect²⁴ (11 April to 9 May), the SYI broadband albedo transect²⁴ (29 April to 9 May), and a number of vibrating wire ice stress gauges².

Project Name	Date	Scan Area(s)	# SingleScans	Location
mosaic_01_101819.RiSCAN	2019-10-18	Snow1	2	84.79°N, 133.09°E
mosaic_01_102019.RiSCAN	2019-10-20	Snow1	5	84.97°N, 132.76°E
mosaic_01_102519.RiSCAN	2019-10-25	Snow1	5	85.44°N, 128.15°E
mosaic_01_110119.RiSCAN	2019-11-01	Snow1	7	85.80°N, 122.97 ° E
mosaic_02_110619.RiSCAN	2019-11-06	Snow2	7	85.93°N, 117.74°E
mosaic_01_110819.RiSCAN	2019-11-08	Snow1	3	85.92°N, 116.53°E
mosaic_02_111319.RiSCAN	2019-11-13	Snow2	8	86.10°N, 117.84°E
mosaic_01_111519.RiSCAN	2019-11-15	Snow1	8	86.16°N, 118.20°E
mosaic_01b_061219.RiSCAN.RiSCAN.RiSCAN	2019-12-06 ^a	Snow1, Snow2	12	86.14°N, 122.25°E
mosaic_01_122719.RiSCAN	2019-12-27	Snow1	5	86.68°N, 115.95°E
mosaic_rov_040120.RiSCAN	2020-01-04	ROV	6	86.99°N, 115.44°E
mosaic_01_040120.RiSCAN	2020-01-04 ^a	Snow1	7	87.01°N, 115.35°E
mosaic_rov_110120.RiSCAN	2020-01-11	ROV	5	87.21°N, 111.40°E
mosaic_01_180120.RiSCAN	2020-01-18	Snow1	8	87.41°N, 98.27°E
mosaic_rov_190120.RiSCAN	2020-01-19	ROV	8	87.40°N, 98.36°E
mosaic_rov_250120.RiSCAN	2020-01-25	ROV	11	87.41°N, 92.79°E
mosaic_01_290120.RiSCAN	2020-01-29	Snow1	8	87.47°N, 95.18°E
mosaic_rov_040220.RiSCAN	2020-02-04	ROV	9	87.48°N, 95.30°E
mosaic_01_060220.RiSCAN	2020-02-06 ^a	Snow1, Snow2	16	87.60°N, 94.04°E
mosaic_01_150220.RiSCAN.RiSCAN	2020-02-15	Snow1	8	88.07°N, 79.87°E
mosaic_rov_220220.RiSCAN.RiSCAN	2020-02-22	ROV	9	88.58°N, 64.54°E
mosaic_01_280220.RiSCAN	2020-02-28	Snow1	7	88.34°N, 33.99°E
mosaic_rov_110320.RiSCAN	2020-03-11	ROV	1	88.27°N, 31.95°E
mosaic_rov_140320.RiSCAN.RiSCAN	2020-03-14	ROV	4	87.67°N, 24.59°E
mosaic_rov_180320.RiSCAN	2020-03-18	ROV	4	87.14°N, 17.28°E
mosaic_01_220320.RiSCAN	2020-03-22	Snow1	6	86.23°N, 15.70°E
mosaic_02_260320.RiSCAN	2020-03-26	Snow2	5	86.00°N, 13.09°E
mosaic_02_300320.RiSCAN.RiSCAN	2020-03-30	Snow2	5	85.38°N, 13.20°E
mosaic_02_040420.RiSCAN	2020-04-04 ^a	Snow2, ROV	11	84.67°N, 12.89°E
mosaic_01_080420.RiSCAN	2020-04-08	Snow1	4	84.48°N, 14.64°E
mosaic_01_080420b.RiSCAN	2020-04-08	Snow1	3	84.48°N, 14.64°E
mosaic_02_110420_rov.RiSCAN	2020-04-11	ROV	6	84.33°N, 14.75°E
mosaic_02_130420_2.RiSCAN	2020-04-13	Snow2	6	84.29°N, 14.96°E
mosaic_rov_170420.RiSCAN	2020-04-17	ROV	6	84.41°N, 13.67°E
mosaic_rs_170420.RiSCAN	2020-04-17	RS	1	84.41°N, 13.67°E
mosaic_rov_220420.RiSCAN	2020-04-22	ROV	6	84.13°N, 15.89°E
mosaic_rs_220420.RiSCAN	2020-04-22	RS	1	84.13°N, 15.89°E
mosaic_02_230420.RiSCAN	2020-04-23	Snow2	6	84.08°N, 16.05°E
mosaic_01_250420.RiSCAN.RiSCAN	2020-04-25	Snow1	4	84.00°N, 15.59°E
mosaic_01_260420.RiSCAN	2020-04-26	Snow1	3	83.93°N, 15.51°E
mosaic_rov_290420.RiSCAN	2020-04-29	ROV	7	84.02°N, 17.24°E
mosaic_02_300420.RiSCAN	2020-04-30	Snow2	6	83.94°N, 17.42°E
mosaic_rs_300420.RiSCAN	2020-04-30	RS	1	83.94°N, 17.42°E
mosaic_01_030520.RiSCAN	2020-05-03	Snow1	4	83.89°N, 17.96°E
mosaic_met_040520.RiSCAN	2020-05-04	MET	2	83.91°N, 18.35°E
mosaic_rov_02_090520.RiSCAN	2020-05-09	Snow2, ROV	12	83.79°N, 14.09°E

Table 1. Summary of repeat TLS measurements from October 2019 to May 2020 on MOSAiC. ^aSome SingleScans were collected on the following day.

