DIGITAL TOOLS FOR ANALOG DATA: RECONSTRUCTING THE FIRST ICE-PENETRATING RADAR SURVEYS OF ANTARCTICA AND GREENLAND

Thomas O. Teisberg[1], *Dustin M. Schroeder*[1,2]

1) Department of Electrical Engineering, Stanford University,
2) Department of Geophysics, Stanford University

ABSTRACT

During the 1970s, the first large-scale ice-penetrating radar surveys were conducted over Antarctica and Greenland through a collaboration between the Scott Polar Research Institute at the University of Cambridge, the National Science Foundation, and the Technical University of Denmark [1]. This pioneering work represents the earliest available large-scale measurements of the sub-surface conditions of Earth's two major ice sheets. In many cases, data collected as part of this survey pre-dates other available data in the area by decades. As we experience increasingly rapid changes in Antarctica and Greenland, understanding the historical states of these ice sheets takes on increased importance. Although 45-55 years is a short time in the evolution of continent-scale ice sheets, this dataset is unique in that it represents direct measurements over massive spatial scales. Unfortunately, integration of this data into models has proven extremely difficult due to unique challenges in how the data was collected and stored. We introduce our efforts to build digital tools for this analog dataset to bring this data into the modern era and make it accessible to researchers.

***Index Terms*—** ice-penetrating radar, datasets, open science, data portals

1. INTRODUCTION

The difficulty of working with analog data formats has discouraged scientific usage of archival data since the availability of large amounts of modern digital data with comparatively easy access and better curation of metadata. As a result, the more than 1000 rolls of optical film containing radar data from Antarctica and Greenland collected primarily in the 1970s sat in archive, mostly unused, until it was re-scanned at higher quality and publicly released in 2019 [2] (see [3] for more details on this process).

Despite its public release, however, the difficulty of working with this data has continued to prevent widespread usage. The dataset presents some unique challenges not common to modern data: related data are stored completely separately and must be tracked down across multiple rolls of film and scanned paper documents, relevant documents include handwritten and typewritten notes which are difficult to automatically convert to digital data without manual quality control, and the data is geolocated using separately recorded data from an inertial navigation system (as the survey was done before the availability of GPS satellites).

In order to facilitate more widespread use of this valuable data, we developed a website, available at `https://www.radarfilm.studio/`, to provide easier access to the data and to facilitate our efforts to provide manual quality control of aspects of the dataset.

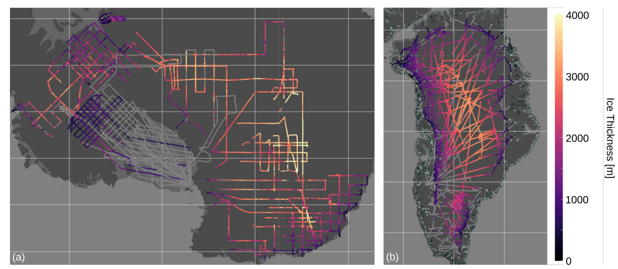

Fig. 1. Currently available historical radar data lines in (a) Antarctica and (b) Greenland. The colorbar shows picked ice thickness. Note that thickness data is from picking done contemporary with the dataset collection and has not yet been reviewed. Radar data available without ice thickness information is shown in gray.

2. DATASET DESCRIPTION

The available dataset spans about a decade of survey flights over Antarctica and Greenland by US-provided C-130 aircraft through the previously-described collaborative effort. The currently available data coverage is shown in Figure 1.

The radar was a coherent, chirped system operating with two channels centered at 60 MHz and 300 MHz [4, 1, 5]. Data from each channel was processed using an analog implementation of pulse compression. The resulting data was rastered onto 35 mm optical film using modified cathode ray tube displays. Three distinct representations of this pulse compressed data could be stored (see Figure 2 simultaneously, however in many cases only one or two were collected and/or preserved. These representations include Z-scopes, more commonly known as radargrams today, and A-scopes, which plot the power return as a function of delay time. Because Z-scope data was passed through an analog dynamic range compression circuit (similar to log compression frequency used in the display of radargrams), the uncompressed A-scopes play a key role in facilitating radiometric calibration of the data [6].

