**Abstract**

SnowPilot (https://snowpilot.org/) is a free, open-source software designed to help users graph, record, and store snowpit data. The SnowPilot database currently contains data from over 65,000 snowpits, collected by avalanche professionals and snow scientists 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 are developing 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.

Additionally, we are creating open-access Jupyter notebooks that demonstrate the utility of this library to investigate snow science questions. A preliminary application of these tools includes a graphical representation of the relationship between hand hardness and primary grain form of snow pit layers from snowpits in Montana. The associated Jupyter notebook outlines the process and methodological approach taken when utilizing these tools to explore the research question.

1. Introduction
   1. SnowPilot
      1. Example png of SnowPilot Pit
   2. SnowPylot
      1. Print output of same pit from snowpilot
   3. Case study
      1. Introduce dataset: 2020-2024 snow years
2. Data Structure
   1. Include link to Github
   2. Snowpit
      1. Core Info
         1. Written description
         2. Table of interesting results from dataset
            1. Number of pits
            2. Number of unique users
            3. Pie chart of countries
      2. Snow Profile
         1. Written description
         2. Table of interesting results
            1. Number of layers
            2. Number of layers w/properties (hand hardness, grain form)
      3. Stability Tests
   3. WhumpfData
      1. Custom information created for a study
3. Example Analyses
   1. ECT / CT question
   2. Hand hardness vs grain form

6. Discussion and Future Work

* Potential applications
* Limitations and potential improvements
* Future development plans
* Invitation for community involvement

1. Update: Selected fields of interest and Information from dataset

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | **Path to** | **Description** | 2020-2024 Data |
| **Snowpit** |  |  |  | Snowpit | Main Snowpit object |  |
|  | **core\_info** |  |  | Snowpit.core\_info | Object to represent “Core Info” user inputs in SnowPilot |  |
|  |  | pit\_id |  |  | Unique identifier from SnowPilot |  |
|  |  | pit\_name |  |  | User input to SnowPilot |  |
|  |  | Date |  |  | Date of observation |  |
|  |  | comment |  |  | User input to SnowPilot |  |
|  |  | caaml\_version |  |  | Version of CAAML schema used in XML |  |
|  |  | **User** |  |  | SnowPilot User Information |  |
|  |  |  | Operation\_id |  | Operation ID from SnowPilot for professional organizations |  |
|  |  |  | Operation\_name |  | Operation Name from SnowPilot for professional organizations |  |
|  |  |  | Professional |  | Boolean indicating if user is professional |  |
|  |  |  | User\_id |  | SnowPilot User identifier |  |
|  |  |  | username |  | SnowPilot username of the user |  |
|  |  | **Location** |  |  | Location of the Observation |  |
|  |  |  | Latitude |  | Decimal degrees |  |
|  |  |  | Longitude |  | Decimal degrees |  |
|  |  |  | Elevation |  | [value, units] |  |
|  |  |  | Aspect |  | Slope Aspect |  |
|  |  |  | slope\_angle |  | [value, units] |  |
|  |  |  | Country |  | Country Name |  |
|  |  |  | Region |  | Region name (State if in US) |  |
|  |  |  | Pit\_near\_avalanche |  | Boolean |  |
|  |  |  | Pit\_near\_avalanche\_location |  | Location description if near avalanche |  |
|  |  | **Weather Conditions** |  |  | Weather Conditions at time of Observation |  |
|  |  |  | Sky\_cond |  |  |  |
|  |  |  | Sky\_cond\_desc |  |  |  |
|  |  |  | Precip\_ti |  |  |  |
|  |  |  | Precip\_ti\_desc |  |  |  |
|  |  |  | Air\_temp\_pres |  |  |  |
|  |  |  | Wind\_speed |  |  |  |
|  |  |  | Wind\_speed\_desc |  |  |  |
|  |  |  | Wind\_dir |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

3. Analysis Capabilities

3.1 Core Information Analysis

* Metadata analysis (pit counts, user statistics)
* Geographic distribution of snow pits
* Professional vs. non-professional contributions
* Temporal patterns in data collection

