The Journal of Open Source Software

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

Courtney M. L. Batterson [1*], **Richard A. Urata** [1*¶], **Alexandre M. Kling** [4], **Victoria L. Hartwick** [3], **and Melinda A. Kahre** [2]

1 Bay Area Environmental Research Institute (BAERI), United States **2** NASA Ames Research Center, United States **3** Southwest Research Institute (SwRI), United States **4** Astera Institute, United States ¶ Corresponding author * These authors contributed equally.

DOI: 10.21105/joss.11057

Software

- Review
- Repository
- Archive

Editor: Arfon Smith

Reviewers:

- @MansurJisan-NCO
- @gampnico
- @shenyulu
- @fzhao70

Submitted: 19 May 2025
Published: 20 November 2026

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

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 command-line tools that manipulate netCDF files to produce secondary datasets and figures useful for science and engineering. By converting output from several Mars GCMs and reanalyses to a common format, CAP facilitates inter-model and model-to-observation comparison for Mars. 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. 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, Haberle, et al., 2023), CAP focuses on 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, the latter built on the GFDL AM4 framework (Zhao et al., 2018)), NASA Goddard ROCKE-3D (Way et al., 2017), the Laboratoire de Météorologie Dynamique (LMD) Mars Planetary Climate Model (PCM) (Forget et al., 1999), MarsWRF (Newman et al., 2019; Richardson et al., 2007), the Tohoku University DRAMATIC Mars GCM (Kobayashi et al., 2026; Kuroda et al., 2005), the Max Planck Institute Mars GCM (Hartogh et al., 2005), and GEM-Mars (Neary & Daerden, 2018), as well as the OpenMARS (Holmes et al., 2020) and EMARS (Greybush et al., 2019) reanalyses. CAP was originally developed as a suite of tools specific to the NASA Ames MGCM (Bertrand et al., 2020; Haberle et al., 2019; Urata et al., 2023), to standardize everything from plots in Mars conventions (e.g., solar longitude (L_s)) to vertical interpolation and variable derivations, which depend on Mars constants, such as gravity and the specific heat of the atmosphere (C_p), that can vary among individual analysis routines. Such tools are a need shared across the Mars climate modeling community. After receiving several community requests, we decided to add compatibility with other MGCMs; CAP reads output from the NASA Ames MGCM directly and, through its `MarsFormat` tool, converts output from PCM, MarsWRF, OpenMARS, and EMARS to Ames MGCM-like output files for compatibility with CAP (see the MarsFormat documentation at https://amescap.readthedocs.io). Model-specific variable and dimension names are defined in a user-editable variable dictionary, so additional models can be supported without changes to

the rest of the pipeline.

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, degrees Celsius) 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.
- Output from a simulation of one Mars year ranges from ~10 GB to ~10 TB, depending on the fields saved, time sampling, and resolution. Such files require memory-aware processing, which is particularly challenging for users without access to clusters or supercomputers.
- Domain-specific knowledge is required to derive secondary variables, manipulate complex data structures, and visualize results. Such information is not always publicly available, and many MGCMs do not output data in self-describing formats like netCDF.

CAP provides libraries and command-line executables for file manipulation, analysis, and visualization. It automates routine and sophisticated post-processing for experienced modelers and removes technical roadblocks for new users of MGCM data.

State of the Field

General-purpose netCDF viewers such as Panoply (Schmunk, 2024), Ncview (Pierce, 2024), the Grid Analysis and Display System (GrADS; Center for Ocean-Land-Atmosphere Studies, 2026), and ParaView (Kitware Inc., 2023) offer quick visualization of gridded data, and some (e.g., GrADS expressions and ParaView's Python interface) support user-defined computations. For Earth science, Python packages such as xarray (Hoyer & Hamman, 2017), Iris (Iris contributors, n.d.), GeoCAT (Visualization & Analysis Systems Technologies, 2020), and ESMValTool (Righi et al., 2020) provide labeled-array analysis, model evaluation, and plotting, while planetary tools such as planetoplot (Spiga, n.d.), aeolus (Sergeev & Zamyatina, 2021), and wrf-python (Ladwig, 2017) address plotting or diagnostics for particular models. CAP builds on this ecosystem: it uses NumPy (Harris et al., 2020), Matplotlib (Hunter, 2007), netCDF4 (Unidata, 2024), and xarray rather than reimplementing array storage, file input/output, or plotting, and it adapts the wrf-python approach to destaggering MarsWRF winds.

What CAP adds is a Mars-specific layer that these packages do not provide as a single workflow: conversion of several MGCM and reanalysis formats to one convention; Mars time handling (sols, L_s, and local true solar time); derived variables that need MGCM vertical coordinates and Mars constants (e.g., mass streamfunction, column-integrated and vertically differentiated quantities); vertical interpolation from hybrid sigma-pressure levels to standard pressure, altitude, or height above ground; and template-driven plotting, all behind one command-line interface. CAP is a complete toolkit for Mars GCMs, providing Mars-specific plotting and vertical interpolation as well as derived variables computed with Mars constants, and addresses a longstanding need for non-Earth analysis codebases in the planetary GCM community. We built CAP as a separate package, rather than extending one of the packages above, because the Earth-oriented packages assume Earth calendars, constants, and model conventions (e.g., CF-compliant time coordinates and Earth model output standards); instead, CAP builds on their array, file input/output, and plotting layers and adds the Mars-specific conventions on top.

