The Journal of Open Source Software

ndbc-api: Accelerating oceanography and climate science research with Python

Christopher David Jellen [1]

1 United States Naval Academy, USA

DOI: 10.21105/joss.07406

Software

- Review
- Repository
- Archive

Editor: Taher Chegini

Reviewers:

- @rwegener2
- @ks905383

Submitted: 02 September 2024
Published: 04 December 2024

Summary

The National Data Buoy Center (NDBC) and its partners are an essential source of marine meteorological and oceanographic data (National Data Buoy Center (NDBC), 2023). The `ndbc-api` Python package is an open-source tool designed to streamline the acquisition, synthesis, and analysis of this data. It provides a Python API for accessing real-time and historical observations from a network of buoys, coastal stations, and deployments. This package simplifies the process of retrieving, parsing, and organizing NDBC data. Traditional HTTP access methods to station-specific text files poses a challenge when dealing with multiple stations or extended time ranges. The `ndbc-api` provides a Python API which retrieves, parses, and merges this data, returning Python-native data structures. This package empowers researchers and practitioners in oceanography, meteorology, and related fields to efficiently integrate NDBC data into their workflows, accelerating research in climate science and oceanography.

Statement of need

The National Oceanic and Atmospheric Association's National Data Buoy Center maintains marine monitoring and observation stations around the world (National Data Buoy Center (NDBC), 2024). These stations report atmospheric, oceanographic, and other meteorological data at regular intervals to the NDBC (National Data Buoy Center (NDBC), 2023, 2024). Measurements are made available over HTTP through the NDBC's data service. Measurements are typically distributed as quality-controlled `utf-8` encoded, station-by-station, fixed-period text files (National Data Buoy Center (NDBC), 2023). While the data collected and maintained by the NDBC is critical to oceanography and climate science researchers, the mode of access adds complexity to their workflows. These challenges are particularly pronounced when working with long-duration data, data from multiple stations, or data with a high proportion of missing measurements.

The `ndbc-api` addresses these critical gaps by providing a streamlined, programmatic interface to the NDBC's data service. By abstracting the complexities of file-based access, handling of missing measurements, and cross-station joins, the `ndbc-api` package lowers the barriers to obtaining and using the NDBC's global oceanographic and meteorological data in scientific research. Researchers specify their stations, data modalities (for example, `adcp` for acoustic doppler current profiler measurements or `cwind` for continuous winds data), and time ranges of interest, and the package returns the processed NDBC data either as a Pandas `DataFrame` object or as an xarray `Dataset` object (Hoyer & Hamman, 2017; Mckinney, 2010; The pandas development team, 2020). The package maps missing measurements from their varied text-based identifiers such as `99`, `999`, or `MM`, into a single missing measurement representation of `nan`. The challenge of aligning and joining data across measurements and stations is similarly handled before the final object is returned to the user. By exposing station metadata and

search functionality alongside data retrieval methods, researchers are also able to identify the set of stations, based on their NDBC identifier, that were active during a given period, or within some radius of a given location. The combination of efficient identification of the relevant stations, retrieval of the desired data variables, and processing of the data from the NDBC data service make the `ndbc-api` a valuable tool for oceanography and climate science researchers.

Acknowledgements

The authors would like to acknowledge the invaluable contributions of the National Data Buoy Center (NDBC), their teams, and their network of partners. Their efforts in deploying and maintaining the buoys and marine observation stations, collecting and quality-controlling oceanographic data, and making their measurements readily available through their data service are fundamental to advancing our understanding of earth systems. The `ndbc-api` Python package would not be possible without their commitment to open data access and their unwavering support for the scientific and operational communities. We extend our sincere appreciation to everyone involved in this vital endeavor.

References

Hoyer, S., & Hamman, J. (2017). Xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, *5*(1). https://doi.org/10.5334/jors.148

Mckinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, *445*(1), 56–61. https://doi.org/10.25080/majora-92bf1922-00a

National Data Buoy Center (NDBC). (2023). *NDBC web data guide*. https://www.ndbc.noaa.gov/docs/ndbc_web_data_guide.pdf.

National Data Buoy Center (NDBC). (2024). *NDBC active stations*. https://www.ndbc.noaa.gov/wstat.shtml.

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