The Journal of Open Source Software

VirtualShip for simulating oceanographic fieldwork anywhere in the global ocean

Jamie R. C. Atkins[1,2]¶, **Emma E. Daniels**[2], **Nick Hodgskin**[1], **Aart C. Stuurman**[1], **Iury Simoes-Sousa**[2], and **Erik van Sebille**[1,2]

1 Institute for Marine and Atmospheric Research, Utrecht University, the Netherlands **2** Freudenthal Institute, Utrecht University, the Netherlands **3** Woods Hole Oceanographic Institution, Falmouth, MA, USA ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

`VirtualShip` is a Python-based package for simulating measurements as if they were coming from real-life oceanographic instruments, facilitating student training, expedition planning, and Observing System Simulation Experiments (OSSEs). The software exploits the customisability of the open-source `Parcels` Lagrangian simulation framework (Delandmeter & van Sebille, 2019; Lange & van Sebille, 2017) and builds a virtual ocean by streaming data from the Copernicus Marine Data Store on-the-fly, enabling virtual expeditions anywhere on the globe.

Statement of need

Marine science relies on fieldwork for data collection, yet sea-going opportunities are limited due to financial costs, logistical constraints, and environmental burdens. We present an alternative means, namely `VirtualShip`, for training scientists to conduct oceanographic fieldwork in an authentic manner, to plan future expeditions and deployments, and to directly compare observational and instrumentational strategies with model data.

`VirtualShip` goes beyond simply extracting grid-cell values from model output. Instead, it uses programmable behaviours and sophisticated interpolation techniques (with `Parcels` underpinnings) to access data in exact locations and timings, as if they were being collected by real-world instruments. `VirtualShip` shares some functionality with existing tools, such as `OceanSpy` (Almansi et al., 2019) and `VirtualFleet` (Maze & Balem, 2023), but extends capabilities to mesh many different instrument deployments into a unified expedition simulation framework. Moreover, `VirtualShip` exploits readily available, streamable data via the Copernicus Marine Data Store, removing the need for users to download and manage large datasets locally and/or arrange for access to remote servers.

Functionality

`VirtualShip` simulates the deployment of virtual instruments commonly used in oceanographic fieldwork, with emphasis on realism in how users plan and execute expeditions. For example, users must consider ship speed and instrument deployment/recovery times to ensure their expedition is feasible within given time constraints. Possible instrument selections include surface `Drifter`, `CTD` (Conductivity-Temperature-Depth), `Argo float`, `XBT` (Expendable Bathythermograph), underway `ADCP` (Acoustic Doppler Current Profiler), and underway `Underwater_temperature/salinity` probes. More detail on each instrument is available in the documentation.

The software can process data and simulate complex multidisciplinary expeditions. One example is a virtual expedition across the Agulhas Current and the South Eastern Atlantic that deploys a suite of instruments to sample physical and biogeochemical properties (Figure 1). Key circulation features appear early in the track, with enhanced ADCP velocities marking the strong Agulhas Current (Figure 1b) and drifters that turn back toward the Indian Ocean indicating the Agulhas Retroflection (Figure 1c). The CTD profiles capture the vertical structure of temperature and oxygen along the route, including the warmer surface waters of the Agulhas region (Figure 1d, early waypoints) and the Oxygen Minimum Zone in the South Eastern Atlantic (Figure 1e, final waypoints).

