The Journal of Open Source Software

tsp ("Teaspoon"): A library for ground temperature data

Nicholas Brown [1¶]

1 Carleton University, Canada ¶ Corresponding author

DOI: 10.21105/joss.04704

Software

- Review
- Repository
- Archive

Editor: Katy Barnhart

Reviewers:

- @Elchin
- @mdpiper

Submitted: 19 July 2022
Published: 28 September 2022

Statement of need

Permafrost thaw is a major concern both in the context of climate change as part of global carbon feedback systems (Natali et al., 2021) and as geohazard threatening ways of life in northern regions (Calmels et al., 2014; Meredith et al., 2019) and infrastructure (Hjort et al., 2022). Many kinds of data are used to understand permafrost (Wang et al., 2018). Of these, ground temperature time series data are commonly used to understand and predict permafrost thaw, for example through the development and evaluation of numerical models (Melton et al., 2019).

Unfortunately, cleaning, organizing, and managing ground temperature data continues to be an obstacle. A survey of 37 data-producing institutions worldwide suggests that more than 50% of environmental science datasets use 'in-house' text formats to store and distribute data (Bavay et al., 2020). Consequently, most researchers rely on ad-hoc scripts to manage data; new data sources commonly take a day or more to integrate into workflows. Members of the permafrost community describe a lack of capacity as one of the main barriers to better data management and more frequent data publication (Brown et al., 2020).

Adding to this challenge, dataloggers for environmental science data can be configured to output text files in many different styles. Differences in date format, column separator, and metadata structure confound attempts to develop reusable scripts to easily open datalogger files for data exploration or further processing. A script that one scientist uses to read their files can fail when trying to open files that were configured differently, even when data are from a sensor of the same make and model. This makes it harder to share data and to use a common set of tools for quality control and analysis. Similarly, numerical models for permafrost produce output in different file formats and with different structures.

Ultimately, the lack of standardization contributes to permafrost data being less FAIR (findable, accessible, interoperable, and reuseable) (Wilkinson et al., 2016) than they could be. One way to mitigate this is to provide reuseable tools to handle permafrost data. `tsp` is a Python library designed to streamline all aspects of working with permafrost ground temperature data. It does this by meeting three objectives:

1. Increase the ease with which ground temperature time series data from any source can be read into Python.
2. Provide reuseable functions to perform the most common visualization and analysis conducted on permafrost data.
3. Improve permafrost data interoperability by providing methods to output data as common text file formats used by the permafrost community and as more standardized, self-documenting formats such as netCDF.

`tsp` includes a library of file readers designed to handle different model outputs, datalogger exports and database standard formats. It also includes functions to produce common data visualizations to speed up the data exploration phase of a project. `tsp` can be used as a

standalone library for interactive data analysis and exploration or can be integrated into other programs and workflows. It was designed with community participation in mind; contributions are encouraged to make it a tool that can be adopted by the broadest possible audience.

Comparison to similar packages

pandas (McKinney, 2010; The pandas development team, 2020) provides many of the same features as `tsp` and, indeed, is used extensively in this package. However, the heterogeneity of permafrost datasets inevitably means that researchers must continuously re-implement new configuration parameters to read data from new sources, and these configurations are not typically shared among the research community. Additionally, certain tasks are unintuitive for novice Python users such as conversion between "wide" and "tidy" forms of data. Teaspoon providers higher-level functionality to accomplish these tasks. The `Data Integration Tool` (Wilcox et al., 2020) was designed to standardize permafrost data for ingestion into the Global Terrestrial Network for Permafrost (GTN-P) database, but comes with no pre-built templates, and so must be configured by each user. It also is targeted towards data managers rather than data users and provides no visualization capabilities. The R-package `PIC` is designed to compute permafrost indices using meteorological and ground temperature data (Luo et al., 2018). It provides some capability for visualizing station locations as maps but no visualization of the station data itself. Although it is possible to use this package with any data, the implementation and user experience strongly favours the dataset with which it was created, and there is little emphasis on data interoperability.

Summary

TSP

The core of the `tsp` package is the `TSP` class (Figure 1). This provides a representation of the observations (usually of ground temperature) collected at specified times and depths (a Time Series of Profiles). This class provides methods to handle many of the common use-cases needed in permafrost research such as generating visualizations, outputting data, and creating summary statistics (e.g. yearly, monthly, or daily means). Data summaries provide an added layer of value by masking out any periods that don't have sufficient data coverage (defined by a user-selectable combination of the number of observations and the length of any data gaps).