Because the surveys were carried out before the availability of GPS satellites, the positioning was calculated using an inertial navigation system and ground control points, with an estimated accuracy of 5 km [1]. Positioning information and radar data was marked with a Coded Binary Decimal (CBD) internal clock to facilitate alignment the data records.

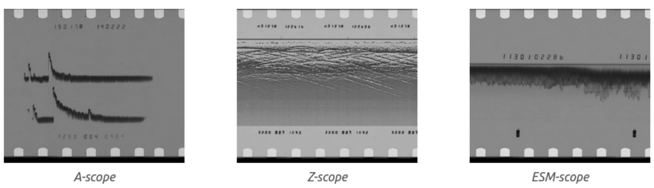

Fig. 2. Three distinct types of film-based data were recorded by the radar instrument. Each type of data recording encodes the same measurement in different ways intended for different scientific uses. Linking these multiple views of the same measurements together represents a significant challenge, however it will enable radiometric analysis as shown in [6].

Fig. 3. Metadata necessary to connect radar data to positioning information and instrument parameters was encoded as text on the film strips. This data has been extracted by optical character recognition (OCR) as part of the preprocessing of the this dataset. Red labels were added to identify components of the metadata.

3. DATA PREPROCESSING

Prior to the development of the radarfilm.studio website, three preprocessing steps were taken on the data:

Preliminary frame stitching: Raw data from the optical film scanner is in the form of single frame images. For a conventional video recording on 35 mm film, these would correspond to individual frames of the video. The recorded data, however, is continuously rastered onto the film and does not align to any particular frame markers. As such, these small slices of data are difficult to work with on their own. Small sections of individual frames from contiguous sections of film were automatically stitched together to produce short "film segments," which we treat as the base unit of data for the purposes of the website and future scientific work.

Optical character recognition (OCR) of metadata on film: Film data is periodically stamped with metadata, including the CBD, flight number, and year of data collection necessary to match radar data with separately recorded positioning information, as shown in Figure 3. This metadata was captured by OpenCV and the Tesseract OCR engine prior to the development of the website. Due to a range of artifacts in the film, such as those shown in Figure 4, and changes in the exact format of the recorded metadata across versions of the radar instrument, the OCR-produced metadata requires manual quality control.

OCR of positioning information: As previously mentioned, the positioning information was stored separately. In most cases, the most up-to-date available record is a typewritten sheet listing CBD numbers, approximate latitude and longitude, and estimates of the ice thickness. These pages have been scanned and processed with a commercial OCR tool (Amazon Textract). While simpler to work with than the film metadata, this data also requires manual review. Please note, in particular, that no efforts have yet been made to review the surface and bed picks.

Fig. 4. Two examples of complications in automatically extracting metadata. Left: Damage on the original film frequently partially obscures important metadata. Right: Handwritten annotations provide potentially valuable information but also obstruct scanning of metadata.

4. WEBSITE DEVELOPMENT

Online portals that facilitate easy discovery of ice-penetrating radar data are still relatively new to the field. Several major providers of modern IPR data have developed dedicated websites over the past few years, while other large modern

