Community Analysis Pipeline: A Python package for processing Mars climate model data

Courtney M. L. Batterson [1*], **Richard A. Urata** [1*¶], **Victoria L. Hartwick** [3*], **Alexandre M. Kling** [1*], and **Melinda A. Kahre** [2*]

1 Bay Area Environmental Research Institute, United States **2** NASA Ames Research Center, United States **3** Southwest Research Institute, United States ¶ Corresponding author * These authors contributed equally.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 19 May 2025
Published: unpublished

Summary

The Community Analysis Pipeline (CAP) is a Python package designed to streamline and simplify the complex process of analyzing the large datasets output by global climate models (GCMs). CAP consists of a suite of tools that manipulate NetCDF files to produce secondary datasets and figures useful for science and engineering. CAP facilitates inter-model and model-to-observation comparison, and it is the first software of its kind to standardize these comparisons. The goal is to enable users of varying levels of programming experience to work more easily with complex data products from GCMs, thereby lowering the barrier to entry into planetary science research.

Statement of need

GCMs perform numerical simulations that describe the evolution of climate systems on planetary bodies. GCMs simulate physical processes within the atmosphere (and, if applicable, within the surface of the planet, ocean, and any interactions between them), calculate radiative transfer within those mediums, and predict the transport of heat and momentum within the atmosphere using a computational fluid dynamics (CFD) solver (the "dynamical core"). GCM data products routinely include surface and atmospheric variables such as wind, temperature, and aerosol concentrations. While GCMs have been applied to planetary bodies in our Solar System (e.g., Earth, Venus, Pluto) and in other stellar systems (e.g., Hartwick et al. (2023)), CAP is extends compatibility to Mars GCMs (MGCMs). Several MGCMs are actively in use and under development in the Mars community, including the NASA Ames MGCM (Legacy and FV3-based versions), NASA Goddard ROCKE-3D, the Laboratoire de Météorologie Dynamique (LMD) Mars Planetary Climate Model (PCM), the Open University OpenMars, NCAR MarsWRF, NCAR MarsCAM, GFDL Mars GCM, Harvard DRAMATIC Mars GCM, Max Planck Institute Mars GCM, and GEM-Mars. To date, CAP is compatible with four of these models (the NASA Ames MGCM, PCM, OpenMars, and MarsWRF) and undergoing development to further extend compatibility to other models.

GCM output is complex in both size and structure, and analyzing the data requires GCM-specific domain knowledge. We highlight the following major challenges for working with MGCM output:

- Files are complex in structure with output fields represented by multiple variables (e.g., air and surface temperature) with varying units (e.g., Kelvin, Celcius) in multi-dimensional structures (e.g., 2–5 dimensions) at a variety of sampling frequencies (e.g., temporally averaged, instantaneous) and on custom horizontal and vertical grids.

- File sizes range from ~10 Gb–10 Tb for simulations describing the Martian climate over a full orbit around the Sun, and vary significantly depending on the number of atmospheric fields being analyzed, time sampling, and horizontal and vertical resolutions. Large files require curated processing pipelines to manage memory storage, which can be particularly challenging for users that do not have access to academic or enterprise clusters or supercomputers for their analyses.
- Domain-specific knowledge is required to derive secondary variables, manipulate complex data structures, and visualize results. Such information is not always publicly available online and many MGCMs do not output data into self-describing formats like netCDF. Working with MGCM data is especially difficult for users unfamiliar with the fields commonly output by MGCMs or the mathematical methods used in climate science.

CAP offers a streamlined workflow for processing and analyzing MGCM data products by providing a set of libraries and executables that facilitate file manipulation and data visualization from the command line. This benefits existing modelers by automating both routine and sophisticated post-processing tasks. It also expands access to MGCM products by removing some of the technical roadblocks associated with processing these complex data products.

State of the Field

CAP is the first software package to provide data visualization, file manipulation, variable derivation, and inter-model or model-to-observation comparison tools for MGCM output in one software suite. Some existing tools perform a subset of these functions, but none of them provide both complex data analysis and visualization tools and they are not designed specifically for climate modeling. Some of the more popular tools include Panoply (Schmunk, 2024), Ncview (Pierce, 2024), Grid Analysis and Display System (GrADS; George Mason University (n.d.)), and Paraview (Kitware Inc., 2023). Each tool offers simple solutions for visualizing NetCDF data. Some provide minimal flexibility for user-defined computations. However, existing tools are either Java or C based, which preclude the use of powerful analysis and visualization packages available in Python such as NumPy and Matplotlib. CAP is the only software package with an open-source Python framework for analyzing and plotting climate model data and performing inter-model or model-to-observation comparison.

Software Design

CAP was originally designed to be an accessible toolkit for manipulating, analyzing, and visualizing data from the NASA Ames MGCM. It is written in Python for its approachability (for developers and users), open-source-friendly design, and integration of third party packages (i.e., NumPy, NetCDF4, Matplotlib). CAP is easily portable to multiple operating systems, which is a key feature of the software as the model data often moves between systems, whether from a cluster to local storage or between users who have differing operating systems. The use of Python over other languages (e.g., C/C++, or Fortran) means CAP sacrifices some computational efficiency for the many benefits of having all of the analysis and visualization tools in a single, open-source package.