Data Processing in RiSCAN. TLS data acquisition and initial post-processing steps were conducted in RiSCAN PRO (Riegl's software for TLS data acquisition and processing: <http://www.riegl.com/products/software-packages/riscan-pro/>) following standard protocols described in the user manual. The following steps were conducted for each Project. First, an arbitrary SingleScan was chosen to be the origin of the Project. This SingleScan was levelled to establish the horizontal axis plane using the VZ1000's onboard inclination sensors. Next, another SingleScan was registered to the first by computing the rigid transformation that minimizes the sum of the least-squares error in the positions of pairs of reflectors (a.k.a. 'keypoints') observed in both SingleScans. We repeated this process until all SingleScans in the Project are registered. Finally, we refined the registration of each SingleScan except for the origin with RiSCAN's 'Multi-Station Adjustment'. In this process,

Reflector	Pole Height (m)	Terrain Relative Height (m)
r01	2.10	1.38
r03	2.03	0.99
r05	2.01	1.10
r09	1.72	1.05
r10	1.71	1.13
r11	1.81	1.51
r12	1.79	1.26
r13	2.89	1.67
r14	2.06	1.60
r15	2.47	1.73
r16	1.96	1.57
r17	1.96	1.54
r18	1.93	1.16
r19	2.06	1.56
r20	1.90	1.44
r21	1.73	1.38
r23	1.99	1.46
r24	1.85	1.55
r28	1.80	1.24
r29	1.77	1.30
r30	1.63	1.21
r31	2.36	1.48
r32	1.98	1.53
r33	2.86	1.49
r34	2.84	1.44
r35	2.09	1.53
r36	2.30	1.47

Table 2. Reflector heights both relative to the local snow surface (Terrain Relative Height) and relative to the approximately the sea surface (Pole Height). Pole Height is in the same reference frame as Fig. 2.

meter-scale planar facets were extracted from each SingleScan. Then, pairs of overlapping facets and pairs of reflectors between the SingleScans were all used as keypoints and an optimization procedure adjusted the rigid transformations of each SingleScan (except the origin) in order to minimize the sum of the least-squares error between all keypoints. Unlike in urban environments where walls, roads, and other human-made objects provide large planar facets, planar surfaces at MOSAiC were mostly meter-scale or smaller wind-scoured snow surfaces. When conducting Multi-Station Adjustment, we found that using a search radius (maximum distance between potential keypoints) between 0.3 and 0.7 m produced the best results. Search radii larger than this tended to match planar facets that were not truly co-planar (e.g., different faces of a snow dune), which causes misalignment.

The data for each Project was stored in a directory with the same name as the Project. All relevant parameters were exported from RiSCAN into open formats. The rigid transformation for each SingleScan is represented by a 4×4 matrix (using homogeneous coordinates²⁵) that is named the Scanner's Own Position (SOP) matrix. The SOP matrices for each SingleScan were exported into tab-delimited, DAT files in the Project directory (e.g., 'ScanPos001.DAT'). We gave each reflector a unique identifier (e.g., 'r01') that was consistent across all of the Projects. We exported the reflector positions (in the Project's reference frame) for each Project into a comma-delimited file (named 'tiepoints.csv') in the Project directory. Finally, we exported the point cloud data itself for each SingleScan in LAS 1.4 format into a subdirectory named 'lasfiles'. LAS 1.4 (<https://github.com/ASPRSorg/LAS>) is an open, community standard for point cloud data that is maintained by the American Society for Photogrammetry and Remote Sensing.

Data processing in pydar. *Overview of pydar.* The overall objective of pydar is to align SingleScans from different Projects in the same Scan Area into a lagrangian, ice-fixed reference frame such that one can observe topographic changes (e.g., snow deposition and erosion) over time. Furthermore, we sought to facilitate re-use of these data by preserving features of the data even when they are unimportant for our use case (e.g., cm-scale point density near the scanner, backscatter reflectance data, etc...) and enabling future researchers to revise our alignment of SingleScans (e.g., if they design a superior alignment procedure). This section provides a brief description of key features of pydar¹⁹ and the steps taken to process the repeat TLS data from MOSAiC. Additional functionality and implementation details can be found in the code documentation. To achieve these goals, pydar has an object-oriented design that mimics the hierarchical structure of the TLS data. And, pydar distinguishes between the spatial relationships of TLS points measured from the same scan position (i.e., within a SingleScan) and the spatial relationships of TLS points measured from different scan positions (i.e., different SingleScans in a Project).

pydar is implemented primarily in Python²⁶, with substantial use of the Numpy²⁷, Scipy²⁸, and VTK²⁹ libraries. Some functionality is implemented in Cython³⁰.

The core of pydar consists of four related classes: `SingleScan`, `Project`, `ScanArea`, and `TiePointList` (in this manuscript we use teletype font to reference classes and methods in code). `SingleScan` objects store the point cloud data for that `SingleScan` in the Scanner's Own Coordinate System (as a `vtkPolyData` object: `SingleScan.polydata_raw`) and the rigid transformation to transform that point cloud into the desired reference frame (as a `vtkTransform` object: `SingleScan.transform`). Separating the spatial information for the point cloud of an entire `Project` into `SingleScans` enables us to adjust the spatial relationships between points from different `SingleScans` (as we do below) without altering the spatial relationships within each `SingleScan`. `SingleScan` also contains methods for filtering the point cloud data (e.g., `FlakeOut`³¹) and reading and writing the data. When we filter TLS points, we set flags in the 'Classification' data field (following LAS 1.4 conventions: <https://github.com/ASPRSorg/LAS>), rather than deleting points. A `Project` object contains the set of `SingleScan` objects for this `Project` (as a dictionary: `Project.scan_dict`) and an object representing the reflector positions (an instance of `TiePointList`). `Project` also contains methods for visualizing the data (e.g., `Project.display_project`), writing data output (e.g., `Project.write_las_pdal`), and converting the point cloud into a surface representation (e.g., `Project.point_to_grid_average_image`). Finally, a `ScanArea` object contains a set of `Project` objects for this Scan Area (as a dictionary: `ScanArea.project_dict`) and methods for aligning the `SingleScans` within those `Projects` (e.g., `ScanArea.z_tilt_alignment_ss`, see below).