Datalogger outputs rarely include depth values. To accommodate this, `tsp` provides the `IndexedTSP` subclass, which treats depth values as indices. Objects of this type behave much like their parent `TSP` class but some features are restricted or give a warning during execution. An `IndexedTSP` can be transformed into a `TSP` by supplying depth information.

To encourage data interoperability by promoting the use of existing file structures and conventions, the `TSP` class and subclasses provides methods to save data in a variety of standard forms used by national or global permafrost databases. More specifically, data can be saved as csv files in the format of the GTN-P database (Biskaborn et al., 2015) or the Northwest Territories Geological Survey (NTGS) Database (Karunaratne et al., 2015). Data can also be saved as a CF-compliant netCDF file using the 'timeSeriesProfile' discrete sampling geometry.

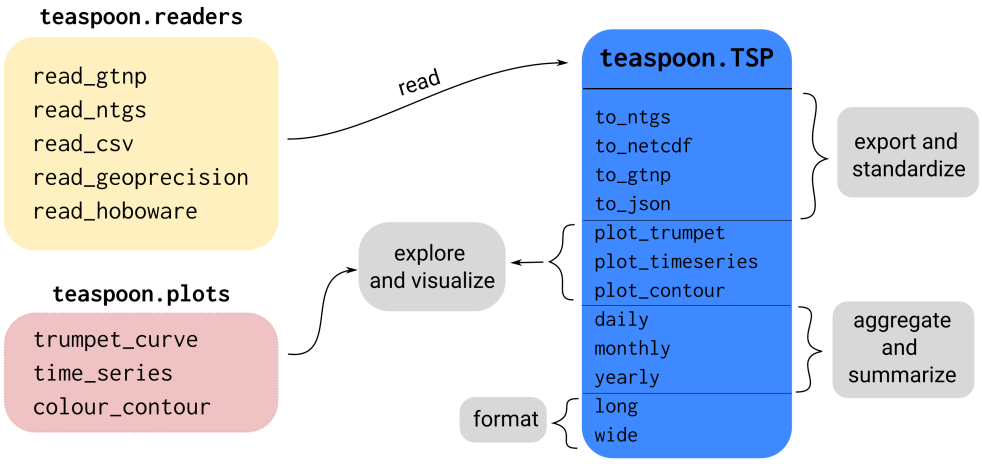

Figure 1: An overview of the TSP class and how it is used to read, write, summarize, and vizualize ground temperature data.

Readers

There are many ways of obtaining permafrost time series profiles including model outputs, dataloggers, and databases. Each of these typically creates text files with a different structure. Accordingly `tsp` provides a library of readers to handle different source datasets. As of version 1.2.0, `tsp` includes readers for dataloggers manufactured by Geoprecision and Onset (HOBOWare) and for ground temperature data downloaded from the GTN-P database (Biskaborn et al., 2015) or from the NTGS database (Karunaratne et al., 2015). It also has readers for model output from GEOtop (Endrizzi et al., 2014) and CLASSIC (Melton et al., 2020).

For datalogger outputs, no attempt was made to create a 'universal' csv reader. Instead, each device manufacturer is treated separately so that as much prior information about the possible variability in output format could be utilized. This increases the reliability and maintainability of each individual reader. To support testing, a collection of example files is included in the library that reflects the possible output configurations of supported dataloggers.

Visualization

`tsp` provides several data visualizations that are commonly used in permafrost data: time series plots, trumpet curves, and colour-contour plots. These are not designed to be publication-ready visualizations (because of the diverse and specialized nature of such visualization) but rather to provide a way to quickly and interactively view data as part of the reasearch and data exploration phases of a project. Much like the ad-hoc scripts used to read in data, basic permafrost data visualization code is often re-written and rarely recycled. To promote reuseability, each plot type is provided as a datasource-agnostic function that accepts the data to be visualized as explicit parameters (instead of relying on implicit information in the DataFrame or CSV structure). Plots can also be generated using the plotting methods of the `TSP` class. These methods handle all necessary data reshaping and make it possible to begin data exploration in as few as three lines of code (Figure 2):

```
from teaspoon import read_geotop
tsp_geotop = read_geotop('model_output.csv')
tsp_geotop.plot_trumpet(title="A trumpet plot using teaspoon", year=2015, max_depth=4)
```