3.2 Snow Profile Analysis

* Layer characterization
* Temperature and density profiles
* Surface conditions
* Statistical analysis of snow properties

3.3 Stability Test Analysis

* Distribution of test types (ECT, CT, PST, RBlock)
* Test result patterns
* Correlation with snow properties
* Spatial and temporal trends

4. Visualization Tools

* Interactive maps using Folium
* Statistical plots and charts
* Time series visualizations
* Custom plotting functions

5. Case Study: 2020-2024 Analysis

* Dataset overview
* Key findings from the analysis
* Notable patterns and trends
* Implications for avalanche safety

6. Discussion

* Benefits of the library for researchers and practitioners
* Limitations and potential improvements
* Future development plans
* Integration with other avalanche safety tools

7. Conclusion

* Summary of key contributions
* Impact on avalanche safety research
* Call for community involvement

**Snowpylot: A Python Library for Snow Pit Data Analysis and Visualization**

**Abstract**

Snow pit observations are fundamental to avalanche safety and research, providing critical information about snowpack structure and stability. While SnowPilot has successfully amassed a vast database of snow pit observations, the tools for analyzing this data have remained limited. This paper presents the SnowPilot Python library, a comprehensive suite of tools for analyzing and visualizing snow pit data. The library, designed to work with the Canadian Avalanche Association Markup Language (CAAML) format, enables researchers and practitioners to process, analyze, and visualize snow pit observations at both individual and large-scale levels. We demonstrate the library's capabilities through a comprehensive analysis of snow pit observations from the 2020-2024 water years, showcasing its ability to handle large datasets and extract meaningful insights. The analysis includes examination of core metadata, snow profile characteristics, stability test results, and their relationships. We also present various visualization tools developed as part of the library, demonstrating how they can be used to better understand and communicate snowpack conditions and trends.

**1. Introduction**

Snow pit observations are fundamental to avalanche safety and research, providing critical information about snowpack structure and stability. Since its inception, SnowPilot has served as a crucial platform for collecting and sharing these observations within the avalanche community. However, while SnowPilot has successfully amassed a vast database of snow pit observations, the tools for analyzing this data have remained limited, primarily relying on manual interpretation and basic statistical methods. This limitation has hindered the potential for large-scale analysis and the extraction of deeper insights from this valuable dataset.

The development of the SnowPilot Python library addresses this gap by providing a comprehensive suite of tools for analyzing and visualizing snow pit data. This library, designed to work with the Canadian Avalanche Association Markup Language (CAAML) format, enables researchers and practitioners to process, analyze, and visualize snow pit observations at both individual and large-scale levels. By automating the data parsing and analysis processes, the library makes it possible to examine patterns and trends across thousands of snow pit observations, potentially revealing new insights about snowpack behavior and avalanche formation.

The importance of this development cannot be overstated in the context of avalanche safety. With climate change affecting snowpack characteristics and avalanche behavior, having robust tools for analyzing historical and current snow pit data becomes increasingly crucial. The ability to process and analyze large datasets efficiently can help identify emerging patterns, validate existing models, and inform avalanche forecasting and safety protocols. Furthermore, by making these analysis tools accessible to the broader avalanche community, the library has the potential to enhance both research capabilities and practical applications in avalanche safety.

**2. Data Structure and Methods**

The SnowPilot library is designed to work with snow pit observations stored in the Canadian Avalanche Association Markup Language (CAAML) format, version 6.0.3. CAAML is an XML-based standard that provides a structured format for recording snow and avalanche observations. This format includes detailed specifications for core metadata such as pit identification, date, and location; snow profile characteristics including layers, temperature, and density; stability test results; weather conditions; and additional observations like whumpf data.