Filtering of TLS data. Wind-blown snow particles were filtered using `FlakeOut`³¹ with the standard parameters (`z_max=3`, `radial_precision=0.005`, `z_std_mult=3.5`, and `leafsize=100`) and assigned the classification flag '65' (the LAS 1.4 standard prescribes that user-defined classifications be greater than 63). Additionally, we manually filtered the logistics area (classification flag '73') from the Snow2 and ROV data because vehicle traffic there substantially disturbed the snow surface. For TLS data that has been processed by pydar, we store the processed data in a subdirectory of the `Project` directory named 'npfiles_archive'. Within the 'npfiles_archive' subdirectories there are subdirectories for each `SingleScan` (e.g., 'ScanPos001'). These subdirectories contain numpy²⁷ files for the point locations in the Scanner's Own Coordinate System ('Points.npy') and each data attribute (e.g., 'Reflectance.npy', 'Classification.npy', etc...). We decided to store the data in this manner for three reasons. First, '.npy' is a space-efficient, open-source format (<https://numpy.org/doc/stable/reference/generated/numpy.lib.format.html>) which is easy to open with widely-available tools. Second, separating the data attributes allows the user to only load the attributes they need into memory, which is useful because the '.las' file for a `SingleScan` is approximately 600 MB. Third, reading '.npy' files into memory in Python²⁶ is considerably faster than reading '.las', which speeds up the overall workflow.

Alignment of TLS data. For our purposes of observing snow accumulation and redistribution, the greatest source of error is bias due to misalignment of `SingleScans` from measurements on different dates. The stochastic errors due to measurement uncertainty within a `SingleScan` are insignificant¹⁵. We developed a three step process to align `SingleScans` from one `Project` (`Project_1`) to another `Project` (`Project_0`):

1. Coarsely align `Project_1` to `Project_0` by minimizing the least-squares error between their reflector positions.
2. Align each `SingleScan` in `Project_1` to the nearest `SingleScan` in `Project_0` using local maxima as keypoints, which reduces tilt biases (a tilt bias of 0.0001 radian creates a vertical error of 0.01 m at 100 m distance from the scanner).
3. Perform a fine-scale vertical alignment for each `SingleScan` in `Project_1` by minimizing modal vertical differences between it and `Project_0`.

For reflector alignment, we labeled each reflector with a consistent name (e.g., 'r01') in RiSCAN. Ice deformation and errors in the VZ1000's reflector search process can shift a reflector relative to the other reflectors. We manually compared the pairwise distances between reflectors and used only the set of reflectors whose pairwise distances between `Projects` changed by 0.02 m or less. This aligned the scans horizontally to within 0.02 m. Typically, this set comprised 4 to 8 reflectors. With this set of reflectors as keypoints, we computed the rigid transformation that minimized the least-squares error between the keypoints³² (implemented in `TiePointList.calc_transformation`). The default version of this rigid transformation calculation (`mode='LS'` in `TiePointList.calc_transformation`) requires at least 3 pairs of keypoints and has six degrees of freedom: translations in the 3 spatial directions and rotations around the 3 unit vectors (roll, pitch, and yaw). However, sometimes when there were just 3 or 4 pairs of keypoints, small vertical errors in the reflector positions produced unrealistic tilts, assessed manually by looking at the vertical differences between the aligned `Projects`. For these cases and when there were only 2 pairs of keypoints, we calculated the rigid transformation without permitting tilt changes (`mode='Yaw'` in `TiePointList.calc_transformation`). Finally, in two cases ice dynamics caused there to be no reflectors whose pairwise distance changed by less than 0.02 m. In these cases, we used a single reflector to determine the translational components of the rigid transformation and manually adjusted the yaw component such that flag posts frozen into the ice (see Fig. 1 for example) aligned to within 0.02 m at their bases. Manual inspection of the results indicated that reflector alignment brings vertical biases within 0.05 m and tilt biases within 0.001 radian.

Local maxima alignment for each `SingleScan` followed the same process as reflector alignment, except that it used local maxima as keypoints instead of reflectors. Local maxima are mostly the crests of ridges, hummocks, or human installations (e.g., poles) and are unlikely to erode or accumulate snow. A pair of local maxima from

the SingleScans was required to be within a set spatial tolerance of each other to be used as keypoints. We defined the tolerance in cylindrical coordinates centered on the scanner's location of the SingleScan that was being aligned. We used a tolerance of 0.0008 radian yaw, 0.001 radian tilt, and 0.1 m radial difference. Local maxima were located on 5×5 m regions (changing the region size to 2×2 m or 10×10 m had no discernable impact on the results). These settings produced several hundred keypoints for each pair of SingleScans from the different Projects. The tilt biases after local maxima alignment were less than 0.0001 radian (i.e., a 1 cm vertical offset across a 100 m distance), determined by manual inspection. Local maxima alignment is implemented in `ScanArea.max_alignment_ss`.

Finally, for the fine-scale vertical alignment, we exploited the fact that numerous field observations at MOSAiC suggested that a plurality of the snow surface did not change on weekly, or even monthly, time-frames. These observations included: snowmobile tracks did not typically get covered by snow except for isolated, distinct snow drifts. The indentations made by the feet of the TLS tripod were often visible when we revisited scan position locations. The circular mark left in the snow by the atmospheric flux chamber measurement could be identified months after the measurement was made. Cm-scale micro-relief on the snow surface observed near the TLS scanner (where point density is very high) appears consistent between scans, unless a snow dune happened to form in that location. Certain distinctive snow features (e.g., barchan dunes) remained unchanged for months. If the plurality of the snow surface does not change between Projects, then the modal vertical difference must be zero. For each SingleScan, we computed the distribution of vertical differences between it and the Project we were aligning it to. We used a raster with 1 m horizontal resolution created by averaging the z-components of the TLS points within each grid cell¹⁶. We used only grid cells with at least 25 points per square meter in the SingleScan and the Project it was being aligned to. The vertical component of the transformation for the SingleScan was set such that the modal difference is zero. Fine-scale vertical alignment is implemented in `ScanArea.z_alignment`. Manual inspection of the results indicated that vertical biases were reduced to within about 0.01 m (see below for Technical Validation).

In Snow1, deformation occurred frequently and throughout the Scan Area, such that there was no core ice floe with multiple reflectors on it that did not experience ice deformation for an extended period of time (as there was for Snow2 and ROV). Alignment steps 2 and 3 are predicated on the ice floe itself remaining the same, and hence could not be applied in this dynamic environment. We include Snow1 for completeness and for use by future researchers (e.g., these ice dynamics would not prevent another researcher from using the Snow1 data to compute aerodynamic drag coefficients). However, we did not validate its vertical alignment to cm-scale accuracy nor do we recommend that the data be used for snow accumulation without further work quantifying ice deformation (which is beyond the scope of this data processing).