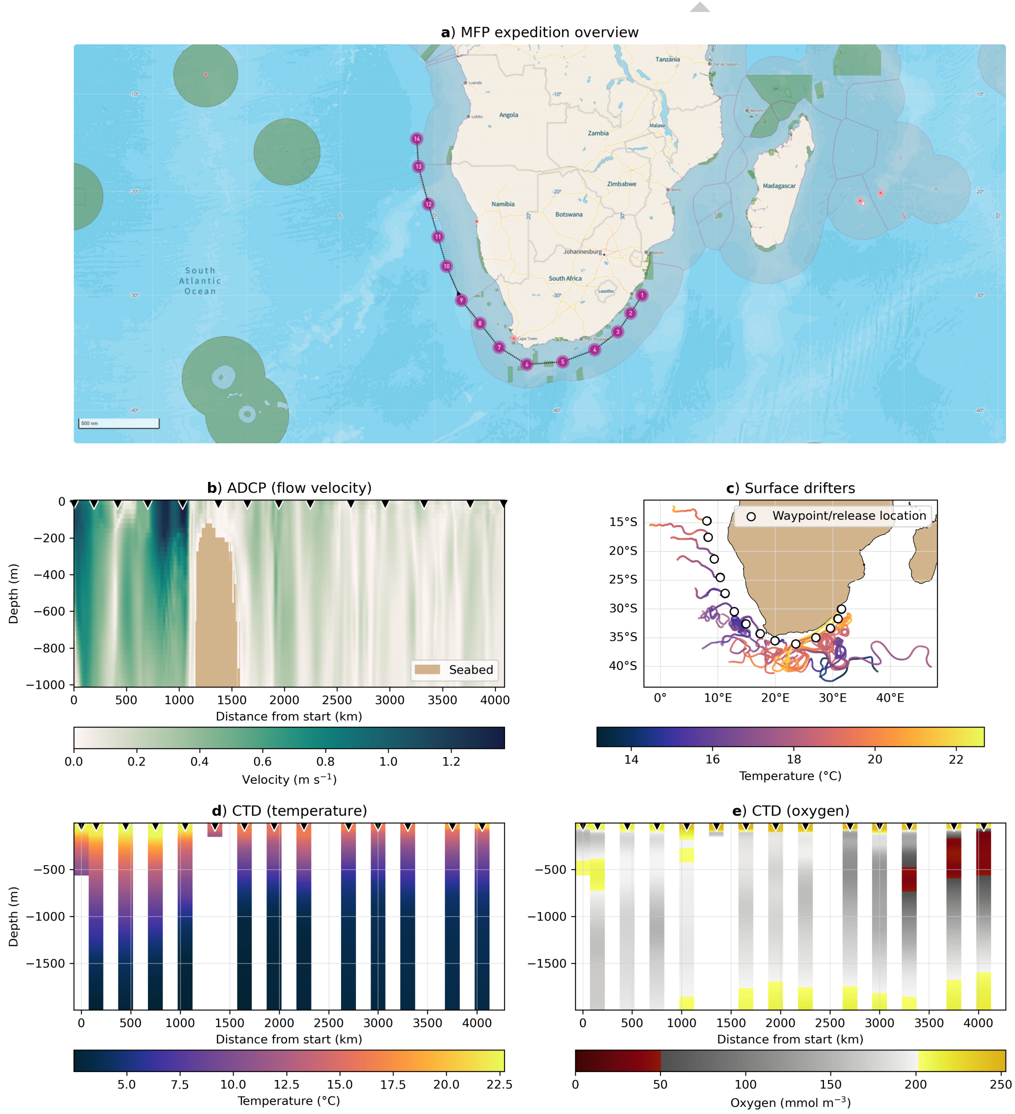

Figure 1: Example VirtualShip expedition simulated in July/August 2023. Expedition waypoints displayed via the NIOZ MFP tool (a), Underway ADCP measurements (b), Surface drifter releases (c; 90-day lifetime per drifter), and CTD vertical profiles for temperature (d) and oxygen (e). Black triangles in b), d) and e) mark waypoint locations across the expedition route, corresponding to the purple markers in a).

The software is designed to be highly accessible to the user. It is wrapped into three high-level command line interface commands (using Click):

1. `virtualship init`: Initialises the expedition directory structure and a `expedition.yaml`

configuration file, which controls the expedition route, instrument choices and deployment timings. A common workflow is for users to first define expedition waypoint locations via the external NIOZ Marine Facilities Planning (MFP) mapping tool. The coordinates can be exported and fed into `init` via the `--from-mfp` flag.

2. `virtualship plan`: Launches a user-friendly Terminal-based expedition planning User Interface (UI), built using `Textual`. This allows users to intuitively modify their expedition waypoint locations, timings and instrument selections.
3. `virtualship run`: Executes the virtual expedition according to the planned configuration. This includes streaming data via the Copernicus Marine Data Store, simulating the instrument beahviours and sampling, and saving the output in `Zarr` format.

A full example workflow is outlined in the Quickstart Guide documentation.