The SnowPilot Python library implements an object-oriented design that mirrors the CAAML data structure. At its core, the SnowPit class serves as the container for all snow pit observations and consists of four main components. The first component, Core Information, manages pit identification, temporal data, user information, location data, weather conditions, and additional metadata. The second component, Snow Profile, handles layer characteristics, grain properties, temperature profiles, density measurements, and surface conditions. The third component, Stability Tests, manages various test types including Extended Column Tests (ECT), Compression Tests (CT), Rutschblock Tests (RBlock), and Propagation Saw Tests (PST), along with their results and characteristics. The fourth component, Whumpf Data, handles custom SnowPilot-specific observations, including cracking characteristics, remote avalanche triggers, and size classifications.

The library includes a robust CAAML parser that handles the XML data format through a systematic process. First, the parser reads CAAML XML files and extracts the relevant data using the ElementTree library. Second, it validates the data structure against the CAAML schema, ensuring required fields are present, data types are correct, and values are within expected ranges. Finally, the validated data is used to instantiate the appropriate Python objects. This process is implemented through functions that can handle both individual files and entire directories of snow pit observations.

For data analysis, the library provides methods that operate at various scales. At the individual pit level, users can perform layer-by-layer examination, stability test interpretation, and profile visualization. For larger datasets, the library supports batch processing capabilities, allowing for multiple pit parsing, statistical aggregation, and temporal and spatial analysis. The library also includes comprehensive data export functionality, enabling conversion to pandas DataFrames and export to common formats like CSV and JSON, facilitating integration with other analysis tools.

The library is implemented in Python 3.x and relies on several key dependencies. XML parsing is handled by the built-in ElementTree library, while data manipulation is managed through pandas. Visualization capabilities are provided through matplotlib and folium, and numerical operations are supported by numpy. This modular design allows for easy extension and modification while maintaining compatibility with the CAAML standard. The implementation follows Python best practices, with clear documentation and type hints, making it accessible to both researchers and practitioners in the avalanche safety community.

**3. Analysis Capabilities**

The SnowPilot Python library provides comprehensive analysis capabilities for snow pit data, enabling researchers and practitioners to examine patterns and trends at multiple scales. These capabilities are demonstrated through the analysis of snow pit observations from the 2020-2024 water years, which includes over 31,000 snow pit observations from 30 countries.

**3.1 Core Information Analysis**

The library enables detailed analysis of snow pit metadata and core information. From the 2020-2024 dataset, we can see that the data comes from a diverse user base, with 3,854 unique users contributing observations. Of these, 19,891 pits were submitted by professional users, while 11,279 were contributed by non-professional observers. This distribution highlights the library's ability to handle data from both professional avalanche centers and citizen scientists.

Geographic analysis reveals the global reach of the dataset, with observations from 30 different countries. The library can track not only the location of observations but also their relationship to avalanche events. Of the analyzed pits, 945 were recorded near avalanche locations, with 480 specifically on avalanche crowns and 240 on avalanche flanks. This spatial information is crucial for understanding the relationship between snow pit observations and avalanche events.

**3.2 Snow Profile Analysis**

The library provides sophisticated tools for analyzing snow profile characteristics. Each snow pit observation includes detailed information about:

* Layer characteristics (depth, thickness, hardness)
* Grain properties (form, size)
* Temperature profiles
* Density measurements
* Surface conditions

The analysis capabilities allow for both individual pit examination and statistical analysis across multiple observations. This enables researchers to:

* Identify common layer sequences
* Track changes in snow properties over time
* Analyze relationships between different snow characteristics
* Compare profiles across different geographic regions

**3.3 Stability Test Analysis**

One of the library's most powerful features is its ability to analyze stability test results. The dataset includes various types of stability tests:

* Extended Column Tests (ECT)
* Compression Tests (CT)
* Rutschblock Tests (RBlock)
* Propagation Saw Tests (PST)

The library can track not only the test results but also their relationship to snow properties and conditions. This enables:

* Statistical analysis of test results
* Correlation between test results and snow properties
* Temporal and spatial patterns in stability
* Comparison of different test types

**3.4 Statistical Analysis**

The library includes comprehensive statistical analysis capabilities, allowing users to:

* Calculate summary statistics for any measured parameter
* Perform temporal analysis of snow properties
* Analyze spatial patterns in snow characteristics
* Identify correlations between different variables

These statistical tools are particularly valuable for:

