SNOWPYLOT: A PYTHON LIBRARY FOR WORKING WITH SNOWPILOT DATA

Mary Kate Connelly[1], Samuel Verplanck[2]

ABSTRACT

SnowPilot (snowpilot.org) is free, open-source software designed to help users graph, record, and store snowpit data (Chabot, Kahrl and Earl, 2016). The SnowPilot database currently contains data from over 65,000 snowpits, collected by snow recreationists and professionals around the world. It is particularly popular among avalanche professionals in the United States. Despite SnowPilot's widespread use to graph and record snowpit data, the SnowPilot database has been underutilized as a research resource. This is due, in part, to the complexity of the data and lack of available tools for accessing specific snowpit properties.

To address these challenges and improve accessibility to the database, we developed SnowPylot (pronounced "*snow-pea-why-let*"). SnowPylot is an open-source Python library that enables researchers to import and structure data from the SnowPilot database within Python, facilitating the use of Python tools and methods for analysis. Users can export CAAML.xml files from SnowPilot, on an individual or batch scale, and use SnowPylot to process the files in Python. (Keywords: SnowPilot, SnowPylot, CAAML, Python, Snow Pit, Data Science)

INTRODUCTION

Snow pit observations are fundamental to both avalanche and hydrologic forecasting, providing critical information about snowpack structure and stability. Since its inception, SnowPilot has served as a powerful platform for collecting and sharing these observations within the avalanche community. However, while SnowPilot has successfully amassed a vast database of snow pit observations (over 65,000), the tools for accessing and analyzing the data have remained limited.

This paper presents SnowPylot v1.1.0, an open-source Python library for accessing and analyzing snow pit data from the SnowPilot database. This library, designed to work with files exported in the Canadian Avalanche Association Markup Language (CAAML) format (Canadian Avalanche Association, n.d.) enables researchers and practitioners to process snow pit observations at both individual and large-scale levels. The information in the CAAML files is parsed into Python objects with a structure that mirrors the CAAML format. Users can leverage Python tools and libraries for analysis and make it possible to examine patterns and trends across thousands of snow pit observations, potentially revealing new insights about snowpack behavior and avalanche activity. In this paper, we demonstrate the library's capabilities through an example analysis of snow pit observations from the 2020-2024 water years, showcasing its ability to handle large datasets and extract meaningful insights.

DATA STRUCTURE AND SAMPLE DATASET

The SnowPylot Python library implements an object-oriented design that mirrors the CAAML data structure. Additional details are available in the README.md and other package documentation on GitHub and PyPI (Connelly & Verplanck, 2025a). For data analysis, the library provides methods that operate at individual and batch scales. At the individual pit level, users can examine and analyze individual layers or stability test results. For larger datasets, the library supports batch processing capabilities, allowing for multiple pit parsing, statistical aggregation, and temporal and spatial analysis. The library also supports conversion to pandas DataFrames and export to common formats like CSV, facilitating integration with other analysis tools.

For the purposes of demonstration, we exported CAAML.xml files from SnowPilot for the 2020-2024 water years and used the SnowPylot library to aggregate summary statistics about the snow pits in the dataset. A Jupyter notebook called "demo_2020-2024.ipynb" contains the full analysis and is in the "demos" folder of the SnowPylot repository on GitHub (Connelly & Verplanck, 2025b).

Paper presented Western Snow Conference 2025
[1] Montana State University, Bozeman, MT, 59718, connellymarykate@gmail.com
[2] Montana State University, Civil Engineering Department, Bozeman, MT, samuelverplank@montana.edu

Snow Pit

The SnowPit class serves as the container for all snow pit observations and consists of four main components: core_info, snow_profile, stability_tests and whump_data. Each CAAML.xml file in the dataset is parsed into a SnowPit object.

Core Info

The core_info class encapsulates core attributes of a snow pit, including the pit's identification details (pit_id, pit_name), temporal information (date), and descriptive elements (comment, caaml_version). There are additional nested objects within the core_info class including a user object that holds information about the person who created the profile (Table 1), a location object that contains geographical and environmental details of where the snow pit was dug (Table 2), and a weather_conditions object that describes the atmospheric conditions at the time of the profile creation.

