The Journal of Open Source Software

CSET: Toolkit for evaluation of weather and climate models

James Frost [1], **James Warner**[1], **Sylvia Bohnenstengel**[1], **David Flack**[1], **Huw Lewis**[1], **Dasha Shchepanovska**[1], **Jon Shonk**[1], **Bernard Claxton**[1], **Jorge Bornemann**[2], **Carol Halliwell**[1], **Magdalena Gruziel**[3], **Pluto ???**[4], **and John M Edwards**[1]

1 Met Office, United Kingdom **2** National Institute of Water and Atmospheric Research, New Zealand **3** Interdisciplinary Centre for Mathematical and Computational Modelling, Poland **4** Centre for Climate Research Singapore, Meteorological Service Singapore, Singapore

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

The *Convective- [and turbulence-] Scale Evaluation Toolkit* (**CSET**) is a community-driven open source library, command line tool, and workflow designed to support the evaluation of weather and climate models at convective and turbulent scales. Developed by the Met Office in collaboration with the Momentum® Partnership and broader research community, CSET provides a reproducible, modular, and extensible framework for model diagnostics and verification. It analyses numerical weather prediction (NWP) and climate modelling output, including from the next-generation LFRic model (Adams et al., 2019), ML models, and observational data and visualises the output in an easily sharable static website to allow the development of a coherent evaluation story for weather and climate models across time and spatial scales.

Statement of need

Evaluation is essential for the model development process in atmospheric sciences. Typically, an evaluation includes both context and justification to demonstrate the benefit of model changes against other models or previous model versions. The verification provides the context or baseline for understanding the model's performance through comparison against observation. The evaluation then demonstrates the benefit through comparison against theoretical expectations or previous or different version of the model and other models for similar application areas using diagnostics derived from model output to explain the context.

Contribution to the field

CSET addresses the need for an evaluation system that supports consistent and comparable evaluation. It gives users easy access to a wide selection of peer-reviewed diagnostics, including spatial plots, time series, vertical profiles, probability density functions, and aggregated analysis over multiple model simulations, replacing bespoke evaluation scripts. To cater for the full evaluation process, CSET provides a range of verification diagnostics to compare against observations and derived diagnostics based on model output, allowing for both physical process-based and impact-based understanding.

The Journal of Open Source Software

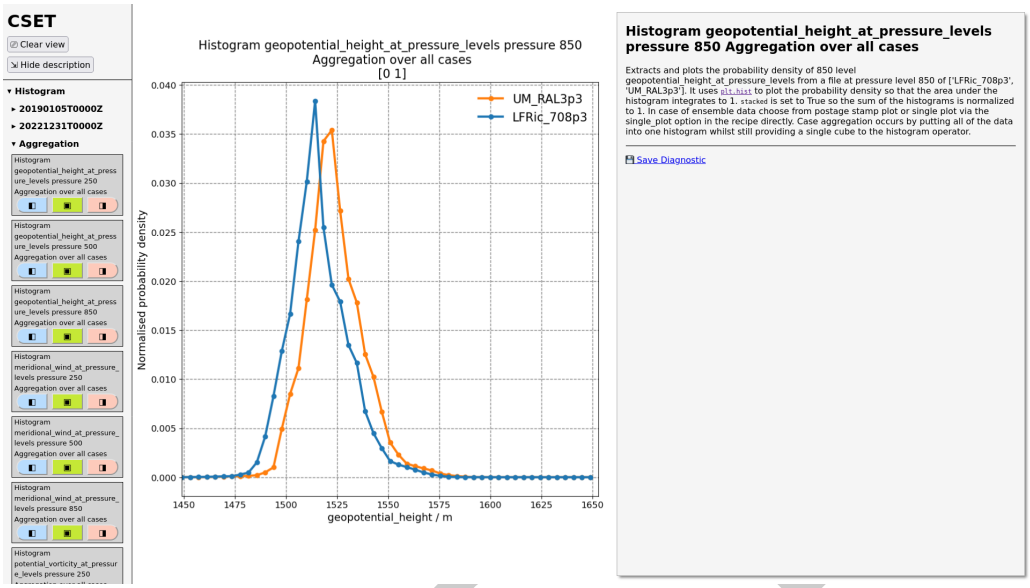

Figure 1: The website produced by CSET. The left column allows for navigating and selecting the displayed diagnostic(s). The main region of the interface displays a particular diagnostic, and documentation to aid interpretation.

The verification side of CSET utilises the Model Evaluation Tools (METplus) verification system (Prestopnik et al., 2025) to provide a range of verification metrics that are aligned with operational verification best practices. The justification side of CSET consists of a range of diagnostics derived from model output. These derived diagnostics include process-based diagnostics for specific atmospheric phenomena and impact-based diagnostics that can be used to understand how model changes will affect customers.

Compared to alternative open source evaluation tools, such as ESMValTool (Andela et al., 2025), CSET is more focused on weather-relevant time scales and evaluating models towards a goal of operational usage.

Design

CSET is build using operators, recipes and a workflow:

- **Operators** are small python functions performing a single task, such as reading, writing, filtering, executing a calculation, stratifying, or plotting.
- **Recipes** are YAML files that compose operators together to produce diagnostics, such as a wind speed difference plot between two model configurations.
- The **Workflow** runs the recipes across a larger number of models, variables, model domains and dates, collating the result into a website.