For convenience and to avoid needing to recompute the transformations, the rigid transformation aligning each SingleScan has been written out to a '.npz' file that can be loaded directly in pydar. In the Project's directory, there is a subdirectory named 'transforms', within which there are subdirectories for each SingleScan (e.g., 'ScanPos001'). The transformation is in this subdirectory and is named 'current_transform.npz'. Finally, because the ROV and Snow2 Scan Areas were connected, we decided to place them within the same lagrangian, ice-fixed reference frame (for which the origin happens to be near the Remotely Operated Vehicle tent²).

Surface Reconstruction from Point Clouds. Many applications of topographical data—including measurement of snow accumulation—require surfaces or gridded data rather than point clouds (the format of TLS data). To produce gridded surfaces, we used gaussian process regression³³. Also known as kriging, gaussian process regression is an interpolation technique that provides the best linear unbiased estimate (minimizes least-squares error) of a parameter (e.g., surface height) at unsampled locations (e.g., a regular grid of points) given nearby measurements (e.g., TLS points) and the covariance function (also referred to as the kernel)^{33,34}. It has previously been applied to TLS data on Arctic sea ice³⁵. The vertical uncertainty in an individual TLS data point increases with distance from the scanner due to the divergence of the laser beam (0.3 mrad for the VZ1000) and is represented as gaussian noise¹⁵. Because the TLS collects useful data up to 200 m from the scanner, the vertical uncertainties in individual data points varies by an order of magnitude. One advantage of gaussian process regression, is that it can factor in the vertical uncertainty of each point when interpolating.

We chose to use an exponential covariance function because it is the simplest covariance function that can represent continuous, non-differentiable (i.e., not-smooth) surfaces³⁴. We chose this because wind-driven spatially variable snow deposition and erosion produce rough snow bedforms on horizontal scales of 10 cm to several meters³⁶. The exponential covariance function contains two hyperparameters: the 'range', defined as the distance at which the correlation between two points is less than 5%; and the 'sill', defined as the variance between two uncorrelated points (i.e., points further apart than the range). We can estimate appropriate values of the hyperparameters from the data itself, by optimizing the marginal likelihood of the gaussian process³⁴. On the scale of our scan areas (several hundred meters across) the snow and ice topography varies from rough areas of pressure ridges and rubble to smooth areas on level ice. To account for this spatial variability, we divided the domain into a grid of non-overlapping 1.2 m by 1.2 m subdomains. For each subdomain we estimated the sill from the variance of TLS points within a 5 m radius of the center of the subdomain and the range via marginal likelihood optimization using GPyTorch³⁷ with a KeOps³⁸ kernel on an NVIDIA Quaddro P2000 GPU. The size of the subdomains was chosen such that they were significantly smaller than the spatial scales of the pressure ridges (which were at least 10 s of meters) and balancing GPU memory limitations with computational time. After estimating the hyperparameters, we interpolated the gaussian processes on a regular grid with 10 cm spacing. This surface reconstruction process is implemented in `Project.merged_points_to_image`.

Data Records

Repeat TLS data in Table 1 are available at the Arctic Data Center²¹. The top level of the archive contains a directory for each Scan Area: Snow1, ROV (which includes the Snow2 projects since they are in the same reference frame), RS (April Remote Sensing site), and MET (single Project focused on Met City on 4 May). Within these directories is a subdirectory for each Project that contains all data records for that Project (as described in the Methods section). An illustrative directory tree for a Scan Area is shown below.

```
[Scan Area Name]
├── [Project Name 0]
│   ├── lasfiles
│   │   ├── ScanPos001 - SINGLESCANS - ... .las
│   │   ├── ScanPos002 - SINGLESCANS - ... .las
│   │   └── ...
│   ├── npyfiles_archive
│   │   ├── ScanPos001
│   │   │   ├── Points.npy
│   │   │   ├── Classification.npy
│   │   │   ├── Reflectance.npy
│   │   │   └── ...
│   │   ├── ScanPos002
│   │   │   ├── Points.npy
│   │   │   ├── Classification.npy
│   │   │   ├── Reflectance.npy
│   │   │   └── ...
│   │   └── ...
│   ├── transforms
│   │   ├── ScanPos001
│   │   │   └── current_transform.npy
│   │   ├── ScanPos002
│   │   │   └── current_transform.npy
│   ├── tiepoints.csv
│   ├── ScanPos001.DAT
│   ├── ScanPos002.DAT
│   └── ...
├── [Project Name 1]
│   └── ...
├── [Project Name 2]
│   └── ...
└── ...
```

Technical Validation

Alignment Validation. We qualitatively assessed the alignment results by examining the patterns of snow accumulation and redistribution in comparison with the locations of the scan positions. In general, after the full alignment process we did not find artifacts due to either the distance from the nearest scan position or on regions observed from different scan positions. In contrast such artifacts were readily apparent when conducting only reflector alignment, indicating that local maxima and fine-scale vertical alignment steps improved the alignment for observing snow processes. To quantitatively assess the uncertainty in our alignment procedure, we used a Bayesian statistical model to compare changes in snow thickness measured by TLS with manual measurements of snow thickness changes at snow thickness stakes in the Ridge Ranch mass balance site³⁹, while accounting for the uncertainties in the individual measurements. The Ridge Ranch mass balance site was located on level, first-year ice within the ROV Scan Area from 19 January to 18 March (at which point ice deformation relocated Ridge Ranch outside of the TLS measurement area). Ridge Ranch included nine snow thickness stakes, arranged in a cross, with approximately eight meters between stakes. Each stake was frozen into the ice and had a metric length scale marked on its side. Snow thickness was measured by manually observing the location of the snow surface on this metric scale with an accuracy of 0.01 m. Changes in snow thickness at each stake can be determined by comparing repeat measurements. Because the stake was permanently frozen into the ice, these measurements directly recorded changes in the snow surface height. They are unaffected by the changes in snow or snow-ice interface properties that may change the penetration depth of a snow thickness probe⁴⁰. Manual measurements at Ridge Ranch were made on 28 January, 5 February, and 7 March. ROV TLS measurements were made on 25 January, 4 February, and 11 March (Table 1). Observers in the field did not observe snow accumulation or drifting snow at Ridge Ranch during 25–28 January, 4–5 February, and 7–11 March. From these three TLS measurements and

Evaluation Period	First TLS Measurement	Second TLS Measurement	First Manual Measurement	Second Manual Measurement
1	2020-01-25	2020-02-04	2020-01-28	2020-02-05
2	2020-02-04	2020-03-11	2020-02-05	2020-03-07

Table 3. Evaluation periods for validating TLS alignment.

three manual stakes measurements, we defined two evaluation periods (Table 3) in which the TLS and the manual measurements should observe the same change in the surface at each stake, if there were no measurement noise or bias due scan misalignment.