Table 1. Summary of some core_info.user attributes for the 2020-2024 dataset

Total pits	31,170
Unique Users	3,854
Pits submitted by Avalanche Professionals	19,891
Pits submitted by non-professionals	11,279

Table 2. Summary of some core_info.location attributes for the 2020-2024 dataset

Unique Countries	30
Pits Near and Avalanche	945
Pit on Avalanche Crown	480
Pits on Avalanche Flank	240

Snow Profile

The snow_profile class contains basic profile information like measurement direction, profile depth, and height of the snowpack. The class also contains information about snow pit layers (stored in a list of layer objects), temperature observations (stored in a list of temp_obs objects), and density observations (stored in a list of density_obs objects). In addition, this class houses a subclass containing information about the surface conditions (surface_condition) and indicates the layer of concern if specified. Some summary statistics about the 2020-2024 dataset using snow_profile attributes are shown in Table 3.

Table 3. Summary of some snow_profile data available for 2020-2024 dataset

Total Pits	31,170
Layers	232,718
Layers with Primary Grain Form	190,805
Layers with Primary Grain Size	109,034
Pits with Temperature Profile	14,449
Temperature Observations	134,518
Pits with Density Profile*	891
Density Observations*	7,731
Pits with Foot Penetration Observation	19,811
Pits with Ski Penetration Observation	11,061

* This number only represents snow pits where a density profile was input separately from layer observations.

Stability Tests

The stability_tests class is a dataclass that serves as a container for different types of snow stability tests. It contains four lists as attributes, each representing a different type of stability test (Extended Column Test, Compression Test, Rutschblock Test and Propagation Saw Test). Some summary statistics about the 2020-2024 dataset using information from stability_tests are shown in Table 4 and Table 5.

Table 4. Summary of some stability_tests data available for 2020-2024 dataset

Total Pits	31,170
Pits with Stability Test Results	28,151
Percentage of Pits with Stability Test Results	90.31%

Table 5. Summary of some stability_tests data available for 2020-2024 dataset by test type

	ECT	CT	PST	RBlock
Total Pits with Test Results	21,092	18,022	3,102	121
Total Test results	29,247	32,150	3,719	134
Percentage of Pits with Test Results	67.67	57.82%	9.95%	0.39%

Whumpf Data

Whumpf_data is a custom, SnowPilot-specific set of fields that was created for a study and allows certain users to capture information about whumpf observations. At the time of this writing, there are 306 records in the SnowPilot database that contain whumpf_data.

EXAMPLE ANALYSIS

What is the relationship between hand hardness and primary grain form (Greene et al., 2022) of snow pit layers?

Steps in Analysis:

1. Parse all files in dataset into SnowPit objects
2. Use SnowPit objects to create a list of layer objects for all layers in all pits
3. Record hand hardness and primary grain form for each layer
4. Map hand hardness to hand hardness group
5. Plot results

Results:

SnowPilot allows the user to enter hand hardness on a scale that includes +/- for each group, for example: F-, F, and F+. We found that the +/- designations were used less and chose to map all sub designations to the parent hand hardness group. Figure 1 shows a clear correlation between hand hardness and primary grain form, which matches our understanding and observations in the field.