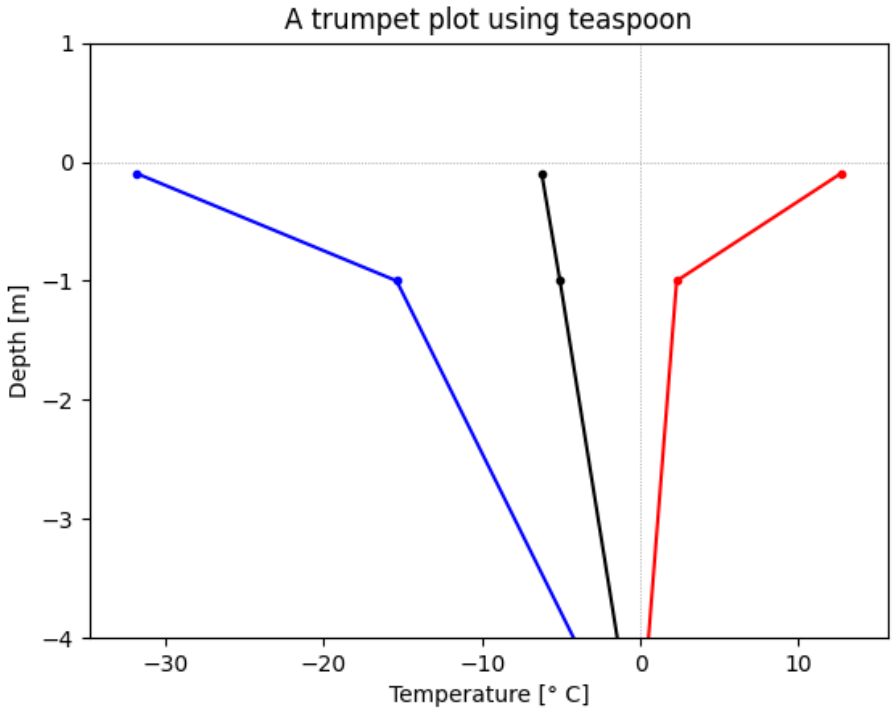

Figure 2: An example of a trumpet plot made with teaspoon, showing the maximum, minimum and mean ground temperature over a year.

Standardizing data processing

One of the motivations for the `tsp` package is the need for rapid, effective QA/QC of ground temperature time series from both sensor data and model output. Too often, researchers re-write code for either reading their data into data frames or for making generic plots for visual inspection (often the first and only check performed on a dataset). This version of `tsp` already streamlines these two phases of the QA/QC process.

In addition, this software supports the standardization of ground temperature data processing techniques (including QA/QC and gap-filling) in three ways. First, it provides a common starting point for data from various sensors and databases. This means techniques that are developed by one research group using `tsp` can be more easily adopted by others. Second, it provides a framework for adding generic format-agnostic functions in a separate module of the `tsp` package to provide a 'one-stop-shop' for temperature data handling. The plotting functions exemplify this framework in that they explicitly require only the essential data as parameters. Finally, the TSP class itself provides a way to streamline running those functions by doing the data manipulation behind the scenes. By providing an easy solution in the `TSP` class, more people are likely to adopt the same techniques, which may become a *de facto* standard.

Resources and Contributions

Contributions to `teaspoon` are encouraged. Possible contributions include: support for more dataloggers, examples of datalogger text file outputs to represent variability for testing, new visualizations, new techniques for data cleaning and QA/QC, or the addition of relevant parameters to existing visualizations.

Links to the code repository, documentation, and information for contributors is provided below:

- Repository - https://gitlab.com/permafrostnet/teaspoon/
- Documentation - https://permafrostnet.gitlab.io/teaspoon/
- Contributing - https://permafrostnet.gitlab.io/teaspoon/source/contributions.html

Acknowledgements

This project was supported by funding from the National Sciences Engineering and Research Council (NSERC) through a Strategic Partnership Grant for Networks to NSERC PermafrostNet.

References

Bavay, M., Fiddes, J., & Godøy, Ø. (2020). Automatic Data Standardization for the Global Cryosphere Watch Data Portal. *Data Science Journal*, *19*. https://doi.org/10.5334/dsj-2020-006

Biskaborn, B., Lanckman, J.-P., Lantuit, H., Elger, K., Streletskiy, D., Cable, W., & Romanovsky, V. (2015). The new database of the Global Terrestrial Network for Permafrost (GTN-P). *Earth System Science Data*, *7*. https://doi.org/10.5194/essd-7-245-2015

Brown, N., Gruber, S., P., P., & Stewart-Jones, E. (2020). *Permafrost Data Workshop 2020: Final Report*. Carleton University. https://doi.org/10.22215/pn/10120001

Calmels, F., Laurent, C., Brown, R., Pivot, F., & Ireland, M. and. (2014). How permafrost-related landscape changes may impact food security of Jean Marie River First Nation, NWT. Paper presented at Arctic Change 2014, 8–12 December, 2014, Ottawa, ON. *Proceedings of GeoQuebec 2015, 68th Canadian Geotechnical Conference and 7th Canadian Permafrost Conference*.