We define $S_{k,i}$ as the change in the snow surface observed by manual stake measurements at stake i over an evaluation period k . Mathematically, $S_{k,i}$ is the true change in the snow surface $r_{k,i}$ plus noise due to measurement error $n_{k,i}$ (Eq. 1). We consider the measurement accuracy of an individual stake reading (0.01 m) to represent two standard deviations of the measurement noise. This implies that the standard deviation of the measurement noise for a single stake reading is 0.005 m. Thus, we represent $n_{k,i}$ as an instance of a zero-mean, normally distributed random variable with a variance $\sigma_s^2 = 2 \cdot (0.005\text{m})^2$. Note that each $S_{k,i}$ is the difference of two independent measurements, hence the multiplication by two in the variance. We define $t_{k,i}$ as the change in the snow surface observed by TLS for each stake and evaluation period. $t_{k,i}$ is the true change $r_{k,i}$ plus measurement noise $m_{k,i}$ minus a constant bias for the evaluation period b_k due to scan misalignment (Eq. 3). To quantify the change observed in the TLS data at a stake, we looked at the mean vertical distance for all exclusive pairs of horizontally closest points within 10 cm of the stake (excluding the stake itself) in the scans at the beginning and end of the evaluation period. The measurement noise $m_{k,i}$ is represented as an instance of zero-mean, normally distributed random variable with a variance ($\sigma_{t,k,i}^2$) determined by the vertical uncertainty in each TLS point due to the laser beam spreading with distance¹⁵.

$$s_{k,i} = r_{k,i} + n_{k,i} \quad (1)$$

$$n_{k,i} \sim N(0, \sigma_s^2) \quad (2)$$

$$t_{k,i} = r_{k,i} + m_{k,i} - b_k \quad (3)$$

$$m_{k,i} \sim N(0, \sigma_{t,k,i}^2) \quad (4)$$

To quantify how misaligned our TLS measurements may be, we computed the posterior distribution of the bias, b_k , given our measurements at each stake. We define $y_{k,i}$ as the difference between the stake measurement $s_{k,i}$ and the TLS measurement $t_{k,i}$ for each stake and evaluation period (Eq. 5). $y_{k,i}$ is equal to the bias for the evaluation period b_k plus the difference in the measurement noise for each measurement, which we denote by $g_{k,i}$. The measurement noise for the TLS and stake measurements are independent and normally distributed, allowing us to represent their difference as a zero-mean, normally distributed random variable whose variance is the sum of the variance of each measurement noise. Thus, each difference $y_{k,i}$ is an instance of a normally distributed random variable given by:

$$y_{k,i} = s_{k,i} - t_{k,i} = b_k + g_{k,i} \sim N(b_k, \sigma_{g,k,i}^2) \quad (5)$$

$$g_{k,i} = n_{k,i} - m_{k,i} \sim N(0, \sigma_{g,k,i}^2) \quad (6)$$

$$\sigma_{g,k,i}^2 = \sigma_s^2 + \sigma_{t,k,i}^2 \quad (7)$$

Applying Bayes Rule, the posterior probability density of b_k given the data for an evaluation period $p(b_k|y_k)$ is proportional to the prior probability density $p(b_k)$ multiplied by the likelihood $p(y_k|b_k)$ (Eq. 8). y_k denotes the set of all observations for evaluation period k : $y_k = \{y_{k,1}, y_{k,2}, \dots, y_{k,N}\}$. Equation 5 indicates that the likelihood of a single observation is gaussian. Conditioned on the bias, the observations are independent. Hence the likelihood of all observations is their product (Eq. 9). Finally, we chose to represent our prior for the bias as a normal distribution with a mean of zero (we do not expect there to be any bias) and a variance of $\tau_0^2 = (0.02\text{m})^2$. This variance was chosen because a bias of greater than 0.04 m (twice the standard deviation of our prior) would be obvious on manual inspection of the data, and was not observed. Moreover, the variance of the uncertainty in the observations is lower than the variance of this prior. So increasing the variance of the prior has little impact on the results (i.e., the information content of the observations is considerably higher than this prior). This yields the following expression for the posterior probability density:

$$p(b_k|y_k) \propto p(b_k)p(y_k|b_k) \quad (8)$$

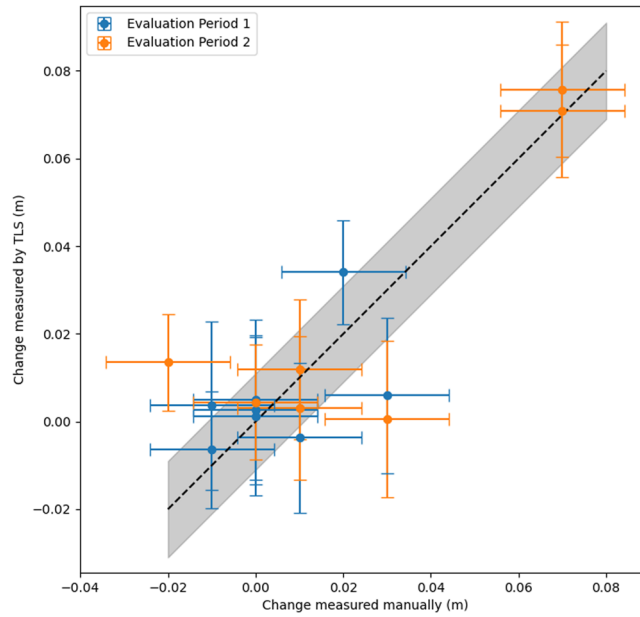


Fig. 3 Comparison of snow surface changes measured manually at snow thickness stakes with the snow surface changes at the same stakes measured by TLS during two evaluation periods (Table 3). Error bars show the 95% confidence intervals of the measurement uncertainty for the manual measurement (Eq. 2) and the TLS measurement (Eq. 4, note that each TLS measurement has an individual uncertainty which depends on the distance of the stake from the scanner and the number of TLS points within 10 cm of the stake). The black, dashed line is the 1:1 line (on which the manual measurement is equal to the TLS measurement). The grey shading above and below line represents the ± 0.011 m uncertainty in our posterior estimate of the bias. No bias between the manual and TLS measurements is visually apparent.