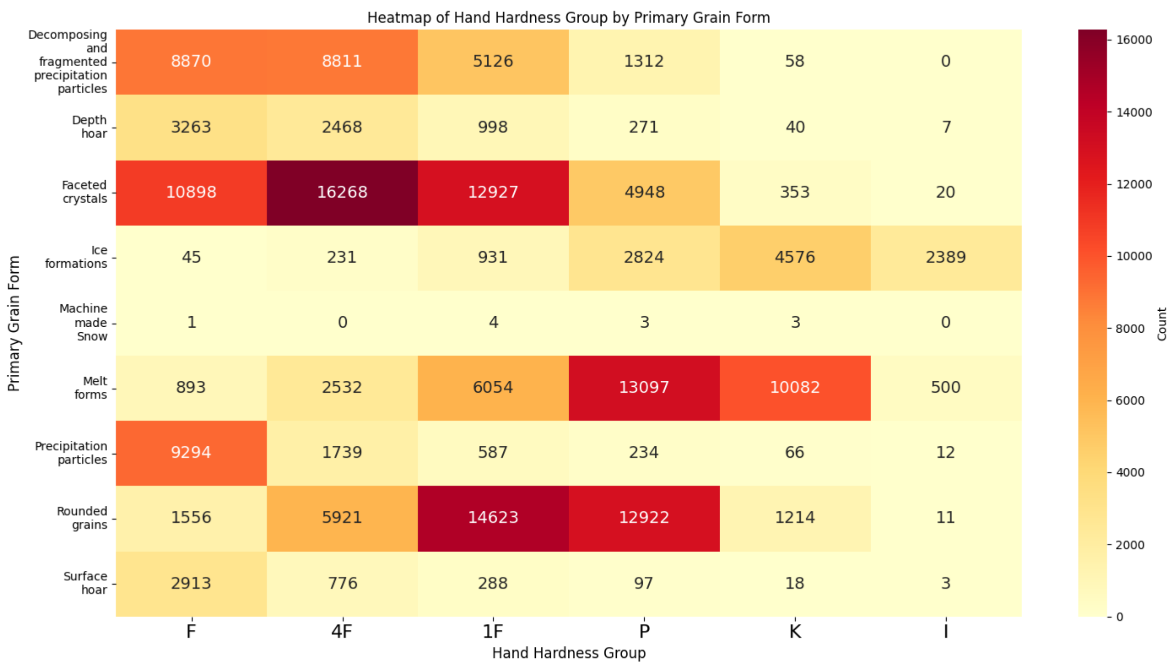

Figure 1. Heatmap of Hand Hardness Group and Primary Grain Form

DATA AND CODE AVAILABILITY

SnowPylot is open-source, and all source code is available on Github (Connelly & Verplanck, 2025c). The CAAML XML files and the example notebooks used for the analyses in this paper are included in the “demos” folder of the Github repository (Connelly & Verplanck, 2025b). The CAAML XML files were downloaded from SnowPilot using the Aviscience Query (SnowPilot, n.d.) on March 31, 2025 and include snow pit observations from the 2020-2024 water years (October 1 to September 30). A small number of individual days are excluded from the dataset due to system errors during export.

DISCUSSION AND FUTURE WORK

By making these tools accessible to the broader avalanche and snow science community, we hope to enhance both research capabilities and practical applications for utilizing data collected in snow pits observations. Some potential applications include snow mechanics and avalanche research, model validation of tools like SNOWPACK (Lehning, 1999), and hydrology and water resource management.

While the available data in SnowPilot is extensive, the data is observational and unvalidated. In our example analysis we found instances where observations and input measurements appeared incorrect. For example, latitude and longitude coordinates that did not fall within the indicated country. While this is an expected challenge with community gathered data, we encourage users of SnowPylot to examine and filter data from the SnowPilot database as appropriate.

Future SnowPylot development will focus on expanding and refining the library's capability. Planned additions include methods to query the SnowPilot database directly, rather than relying on exported files, and improved visualization tools. We encourage community involvement in this development process and invite anyone interested to reach out or submit a pull request from the GitHub repository. We also plan to collaborate with SnowPilot to improve data quality and accessibility. For example, density measurements input as part of a snow pit layer are not currently included in the CAAML.xml export. Adding these measurements to the export will allow them to be accessed by SnowPylot and used for analysis.

ACKNOWLEDGEMENTS

We would like to thank Jim Earl and Doug Chabot for their help in learning about and accessing the SnowPilot database, Philipp Rosendahl, Ron Simenhois and Ryan Mason for their data science and programming help. We would also like to thank the Friends of the Gallatin National Forest for providing financial support to establish and maintain the SnowPilot database and Montana State University Norm Asbjornson College of Engineering for providing financial support for the development of SnowPylot.

REFERENCES

Chabot, D., Kahrl, M., & Earl, J. (2016) SNOWPILOT: ONLINE, UPDATED AND FREE. International Snow Science Workshop.

Canadian Avalanche Association. (n.d.). *CAAML: Canadian Avalanche Association Markup Language*. Retrieved May 7, 2025, from http://caaml.org/

Connelly, M. K., & Verplanck, S. (2025a). *SnowPylot: A Python library for working with SnowPilot data* (v1.1.0) [Python software]. Python Package Index. https://pypi.org/project/snowpylot/

Connelly, M. K., & Verplanck, S. (2025b). *SnowPylot demos: Example analyses and notebooks* [Code repository]. GitHub. https://github.com/connellymk/snowpylot/tree/main/demos

Greene, E., Birkeland, K., Elder, K., McCammon, I., Staples, M., Sharaf, D., ... & Wagner, W. (2022). Snow, Weather, and Avalanches: Observation Guidelines for Avalanche Programs in the United States, Denver, Colorado, 4th edn., American Avalanche Association. ISBN 78-0-9760118-1-1.

Connelly, M. K., & Verplanck, S. (2025c). *SnowPylot* [Code repository]. GitHub. https://github.com/connellymk/snowpylot

SnowPilot. (n.d.). *Avalanche Science Database Query Tool*. Retrieved May 7, 2025, from https://snowpilot.org/avscience-query

Lehning, M., Bartelt, P., Brown, B., Russi, T., Stöckli, U., & Zimmerli, M. (1999). Snowpack model calculations for avalanche warning based upon a new network of weather and snow stations. Cold Regions Science and Technology, 30(1–3), 145–157. https://doi.org/10.1016/S0165-232X(99)00022-1