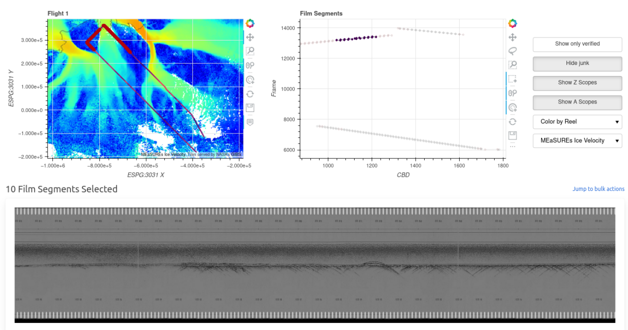

Fig. 5. An example of a page for an individual flight line, in this case a flight that extends onto the Filchner–Ronne Ice Shelf (FRIS). The top left shows a map of the flight line with a user-selectable basemap (currently showing surface velocity). The center-top plot shows available film segments, plotted across CBD number and frame number (from the original scanning). This view allows the user to separate out different types of data (i.e. A-scopes and Z-scopes) associated with the flight and quickly identify potentially mislabeled data. Beneath the plots, a preview of the selected film segments is automatically generated to create a radargram-like view. In this instance, it shows a basal channel on the FRIS that radar evidence suggests has persisted for 40 years [2].

datasets remain without any user-friendly access method. Of the existing data portals, there are a mix of mostly scratch-built solutions [7] and portals that leverage existing packages such as ArcGIS's interactive mapping tools [8]. Due to the unique requirements of this archival data set and the emphasis on support for updating metadata, we choose to build our own solution. The code for the entire website is open-source and available on GitHub at `https://github.com/radioglaciology/radarfilmstudio`.

The primary purpose of the website is to put the available data in geographic context. As such, the most important feature are the maps, shown in Figure 1, showing the location of flight lines where data exists. The base layer of these maps may be changed to show surface velocity or satellite imagery for added context. Clicking on a flight line directs users to a page showing that flight line in more detail, an example of which is shown in Figure 5.

Individual film segments interactively selected in the top plots are stitched into a preview radargram-like view for quick looks at the data. In addition, each film segment is listed below this preview for more careful inspection, with options for registered users to update any incorrect metadata or add notes, as shown in Figure 6. Updates may be made by users directly on the website or by scripts. Regardless of the source, the entire version history of each film segment is tracked to preserve the history of its metadata.

File paths permanently assocaited with each film segment uniquely link that unit of data back to permanent digital

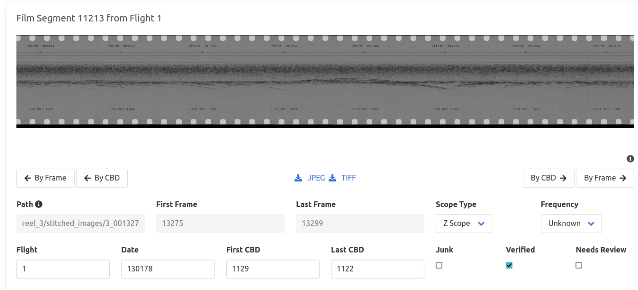

Fig. 6. "Film segments" make up the base unit of data for the website. Each selected film segment is displayed as an entry like the one shown here. Authorized users may directly update metadata through this interface.

archives stored by (and publicly available through) Stanford Libraries [3]. Work is ongoing to ensure that associated metadata is stored in an equally permanent manner.

4.1. Technical architecture and implementation

The website is open-source, along with relevant preprocessing code. The core feature is a PostgreSQL database containing entries for each film segment. For all new data, this is initially populated by the results of the OCR pre-processing. As metadata is updated, either through the website or by a custom script, these changes are saved in the database. A second database records each change, along with the user or script that made the change and a timestamp.

The site itself is powered by the open-source Flask web framework. The plots and maps are created using the Bokeh and HoloViews data visualization libraries, facilitating dynamic interactions between multiple data visualizations on the same page. These linkages are critical in places where the data relations are sometimes complex. A single CBD-linked location on the map, for example, may correspond to one or multiple film segments depending on the configuration of the radar and which sets of optical film were saved). Using our website, the user may select data from either the map or the film segments plot directly, either of which will automatically highlight the matching data from the other.