Evaluation Period	Expectation of Posterior Density of Bias	95% Credible Interval
1	-0.001 m	[-0.009 m, 0.007 m]
2	-0.004 m	[-0.011 m, 0.004 m]

Table 4. Scan misalignment bias.

$$\propto p(b_k) \prod_{i=1}^N p(b_k|y_k) \tag{9}$$

$$\propto \exp\left(-\frac{b_k^2}{2\tau_0^2}\right) \prod_{i=1}^N \exp\left(-\frac{1}{\sigma_{g,k,i}^2}(y_{k,i} - b_k)^2\right) \tag{10}$$

Algebraic simplification of Eq. 10 yields that the posterior density $p(b_k|y_k)$ is a gaussian (Eq. 11) whose mean is the mean of the prior and the observations, each weighted by the inverse of their variance (Eq. 12). The variance of $p(b_k|y_k)$ is determined from the sum of the inverse variances of the prior and the observations (Eq. 13).

$$p(b_k|y_k) = N(\mu_k, \tau_k^2) \tag{11}$$

$$\mu_k = \tau_k^2 \sum_{i=1}^N \frac{y_{k,i}}{\sigma_{g,k,i}^2} \tag{12}$$

$$\frac{1}{\tau_k^2} = \frac{1}{\tau_0^2} + \sum_{i=1}^N \frac{1}{\sigma_{g,k,i}^2} \tag{13}$$

We use the posterior density of the bias to establish a 95% credible interval (the interval between the 2.5th and 97.5th percentiles in the distribution of a parameter) for the bias in our alignment procedure.

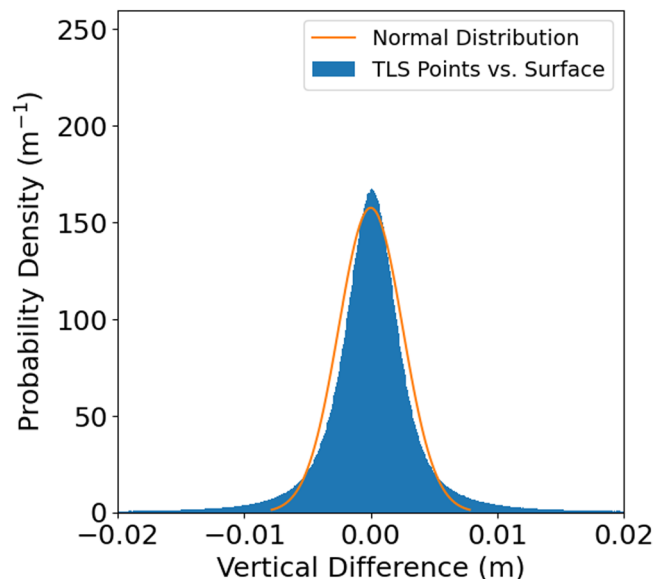


Fig. 4 Histogram of vertical differences between TLS data points and reconstructed surface for a 250 m x 65 m region of sea ice containing a large ridge and level ice measured on 26 March 2020. The overlain curve shows a normal distribution with the same median absolute deviation as the data. The 2.5th percentile to the 97.5th percentile of the differences roughly matches a normal distribution but the tails are more extreme.

To validate the TLS alignment, we compared the snow surface change observed by TLS and manual measurements (Fig. 3) at the Ridge Ranch mass balance site⁴¹ during evaluation periods 1 and 2 (Table 3). Each evaluation period included at least one snow accumulation and redistribution event. The evaluation periods collectively span 25 January to 11 March. Manual measurements indicated that most stakes experienced little to no change in the snow surface in either evaluation period. The largest snow accumulation observed manually at any of the stakes occurred during evaluation period 2, when 0.07 m of snow accumulated at two stakes (Fig. 3). The TLS measurements, too, observed approximately 0.07 m of change at the same two stakes and little to no change at the others (Fig. 3). With these data, we compute the posterior density of the bias (Table 4) following Eqs. 1–12. The estimated mean biases for evaluation period 1 and 2 are -0.001 m and -0.004 m, respectively. And the minimum and maximum 95% credible interval bounds are -0.011 m and 0.007 m respectively. Thus, we conclude that the bias due to scan misalignment is less than 0.011 m. We stress that these numerical values are particular to the ROV and Snow2 scan areas at MOSAiC (regular ice deformation in the Snow1 area will require future work to correct for). Future TLS measurement campaigns should conduct *in-situ* validation measurements for their specific measurement sites.

Surface Reconstruction Validation. We validated our gaussian process regression approach to surface reconstruction by examining the differences between the vertical components of TLS data points and the reconstructed surface. Figure 4 shows an example of the distribution of these differences for a 250 m x 65 m region reconstructed on a 10 cm grid containing a large ridge and level ice. The mean difference is 2.5×10^{-5} m, the median difference is 6.5×10^{-5} m, the standard deviation of the differences is 0.0054 m and the median absolute deviation, a metric of the variability of data that is robust to extreme values⁴², for these data is 0.0017 m. This is similar in magnitude to the median standard deviation of the vertical uncertainty due to laser beam divergence: 0.0026 m. These results combined with manual inspection of the differences suggest that almost all of the differences between the TLS points and the reconstructed surface can be attributed to the divergence of the laser beam with distance from the scanner. A small fraction (less than 5%) of the differences are due to areas with high surface roughness on horizontal length scales of less than the 10 cm grid spacing. These rough areas cause the tails of the distribution of differences to be more extreme than a normal distribution (Fig. 4). For the purpose of assessing snow accumulation, the uncertainties in our surface reconstruction approach are insignificant compared to the alignment uncertainties (0.011 m, see above). However, we caution that applications involving surface roughness may want to further develop surface reconstruction techniques.