CAP tools are grouped by functionality for an intuitive experience. For example, file manipulation functions (like splitting a dataset by date or slicing by latitude) are stored in MarsFiles, functions that add diagnostic variables are in MarsVars, and plotting is handled by MarsPlot. The main executables (MarsFiles, MarsVars, MarsPlot, MarsInterp, MarsFormat, MarsPull, MarsCalendar) contain readable instructions and function definitions, many of which are accessible via the command line with the `-h` flag, while the more complicated backend functions are stored in separate files (e.g., FV3_utils.py, Ncdf_wrapper.py).

Research Impact Statement

CAP has been used in multiple research projects that have been published and/or presented at conferences worldwide (e.g., Urata et al. (2025); Batterson et al. (2023); Hartwick, Haberle, et al. (2022); Hartwick, Toon, et al. (2022); K. Steakley et al. (2023); K. E. Steakley et al. (2024); Kahre et al. (2022); Kahre et al. (2023); Nagata et al. (2025); Hartwick & Kahre (2024)).

AI Use Disclosure

Claude (Sonnet 4.5) was used to assist with software execution error analysis and for the development of the automated testing suite via GitHub Actions. Quality and correctness of AI-generated content was verified by comparing results to expected output, and was tested by multiple team members.

Acknowledgements

This work is supported by the Planetary Science Division of the National Aeronautics and Space Administration as part of the Mars Climate Modeling Center funded by the Internal Scientist Funding Model.

References

Batterson, C. M. L., Kahre, M. A., Bridger, A. F. C., Wilson, R. J., Urata, R. A., & Bertrand, T. (2023). Modeling the B regional dust storm on Mars: Dust lofting mechanisms predicted by the new NASA Ames Mars GCM. *Icarus*, *400*, 115542–115542. https://doi.org/10.1016/j.icarus.2023.115542

George Mason University. (n.d.). *GrADS home page*. COLA GMU. http://cola.gmu.edu/grads/

Hartwick, V. L., Haberle, R. M., & Kahre, M. A. (2023). Stabilization of tenuous atmospheres by atmospheric dust in the HZ of M-dwarf stars. *American Geophysical Union, Fall Meeting*.

Hartwick, V. L., Haberle, R. M., Kahre, M. A., & Wilson, R. J. (2022). The dust cycle on Mars at different orbital distances from the sun: An investigation of impact of radiatively active dust on land planet climate. *The Astrophysical Journal*, *941*(1). https://doi.org/10.3847/1538-4357/ac9481

Hartwick, V. L., & Kahre, M. A. (2024). Modeling the impact of atmospheric dust on the atmospheric stability & climate of tidally locked Mars-like exoplanets with the NASA Ames FV3-based Mars global climate model. *10th International Conference on Mars*.

Hartwick, V. L., Toon, O. B., Kahre, M. A., Pierpaoli, O., & Lunduist, J. M. (2022). Assessment of wind energy resources for future manned missions to Mars. *Nature Astronomy*. https://doi.org/10.1038/s41550-022-01851-4

Kahre, M. A., Wilson, R. J., Brecht, A. S., Haberle, R. M., Harman, S., Urata, R., Kling, A., Steakley, K. E., Batterson, C. M., Hartwick, V., & Gkouvelis, L. (2022). Update and status of the Mars climate modeling center at NASA Ames Research Center. *7th Workshop on Mars Atmosphere Modelling and Observations*.

Kahre, M. A., Wilson, R. J., Urata, R. A., Batterson, C. M., Kling, A., Brecht, A. S., Steakley, K., Hartwick, V., & Harman, C. E. (2023). The NASA Ames Mars Global Climate Model: Benchmarking Publicly Released Source Code and Model Output. *AGU Fall Meeting Abstracts*, *2023*, P51E–2741.

Kitware Inc. (2023). *Core feature of ParaView*. ParaView. https://www.paraview.org/core/

Nagata, N., Liu, H., Nakagawa, H., Jain, S., Rafkin, S., & Hartwick, V. (2025). Impacts of 2018 global dust storm on atmosphere-ionosphere coupling on Mars: With MAVEN & NASA Ames Mars GCM. *Japan Geoscience Union Meeting*.

Pierce, D. W. (2024). *Ncview*. David W. Pierce, Scripps Institution of Oceanography. https://cirrus.ucsd.edu/ncview/

Schmunk, R. B. (2024). *NASA GISS: Panoply 5 netCDF, HDF and GRIB data viewer*. NASA Goddard Institute for Space Studies. https://www.giss.nasa.gov/tools/panoply/

Steakley, K. E., Kahre, M. A., Haberle, R. M., & Lee, R. (2024). Sensitivities of the Water Cycle in a Post-Impact H2-Rich Early Mars Environment. *LPI Contributions*, *3007*, 3442.

Steakley, K., Hartwick, V., Kahre, M. A., & Haberle, R. M. (2023). Cloud Condensation Nuclei in the Early Martian Atmosphere. *AGU Fall Meeting Abstracts*, *2023*, P41D–07.

Urata, R., Bertrand, T., Kahre, M., Wilson, R., Kling, A., & Wolff, M. (2025). Impact of a bimodal dust distribution on the 2018 Martian global dust storm with the NASA Ames Mars global climate model. *Icarus*, *429*, 116446. https://doi.org/10.1016/j.icarus.2024.116446