In order to place the data in full context, the user can select between different base maps for each map plot. These are provided using tiles from NASA's Global Imagery Browse Services (GIBS). All geographic data is plotted in polar projections (EPSG:3031 and EPSG:3413) to provide comparatively realistic aspect ratios, as compared to the common Web Mercator (EPSG:3857) projection. This apparently trivial implementation detail remains a significant challenge for polar-focused web maps, as many web mapping packages do not support these projections.

5. FUTURE WORK

radarfilm.studio is an ongoing project that currently supports multiple scientific projects. In order to continue supporting these projects and to support new research, there are a number of planned new efforts:

Public data access API: Currently data can be manually downloaded from the website, and, while this process can be automated, this workflow could be improved. We plan to develop a public API for accessing both the radar data and the metadata. As much as possible, we will develop this in coordinate with other data providers to allow for interoperability.

Pixel-level geoferencing: Our current metadata is tied to each film segment. This provides sufficient resolution for most qualitative work and a solid basis for manual alignment of small segments. Ideally, however, we could provide both geolocation and reflection time on a per-pixel basis in every image, at a similar level to modern data. This could facilitate radiometric analysis of the data, among other quantitative radar analysis approaches.

Finding modern data comparisons: With the increasing availability of open access to modern radar data, we plan to provide overlays on the maps showing the locations of modern data for comparison purposes. In a few instances, these overlaps have been manually identified and yielded interesting results [2].

6. ACKNOWLEDGEMENTS

This paper reports primarily on the radarfilm.studio website created to facilitate easier access to this historical dataset. As briefly described above, the scanning, reconstruction, and metadata curation processes have required the help of numerous collaborators on this project. For a full list of those involved in bringing this dataset back to life, please see `https://www.radarfilm.studio/docs/citation`

7. REFERENCES

[1] D.J. Drewry, S.R. Jordan, and E. Jankowski, "Measured Properties of the Antarctic Ice Sheet: Surface Configuration, Ice Thickness, Volume and Bedrock Characteristics," *Annals of Glaciology*, vol. 3, pp. 83–91, 1982.

[2] Dustin M. Schroeder, Julian A. Dowdeswell, Martin J. Siegert, Robert G. Bingham, Winnie Chu, Emma J. MacKie, Matthew R. Siegfried, Katherine I. Vega, John R. Emmons, and Keith Winstein, "Multidecadal observations of the Antarctic ice sheet from restored analog radar records," *Proceedings of the National Academy of Sciences*, vol. 116, no. 38, pp. 18867–18873, Sept. 2019.

[3] Stanford Libraries, "Antarctic Radar Film Digitization Project - Spotlight at Stanford," https://exhibits.stanford.edu/radarfilm.

[4] G. De Q. Robin, D.J. Drewry, and D.T. Meldrum, "International Studies of Ice Sheet and Bedrock," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 279, no. 963, pp. 185–196, 1997.

[5] Robert G. Bingham and Martin J. Siegert, "Radio-Echo Sounding Over Polar Ice Masses," *Journal of Environmental and Engineering Geophysics*, vol. 12, no. 1, pp. 47–62, Mar. 2007.

[6] Dustin M. Schroeder, Anna L. Broome, Annabel Conger, Acacia Lynch, Emma J. Mackie, and Angelo Tarzona, "Radiometric analysis of digitized Z-scope records in archival radar sounding film," *Journal of Glaciology*, vol. 68, no. 270, pp. 733–740, Aug. 2022.

[7] Kyle Purdon, *OpenPolarServer (OPS) - An Open Source Spatial Data Infrastructure for the Cryosphere Community*, Ph.D. thesis, University of Kansas, 2014.

[8] Alice C. Frémand, Julien A. Bodart, Tom A. Jordan, Fausto Ferraccioli, Carl Robinson, Hugh F. J. Corr, Helen J. Peat, Robert G. Bingham, and David G. Vaughan, "British Antarctic Survey's aerogeophysical data: releasing 25 years of airborne gravity, magnetic, and radar datasets over Antarctica," *Earth System Science Data*, vol. 14, no. 7, pp. 3379–3410, July 2022.