

MOODA: The Module for Ocean Observatory Data Analysis and Harmonization

Raul Bardaji¹, Ivan Rodero², Carlos Rodero³, Jaume Piera³, Juanjo Dañobeitia^{1,2}

¹ Marine Technology Unit – CSIC, Passeig Marítim de la Barceloneta, 37-49, Barcelona, Spain, bardaji@utm.csic.es,

² EMSO ERIC Central Management Office, Via Giunio Antonio Resti, 63, Rome, Italy, ivan.rodero@emso-eu.org and juanjo.danobeitia@emso-eu.org

³ Marine Science Institute – CSIC, Passeig Marítim de la Barceloneta, 37-49, Barcelona, Spain, rodero@icm.csic.es and jpiera@icm.csic.es

Abstract – The development of straightforward tools in data analysis plays a significant role in the available accumulative data from marine observatories. A large number of different variables recorded with different formats in marine observatories require methodologies that allow analysis and integration of these data automatically. In this paper, we present the MOODA open-source Python package, which provides an extensive range of procedures for harmonization and analysis of data from marine observatories, including feature extraction, quality control, filtering features, and visualization tools. We present the key aspects of the design and implementation of the package (mooda v1.x). MOODA is an integral component of the European Multidisciplinary Seafloor and water-column Observatory (EMSO) data management platform to harmonize and manage data from the different marine observatories. MOODA's dynamic development model based on user feedback achieves a continuous enhancement and integration of the library.

Keywords – Python, data science, data harmonization, EMSO ERIC

I. INTRODUCTION

The European Multidisciplinary Seafloor and water-column Observatory (EMSO) European Research Infrastructure Consortium (ERIC) [1] is a large-scale, distributed, marine Research Infrastructure (RI). EMSO ERIC consists of ocean observation systems for long-term, high-resolution, (near) real-time monitoring of environmental processes, including natural hazards, climate change, and marine ecosystems. EMSO ERIC observatory nodes are at key environmental sites across European seas, from the North Atlantic, through the Mediterranean, to the Black Sea. To analyze and harmonize the data from the different observatories, we developed MOODA, the Module for Ocean Observatory Data Analysis.

MOODA is a Python library that facilitates data access, transport, and analysis by the scientific community and stakeholders. Some of MOODA's features are:

- Direct data access with sophisticated query capabilities
- Data filtering methods based on metadata information
- Complex visualization tools
- Summary reports of the validated data generated from a specific query, including event annotations
- Specific data analysis tools for different scientific disciplines

We have added algorithms to open and analyze data from European and International marine data sources (e.g., EMODnet, Copernicus CMEMS, NSF Ocean Observatories Initiative [2,3], Ocean Networks Canada) and raw data from oceanographic instrumentation. MOODA is open source, adaptable, and scalable, which allows contributions from researchers and developers from all the disciplines associated with the marine observatories. In this contribution, we present the code package design and its main characteristics.

II. MOODA FEATURES

MOODA consist of a set of modules. These modules are classified into tree types:

- Analysis modules. Libraries that contain a set of functions and classes for processing and analyzing data. For example, `mooda.WaterFrame.min(parameter)`, that returns the minimum value, time and location (if the information is available) of the input parameter.
- Input/output modules. Modules that contain a set of subclasses for reading data from a particular marine observatory or instrument. Module functions consist of translating the input observatory data to the standardized data frames

(i.e., WaterFrames) for more accessible analysis. E.g., `mooda.read_nc()`, that allows to open datasets from a NetCDF.

- Plot modules. Modules that contain a set of classes and functions to create plots. For example, `mooda.WaterFrame.plot_timeseries(parameter)`, that makes a plot of the input parameter with a trace of the values of the time series and a shadow trace of the standard deviation of the values (see Figure 1).

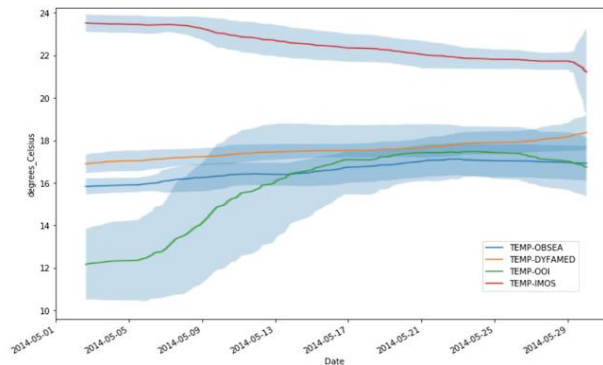


Figure 1. Example of plot generated with `mooda.WaterFrame.plot_timeseries()`

- A differential issue between the MOODA package and other existing data analysis packages, which adds significant value, is the integration of data quality control management. The MOODA package generates data quality control flags based on the QUAL - GTSPP data quality codes [4] and OceanSites quality control flags [5].

MOODA can be downloaded and installed with the popular Python package management system “pip” by using the command “pip install mooda” or using the source code available in Github [6].

III. CONCLUSIONS

MOODA is an open-source Python library that implements a wide range of marine data analysis functionalities. The library is useful to open, manage and analyze data from files and data sources from in-situ oceanographic data sources (e.g., EMODnet, Copernicus CMEMS, NSF Ocean Observatories Initiative, Ocean Networks Canada) raw data from oceanographic instrumentation. The package design and implementation model is adaptable and scalable, allowing continuous user input and contributions.

ACKNOWLEDGMENTS

This work is supported by the CSIC Intramural Project EMSO – Laboratorios Submarinos Profundos and EMSO ERIC.

REFERENCES

- [1] J. J. Dañobeitia, S. Pouliquen, T. Johansen, A. Basset, et al., “Towards a comprehensive and integrated strategy of the European Marine Research Infrastructures for Ocean Observations”, *Front. Mar. Sci.* 7:180. doi: 10.3389/fmars.2020.00180.
- [2] L. M. Smith, J. A. Barth, D. S. Kelley, A. Plueddemann, I. Rodero, G. A. Ulses, M. F. Vardaro, R. Weller, “The ocean observatories initiative”, *Oceanography* 31(1), 16–35(2018).
- [3] I. Rodero, M. Parashar, “Architecting the Cyberinfrastructure for National Science Foundation Ocean Observatories Initiative (OOI)”, 7th International Workshop on Marine Technology : MARTECH 2016. "Instrumentation viewpoint". Vilanova i la Geltrú: SARTI, 2016, p. 99-101.
- [4] QUAL - GTSPP data quality codes - https://www.nodc.noaa.gov/GTSPP/document/codetbls/gtsppcodes/gtspp_qual.html
- [5] OceanSites data format reference manual - http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf
- [6] MOODA GitHub web site - <https://github.com/rbardaji/mooda>