Endrizzi, S., Gruber, S., Dall'Amico, M., & Rigon, R. (2014). GEOtop 2.0: Simulating the combined energy and water balance at and below the land surface accounting for soil freezing, snow cover and terrain effects. *Geoscientific Model Development*, *7*(6), 2831–2857. https://doi.org/10.5194/gmd-7-2831-2014

Hjort, J., Streletskiy, D., Doré, G., Wu, Q., Bjella, K., & Luoto, M. (2022). Impacts of permafrost degradation on infrastructure. *Nature Reviews Earth & Environment*, *3*(1), 24–38. https://doi.org/10.1038/s43017-021-00247-8

Karunaratne, K., Kokelj, S., M., P., S., W., & Gruber, S. (2015). Metadata protocol for ground temperature records in the northwest territories. *Proceedings of GeoQuebec 2015, 68th Canadian Geotechnical Conference and 7th Canadian Permafrost Conference*, 1–8.

Luo, L., Zhang, Z., Ma, W., Yi, S., & Zhuang, Y. (2018). PIC v1.3: comprehensive R package for computing permafrost indices with daily weather observations and atmospheric forcing over the Qinghai–Tibet Plateau. *Geoscientific Model Development*, *11*. https://doi.org/10.5194/gmd-11-2475-2018

McKinney, Wes. (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt & Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56–61). https://doi.org/10.25080/Majora-92bf1922-00a

Melton, J., Arora, V. K., Wisernig-Cojoc, E., Seiler, C., Fortier, M., Chan, E., & Teckentrup, L. (2020). CLASSIC v1.0: The open-source community successor to the canadian land surface scheme (CLASS) and the canadian terrestrial ecosystem model (CTEM) – part 1: Model framework and site-level performance. *Geoscientific Model Development*, *13*(6), 2825–2850. https://doi.org/10.5194/gmd-13-2825-2020

Melton, J., Verseghy, D. L., Sospedra-Alfonso, R., & Gruber, S. (2019). Improving permafrost physics in the coupled canadian land surface scheme (v.3.6.2) and canadian terrestrial ecosystem model (v.2.1) (CLASS-CTEM). *Geoscientific Model Development*, *12*(10), 4443–4467. https://doi.org/10.5194/gmd-12-4443-2019

Meredith, M., Sommerkorn, M., Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A., Kofinas, G., Mackintosh, A., Melbourne-Thomas, J., Muelbert, M. M. C., Ottersen, G., Pritchard, H., & Schuur, E. A. G. (2019). Polar regions. In H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, & N. M. Weyer (Eds.), *IPCC special report on the ocean and cryosphere in a changing climate* (pp. 203–320). https://doi.org/10.1016/S1366-7017(01)00066-6

Natali, S. M., Holdren, J. P., Rogers, B. M., Treharne, R., Duffy, P. B., Pomerance, R., & MacDonald, E. (2021). Permafrost carbon feedbacks threaten global climate goals. *Proceedings of the National Academy of Sciences*, *118*(21). https://doi.org/10.1073/pnas.2100163118

The pandas development team. (2020). *Pandas-dev/pandas: pandas* (latest) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.3509134

Wang, K., Jafarov, E., Overeem, I., Romanovsky, V., Schaefer, K., Clow, G., Urban, F., Cable, W., Piper, M., Schwalm, C., Zhang, T., Kholodov, A., Sousanes, P., Loso, M., & Hill, K. (2018). A synthesis dataset of permafrost-affected soil thermal conditions for alaska, USA. *Earth System Science Data*, *10*(4), 2311–2328. https://doi.org/10.5194/essd-10-2311-2018

Wilcox, H., Schaefer, K., Jafarov, E., Pulsifer, P., Strawhacker, C., Yarmey, L., & Basak, R. (2020). Data Integration Tool: Permafrost Data Debugging. In *AGU Fall Meeting Abstracts*. https://ui.adsabs.harvard.edu/abs/2017AGUFMIN11D..05W/abstract

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18