Usage Notes

On 18 March, the ice was deforming within the ROV Scan Area while we were collecting TLS data. On 8 April, Snow1 split in two while we were collecting TLS data. Data collected after the deformation began are in a second project ('mosaic_01_080420.RiSCAN'). We recommend caution when using data collected during deformation events. The convention in project names was inadvertently switched from 'MMDDYY' to 'DDMMYY' at the turn of the year (except for 'mosaic_01b_061219.RiSCAN.RiSCAN.RiSCAN' on 6 December). A function is provided in pydar to convert a Project's name to its date (`pydar.mosaic_date_parser`). Sometimes extra 'RiSCAN's were included in the Project name, these have no significance.

Researchers interested in extending the functionality of pydar (e.g., adding feature tracking functionality for ice deformation, point cloud segmentation, etc) are encouraged to contact the corresponding author in case related efforts are underway. We also welcome discussions on potential uses of these data and collaborations with other data products.

Code availability

pydar is available at Zenodo¹⁹ (<https://zenodo.org/record/8120858>).

Received: 13 July 2023; Accepted: 27 December 2023;

Published online: 13 January 2024

References

- Knust, R. Polar Research and Supply Vessel POLARSTERN Operated by the Alfred-Wegener-Institute. *Journal of large-scale research facilities JLSRF* **3**, A119, <https://doi.org/10.17815/jlsrf-3-163> (2017).
- Nicolaus, M. *et al.* Overview of the MOSAiC expedition: Snow and sea ice. *Elementa: Science of the Anthropocene* **10**, 000046, <https://doi.org/10.1525/elementa.2021.000046> (2022).
- Shupe, M. D. *et al.* Overview of the MOSAiC expedition: Atmosphere. *Elementa: Science of the Anthropocene* **10**, 00060, <https://doi.org/10.1525/elementa.2021.00060> (2022).
- Rabe, B. *et al.* Overview of the MOSAiC expedition: Physical oceanography. *Elementa: Science of the Anthropocene* **10**, 00062, <https://doi.org/10.1525/elementa.2021.00062> (2022).
- Stroeve, J. C. *et al.* The Arctic's rapidly shrinking sea ice cover: a research synthesis. *Climatic Change* **110**, 1005–1027, <https://doi.org/10.1007/s10584-011-0101-1> (2012).
- Nandan, V. *et al.* Wind redistribution of snow impacts the Ka- and Ku-band radar signatures of Arctic sea ice. *The Cryosphere* **17**, 2211–2229, <https://doi.org/10.5194/tc-17-2211-2023> (2023).
- Sturm, M. & Massom, R. Chapter 3: Snow in the sea ice system: friend or foe? In Thomas, D. N. (ed.) *Sea Ice* (John Wiley & Sons, Hoboken, NJ, 2017).
- Hunke, E. *et al.* CICE, The Los Alamos Sea Ice Model. Tech. Rep. CICE; 005315WKSTN00, Los Alamos National Lab. (LANL), Los Alamos, NM (United States) (2017).
- Urrego-Blanco, J. R., Urban, N. M., Hunke, E. C., Turner, A. K. & Jeffery, N. Uncertainty quantification and global sensitivity analysis of the Los Alamos sea ice model. *Journal of Geophysical Research: Oceans* **121**, 2709–2732, <https://doi.org/10.1002/2015JC011558> (2016).
- Sturm, M., Perovich, D. K. & Holmgren, J. Thermal conductivity and heat transfer through the snow on the ice of the Beaufort Sea. *Journal of Geophysical Research: Oceans* **107**, 19–17, <https://doi.org/10.1029/2000JC000409> (2002).
- Abraham, C., Steiner, N., Monahan, A. & Michel, C. Effects of subgrid-scale snow thickness variability on radiative transfer in sea ice. *Journal of Geophysical Research: Oceans* **120**, 5597–5614, <https://doi.org/10.1002/2015JC010741> (2015).
- Wagner, D. N. *et al.* Snowfall and snow accumulation during the MOSAiC winter and spring seasons. *The Cryosphere* **16**, 2373–2402, <https://doi.org/10.5194/tc-16-2373-2022> (2022).
- Prokop, A., Schirmer, M., Rub, M., Lehning, M. & Stocker, M. A comparison of measurement methods: terrestrial laser scanning, tachymetry and snow probing for the determination of the spatial snow-depth distribution on slopes. *Annals of Glaciology* **49**, 210–216, <https://doi.org/10.3189/172756408787814726> (2008).
- Deems, J. S., Painter, T. H. & Finnegan, D. C. Lidar measurement of snow depth: a review. *Journal of Glaciology* **59**, 467–479, <https://doi.org/10.3189/2013JoG12J154> (2013).
- Hartzell, P. J., Gadomski, P. J., Glennie, C. L., Finnegan, D. C. & Deems, J. S. Rigorous error propagation for terrestrial laser scanning with application to snow volume uncertainty. *Journal of Glaciology* **61**, 1147–1158, <https://doi.org/10.3189/2015JoG15J031> (2015).
- Trujillo, E., Leonard, K., Maksym, T. & Lehning, M. Changes in snow distribution and surface topography following a snowstorm on Antarctic sea ice. *Journal of Geophysical Research: Earth Surface* **121**, 2172–2191, <https://doi.org/10.1002/2016JF003893> (2016).
- Polashenski, C., Wright, N. & Liston, G. Snow Wind and Time Project: Snow Depth Data Utqiagvik, AK, November 2016 - June 2017. *Arctic Data Center*, <https://doi.org/10.18739/A2C53F17N> (2019).
- Polashenski, C., Perovich, D. & Courville, Z. The mechanisms of sea ice melt pond formation and evolution. *Journal of Geophysical Research: Oceans* **117**, <https://doi.org/10.1029/2011JC007231> (2012).
- Clemens-Sewall, D. [davidclemenssewall/pydar: v1.0.1, Zenodo https://doi.org/10.5281/ZENODO.8120858](https://doi.org/10.5281/ZENODO.8120858) (2023).
- Raphael, I. *et al.* Measurements of sea ice point-mass-balance using hotwire thickness gauges and ablation stakes during the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) Expedition in the Central Arctic (2019–2020). *Arctic Data Center*, <https://doi.org/10.18739/A2NK36626> (2022).
- Clemens-Sewall, D. *et al.* High-Resolution Repeat Topography of Drifting Ice Floes in the Arctic Ocean from Terrestrial Laser Scanning Collected on the Multidisciplinary drifting Observatory for the Study of Arctic Climate Expedition. *Arctic Data Center* <https://doi.org/10.18739/A20863712> (2022).
- Kruppen, T. *et al.* The MOSAiC ice floe: sediment-laden survivor from the Siberian shelf. *The Cryosphere* **14**, 2173–2187, <https://doi.org/10.5194/tc-14-2173-2020> (2020).
- Itkin, P. *et al.* Sea ice and snow characteristics from year-long transects at the MOSAiC Central Observatory. *Elementa: Science of the Anthropocene* **11**, 00048, <https://doi.org/10.1525/elementa.2022.00048> (2023).
- Light, B. *et al.* Arctic sea ice albedo: Spectral composition, spatial heterogeneity, and temporal evolution observed during the MOSAiC drift. *Elementa: Science of the Anthropocene* **10**, 000103, <https://doi.org/10.1525/elementa.2021.000103> (2022).
- Möbius, A. F. Der barycentrische Calcul, ein Hülfsmittel zur analytischen Behandlung der Geometrie (etc.) (Barth, 1827).
- Rossum, G. V. & Drake, F. L. *The Python language reference*. No. Pt. 2 in Python documentation manual / Guido van Rossum; Fred L. Drake [ed.] (Python Software Foundation, Hampton, NH, 2010), release 3.0.1 [repr.] edn.
- Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362, <https://doi.org/10.1038/s41586-020-2649-2> (2020).
- Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261–272, <https://doi.org/10.1038/s41592-019-0686-2> (2020).
- Schroeder, W., Martin, K. & Lorensen, B. *The Visualization Toolkit* (Kitware, 2006).
- Behnel, S. *et al.* Cython: The best of both worlds. *Computing in Science & Engineering* **13**, 31–39 (2010). Publisher: IEEE.
- Clemens-Sewall, D., Parno, M., Perovich, D., Polashenski, C. & Raphael, I. A. FlakeOut: A geometric approach to remove wind-blown snow from terrestrial laser scans. *Cold Regions Science and Technology* **201**, 103611, <https://doi.org/10.1016/j.coldregions.2022.103611> (2022).
- Arun, K. S., Huang, T. S. & Blostein, S. D. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-9**, 698–700, <https://doi.org/10.1109/TPAMI.1987.4767965>. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence (1987).
- Matheron, G. Principles of geostatistics. *Economic Geology* **58**, 1246–1266, <https://doi.org/10.2113/gsecongeo.58.8.1246> (1963).
- Rasmussen, C. E. & Williams, C. K. *Gaussian Processes for Machine Learning* (MIT Press, 2006).