The design provides a flexible software that is easily adaptable by scientists to address model evaluation questions while maintaining traceability.

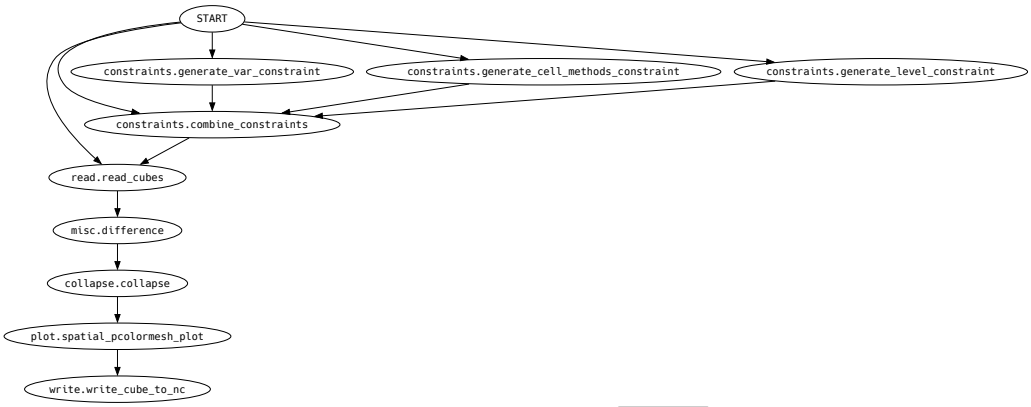

Figure 2: Graph view of a wind speed difference recipe, as produced by `cset graph`. Each node represents an operator, with the arrows showing the flow of data.

The recipes and operators within CSET are well-documented, tested, and peer reviewed, increasing discoverability and giving confidence to users. The documentation covers information on the applicability and interpretation of diagnostics, ensuring they are appropriately used.

CSET has been built with portability in mind. It can run on a range of platforms, from laptops to supercomputers, and can be easily installed from conda-forge. It is built on a modern software stack that is underpinned by Cylc (a workflow engine for complex computational tasks) (Oliver et al., 2018), Python 3, and Iris (a Python library for meteorological data analysis) (Iris contributors, 2025). CSET is open source under the Apache-2.0 licence, and actively developed on GitHub, with extensive automatic unit and integration testing. It aims to be a community-based toolkit, thus contributing to CSET is made easy and actively encouraged with clear developer guidelines to help.

Research usage

Recently, CSET has been the tool of choice in the development and evaluation of the Regional Atmosphere Land Configuration RAL3-LFRic in the Met Office and across the Momentum® Partnership (a cooperative partnership of institutions sharing a seamless modelling framework for weather and climate science and services), as part of the Met Office's Next Generation Modelling System (NGMS) programme to transition from the Unified Model to LFRic. It has helped us to characterise the regional configuration and lead to improvements in our model.

Conclusion

CSET shows the benefits of open source evaluation software. It reduces redundant evaluation diagnostics development and supports easier collaboration across organisations involved in atmospheric model evaluation, helping to build a clear and consistent understanding of model characteristics and model improvement benefits. Major items on CSET's development roadmap are integrating METplus verification into the workflow, and increasing the number of supported observation sources.

The CSET documentation is hosted at https://metoffice.github.io/CSET

Acknowledgements

We acknowledge contributions and support from the Met Office and Momentum® Partnership for this project.

References

Adams, S. V., Ford, R. W., Hambley, M., Hobson, J. M., Kavčič, I., Maynard, C. M., Melvin, T., Müller, E. H., Mullerworth, S., Porter, A. R., Rezny, M., Shipway, B. J., & Wong, R. (2019). LFRic: Meeting the challenges of scalability and performance portability in weather and climate models. *Journal of Parallel and Distributed Computing*, *132*, 383–396. https://doi.org/https://doi.org/10.1016/j.jpdc.2019.02.007

Andela, B., Broetz, B., Mora, L. de, Drost, N., Eyring, V., Koldunov, N., Lauer, A., Mueller, B., Predoi, V., Righi, M., Schlund, M., Vegas-Regidor, J., Zimmermann, K., Adeniyi, K., Arnone, E., Bellprat, O., Berg, P., Billows, C., Blockley, E., … Chun, F. (2025). *ESMValTool* (Version v2.13.0). https://doi.org/10.5281/zenodo.3401363

Iris contributors. (2025). *Iris*. https://doi.org/10.5281/zenodo.595182

Oliver, H. J., Shin, M., & Sanders, O. (2018). Cylc: A workflow engine for cycling systems. *Journal of Open Source Software*, *3*(27), 737. https://doi.org/10.21105/joss.00737

Prestopnik, J., Opatz, J., Gotway, J. H., Jensen, T., Vigh, J., Row, M., Kalb, C., Fisher, H., Goodrich, L., Adriaansen, D., Win-Gildenmeister, M., McCabe, G., Frimel, J., Blank, L., & Arbetter, T. (2025). *The METplus version 6.1.0 user's guide*. Developmental Testbed Center. https://github.com/dtcenter/METplus/releases.