Implementation

Under the hood, `VirtualShip` is modular and extensible. The workflows are designed around `Instrument` base classes and instrument-specific subclasses and methods. This means the platform can be easily extended to add new instrument types. Instrument behaviours are coded as `Parcels` kernels, which allows for extensive customisability. For example, a `Drifter` advects passively with ocean currents, a `CTD` performs vertical profiling in the water column and an `ArgoFloat` cycles between ascent, descent and drift phases, all whilst sampling physical and/or biogeochemical fields at their respective locations and times.

Moreover, the data ingestion system relies on Analysis-Ready and Cloud-Optimized data (ARCO; Stern et al. (2022), Abernathey et al. (2021)) streamed directly from the Copernicus Marine Data Store, via the `copernicusmarine` Python toolbox. This means users can simulate expeditions anywhere in the global ocean without downloading large datasets by default. Leveraging the suite of physics and biogeochemical products available on the Copernicus plaform, expeditions are possible from 1993 to present and forecasted two weeks into the future. There is also an option for the user to specify local `NetCDF` files for data ingestion, if preferred, with the necessary file stuctures and naming conventions outlined in the relevant documentation.

Applications and future outlook

`VirtualShip` has already been extensvely applied in Master's teaching settings at Utrecht University as part of the VirtualShip Classroom initiative. Educational assignments and tutorials have been developed alongside to integrate the tool into coursework, including projects where students design their own research question(s) and execute their fieldwork and analysis using `VirtualShip`. Its application has been shown to be successful, with students reporting increased self-efficacy and knowledge in executing oceanographic fieldwork (Daniels et al., 2025). We encourage researchers to continue to use `VirtualShip` as a cost-effective means to plan future expeditions, as a tool for accessing ocean model data in a realistic manner, and to compare models to observations in a like-for-like manner.

Both the customisability of the `VirtualShip` platform and the exciting potential for new ARCO-based data hosting services in domains beyond oceanography (e.g., atmospheric science) means there is potential to extend VirtualShip (or "VirtualShip-like" tools) to other domains in the future. Furthermore, as the `Parcels` underpinnings themselves continue to evolve, with a future (at time of writing) v4.0 release focusing on alignment with Pangeo standards and `Xarray` data structures (Hoyer & Hamman, 2017), `VirtualShip` will also benefit from these improvements, further enhancing its capabilities, extensibility and compatability with modern cloud-based data pipelines.

Acknowledgements

The VirtualShip project is funded through the Utrecht University-NIOZ (Royal Netherlands Institute for Sea Research) collaboration.

References

Abernathey, R. P., Augspurger, T., Banihirwe, A., Blackmon-Luca, C. C., Crone, T. J., Gentemann, C. L., Hamman, J. J., Henderson, N., Lepore, C., McCaie, T. A., Robinson, N. H., & Signell, R. P. (2021). Cloud-native repositories for big scientific data. *Computing in Science & Engineering*, *23*(2), 26–35. https://doi.org/10.1109/MCSE.2021.3059437

Almansi, M., Gelderloos, R., Haine, T. W. n., Saberi, A., & Siddiqui, A. H. (2019). OceanSpy: A python package to facilitate ocean model data analysis and visualization. *Journal of Open Source Software*, *4*(39), 1506. https://doi.org/10.21105/joss.01506

Daniels, E., Chytas, C., & Sebille, E. van. (2025). The virtual ship classroom: Developing virtual fieldwork as an authentic learning environment for physical oceanography. *Current: The Journal of Marine Education*. https://doi.org/10.5334/cjme.121

Delandmeter, P., & van Sebille, E. (2019). The Parcels v2.0 Lagrangian framework: new field interpolation schemes. *Geoscientific Model Development*, *12*(8), 3571–3584. https://doi.org/10.5194/gmd-12-3571-2019

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

Lange, M., & van Sebille, E. (2017). Parcels v0.9: prototyping a Lagrangian ocean analysis framework for the petascale age. *Geoscientific Model Development*, *10*(11), 4175–4186. https://doi.org/10.5194/gmd-10-4175-2017

Maze, G., & Balem, K. (2023). *Virtual fleet - recovery* (Version v0.1). Zenodo. https://doi.org/10.5281/zenodo.7520147

Stern, C., Abernathey, R., Hamman, J., Wegener, R., Lepore, C., Harkins, S., & Merose, A. (2022). Pangeo forge: Crowdsourcing analysis-ready, cloud optimized data production. *Frontiers in Climate*, *Volume 3 - 2021*. https://doi.org/10.3389/fclim.2021.782909