* Identifying trends in snowpack behavior
* Understanding regional differences
* Tracking changes over time
* Supporting avalanche forecasting

**3.5 Data Integration**

A key strength of the library is its ability to integrate different types of snow pit data. This integration enables:

* Correlation between weather conditions and snow properties
* Analysis of relationships between stability tests and snow characteristics
* Examination of how different factors influence avalanche formation
* Development of more comprehensive models of snowpack behavior

**4. Visualization Tools**

The SnowPilot library includes a comprehensive suite of visualization tools designed to help users understand and communicate snow pit data effectively. These tools range from basic plots to interactive visualizations, enabling users to explore data at various scales and levels of detail.

**4.1 Geographic Visualization**

The library provides tools for creating interactive maps using Folium, allowing users to:

* Plot snow pit locations
* Visualize spatial patterns in snow properties
* Identify clusters of observations
* Analyze geographic distributions of stability tests

**4.2 Statistical Plots**

A variety of statistical visualization tools are available, including:

* Histograms and density plots
* Box plots and violin plots
* Scatter plots and correlation matrices
* Time series plots

**4.3 Profile Visualization**

Specialized tools for visualizing snow profiles include:

* Layer diagrams
* Temperature and density profiles
* Stability test results visualization
* Combined profile plots

**4.4 Custom Plotting Functions**

The library includes customizable plotting functions that allow users to:

* Create publication-quality figures
* Combine multiple types of data
* Export visualizations in various formats
* Generate standardized reports

**5. Case Study: 2020-2024 Analysis**

The capabilities of the SnowPilot library are demonstrated through a comprehensive analysis of snow pit observations from the 2020-2024 water years. This case study showcases the library's ability to handle large datasets and extract meaningful insights.

**5.1 Dataset Overview**

The analysis includes over 31,000 snow pit observations from 30 countries, collected by both professional and non-professional observers. The data spans multiple seasons and geographic regions, providing a rich dataset for analysis.

**5.2 Key Findings**

The analysis revealed several interesting patterns and trends:

* Geographic distribution of observations
* Temporal patterns in data collection
* Relationships between snow properties and stability
* Regional variations in snow characteristics

**5.3 Notable Patterns**

Several notable patterns emerged from the analysis:

* Seasonal trends in snow properties
* Geographic clustering of observations
* Correlation between weather conditions and snow characteristics
* Patterns in stability test results

**5.4 Implications**

The findings have important implications for:

* Avalanche forecasting
* Snowpack modeling
* Research methodology
* Data collection protocols

**6. Discussion**

The development of the SnowPilot Python library represents a significant advancement in the tools available for analyzing snow pit data. The library's capabilities enable new approaches to understanding snowpack behavior and avalanche formation.

**6.1 Benefits**

The library provides numerous benefits for researchers and practitioners:

* Efficient processing of large datasets
* Standardized analysis methods
* Enhanced visualization capabilities
* Improved data integration

**6.2 Limitations**

Current limitations include:

* Data quality variations
* Geographic coverage gaps
* Temporal resolution constraints
* Processing speed for very large datasets

**6.3 Future Development**

Planned improvements include:

* Enhanced machine learning capabilities
* Additional visualization tools
* Improved data validation
* Extended API functionality

**6.4 Integration**

The library's potential for integration with other tools includes:

* Avalanche forecasting models
* Weather data systems
* Geographic information systems
* Research platforms

**7. Conclusion**

The SnowPilot Python library provides a powerful set of tools for analyzing and visualizing snow pit data. Through the analysis of the 2020-2024 dataset, we have demonstrated the library's capabilities and its potential for contributing to avalanche safety research and practice.The library's development represents a significant step forward in the tools available for snow pit analysis. By making these tools accessible to the broader avalanche community, we hope to enhance both research capabilities and practical applications in avalanche safety.Future development will focus on expanding the library's capabilities, improving its integration with other tools, and enhancing its usability for both researchers and practitioners. We encourage community involvement in this development process, as it is through collaboration that we can best serve the needs of the avalanche safety community.