35. Landy, J., Ehn, J., Shields, M. & Barber, D. Surface and melt pond evolution on landfast first-year sea ice in the Canadian Arctic Archipelago. *Journal of Geophysical Research: Oceans* **119**, 3054–3075, <https://doi.org/10.1002/2013JC009617> (2014).
36. Filhol, S. & Sturm, M. Snow bedforms: A review, new data, and a formation model. *Journal of Geophysical Research: Earth Surface* **120**, 1645–1669, <https://doi.org/10.1002/2015JF003529> (2015).
37. Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q. & Wilson, A. G. GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration. *arXiv:1809.11165 [cs, stat]* (2021).
38. Charlier, B., Feydy, J., Glaunès, J. A., Collin, F.-D. & Durif, G. Kernel Operations on the GPU, with Autodiff, without Memory Overflows. *Journal of Machine Learning Research* **22**, 1–6 (2021).
39. Raphael, I., Perovich, D. K., Polashenski, C. M., Clemens-Sewall, D. & Planck, C. An autonomous sea ice mass balance observation network. In *AGU Fall Meeting 2021* (AGU, 2021).
40. Sturm, M. & Holmgren, J. An Automatic Snow Depth Probe for Field Validation Campaigns. *Water Resources Research* **54**, 9695–9701, <https://doi.org/10.1029/2018WR023559> (2018).
41. Raphael, I. *et al.* Manual point-measurements of sea ice mass balance during the MOSAiC Expedition. In *EGU General Assembly Conference Abstracts*, EGU21–3757 (2021).
42. Rousseeuw, P. J. & Croux, C. Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association* **88**, 1273–1283, <https://doi.org/10.1080/01621459.1993.10476408> (1993).
43. Nixdorf, U. *et al.* MOSAiC Extended Acknowledgement. *Zenodo* <https://doi.org/10.5281/zenodo.5541624> (2021).

Acknowledgements

Data used in this manuscript were produced as part of the international MOSAiC project with the tag MOSAiC20192020 and the Project_ID: AWI_PS122_00. We thank all people involved in the expedition of the research vessel Polarstern¹ during MOSAiC in 2019–2020 as listed in the MOSAiC Extended Acknowledgement⁴³. Thank you to Jon Holmgren for fabricating the heater cover for the TLS. Thank you to Eric Brossier, Steven Fons, Jesper Hansen, Thomas Olufson, Martin Radenz, Carl Martin Schonning, Saga Svavarsdóttir and Monica Votvik for their capable assistance in the field. D.C.S. was supported by NSF OPP-1724540, NSF OPP-1724424 and NSF OPP-2138788. I.A.R., and D.P. were supported by NSF OPP-1724540, NSF OPP-1724424 and NSF OPP-2138785. C.P. was supported by NSF OPP-1724540 and NSF OPP-2138785. MO was supported by NSF OPP-1735862. A.R.M. was supported by WSL_201812N1678 and SPI DIRCR-2018-003. I.S.O.M. supported by BMBF Grant 03F0810A, jointly funded by BMBF (DE) and NERC (UK). D.N.W. was supported by SNSF-200020_179130 and SPI DIRCR-2018-003. M.J. was supported by EU-Arice DEARice. P.I. was supported by NSF #1820927 and RCN #287871.

Author contributions

The concept of TLS snow redistribution and accumulation measurements on MOSAiC was conceived, facilitated, and supervised by C.P. D.C.-S. led the data collection from March to May 2020, developed the data processing methods, and wrote the initial draft. I.A.R. led the data collection from October 2019 to February 2020. M.P. provided guidance and advice on the technical validation and surface reconstruction. C.P. designed the heater cover for the TLS. D.P. supervised this project. P.I., M.J., A.J., A.R.M., I.M., M.O., D.W., and R.V. contributed to data acquisition. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.C.-S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024