Software Design

CAP is written in Python for its approachability (for developers and users), open-source-friendly design, and integration of third-party packages (NumPy, netCDF4, xarray, Matplotlib). CAP is portable to Linux, macOS, and Windows, which matters because model data often moves between systems, whether from a supercomputer to local storage or between collaborators. A compiled language such as C++ could also combine analysis and visualization in one package (e.g., with VTK for visualization and Eigen or xtensor for array computation). However, we chose Python because the intended users are planetary scientists and students, many of whom analyze data in Python. Python enables reading, modifying, and extending CAP's functions and plotting templates without a compilation step, and the scientific Python stack already provides netCDF input/output and publication-quality plotting. The cost is lower computational efficiency, which CAP mitigates by vectorizing calculations with NumPy and by processing large files in chunks (e.g., `MarsInterp` interpolates in time chunks and `MarsFormat` can use Dask (Dask Development Team, 2016) for lazy, chunked processing).

Several design choices follow from the structure of Mars model output:

- **Time.** The netCDF Climate and Forecast (CF) conventions define time relative to Earth calendars. Mars days (sols) are about 2.7% longer than Earth days, and a Mars year has about 668.6 sols, so no CF calendar represents Mars time. CAP therefore stores time as sols since the start of the simulation and keeps L_s in a separate `areo` variable. `MarsCalendar` converts between sols and L_s, and `MarsFormat` computes L_s for converted files that lack it. Diurnal ("diurn") files store each hour of the day in universal time, so the local time differs with longitude; `time_shift_calc` in `FV3_utils.py` shifts these fields to uniform local solar time, including an equation-of-time correction when the model provides it, which `MarsFiles` exposes through its `-t` option.
- **Common format.** All tools operate on files in the NASA Ames FV3-based MGCM layout. `MarsFormat` maps other models onto this layout, so that `MarsFiles`, `MarsVars`, `MarsInterp`, and `MarsPlot` need not handle each model separately.
- **Executables and libraries.** Each executable in `bin/` (`MarsPull`, `MarsFiles`, `MarsVars`, `MarsInterp`, `MarsPlot`, `MarsFormat`, `MarsCalendar`, `MarsNest`) is a script that parses its arguments, processes one file at a time, and writes a new file, while shared numerical routines (e.g., vertical coordinates, interpolation, and time shifting in `FV3_utils.py`, and spectral filtering in `Spectral_utils.py`, which optionally uses SHTools (Wieczorek & Meschede, 2018)) are plain functions that can be imported into user scripts. Executables use module-level constants and settings, for example the planetary constants in `MarsVars`, which are reset for each input file from the file's metadata. This keeps each executable readable as a linear script, at the cost of making those routines harder to reuse than the library functions.
- **Writing files.** Reading is functional, using netCDF4 and xarray directly, while writing is object-oriented: the `Ncdf` class in `Ncdf_wrapper.py` wraps a netCDF4 dataset and provides methods to copy dimensions and variables, log new variables with their metadata, and copy large variables in chunks. Tools such as `MarsFiles` (e.g., when concatenating or splitting files) and `MarsInterp` build their output through `Ncdf`, so metadata handling is shared, while the processing itself remains in functions. The related `Fort` class converts binary output from the Legacy NASA Ames MGCM to the same netCDF layout.

CAP is tested with unit tests of numerical routines and integration tests of each executable on synthetic files that mimic each supported model, run on Linux, macOS, and Windows with GitHub Actions.

Research Impact Statement

CAP is developed by the NASA Ames Mars Climate Modeling Center alongside the NASA Ames MGCM. It was the analysis tool taught at the NASA Ames MGCM and CAP Tutorials held in 2021 and 2023. Tutorial exercises are part of CAP's documentation and can now be completed either with the publicly released NASA Ames MGCM simulations, which `MarsPull` downloads from the NASA Advanced Supercomputing Data Portal, or with synthetic files generated by CAP's test suite.

CAP is also used outside the development team. Scientists at Aeolis Research on the MarsWRF team requested compatibility with MarsWRF output directly and contributed code to CAP, adding support for current planetWRF output, native staggered winds, true local solar time, and faster chunked interpolation and conversion. CAP users outside NASA Ames also include researchers at Yale University, the Southwest Research Institute, and San José State University.

Within the Mars Climate Modeling Center, CAP has been used to analyze NASA Ames MGCM simulations. In Urata et al. (2025), CAP shifted output to uniform local time, interpolated it to a standard altitude grid, and was used for the migrating tide analysis; in Batterson et al. (2023), CAP computed the mass streamfunction and derived vertical winds and sedimentation rates. CAP also produced preliminary analysis figures for several studies (Batterson et al., 2023; Hartwick et al., 2022; Hartwick, Toon, et al., 2023; Urata et al., 2025) and figures for conference presentations (Hartwick & Kahre, 2024; Kahre et al., 2022, 2023; Steakley et al., 2023, 2024).

Because CAP has not previously had a DOI or a peer-reviewed description, studies that used it could not cite it formally, and it does not appear in their reference lists. This paper, together with an archived, versioned release, gives CAP a citable reference for future work.

AI Use Disclosure

Claude Sonnet 4.5 (Anthropic) was used to assist with software execution error analysis and for the development of the automated testing suite via GitHub Actions. During the review of this paper, Claude Opus 5.5, via Claude Code, was used to help implement reviewer-requested changes, including the MarsPull download fixes, compact test fixtures, regression tests, and documentation and manuscript revisions. Quality and correctness of AI-generated content were verified by comparing results to expected output, by running the automated test suite, and by manual review from Richard Urata and Courtney Batterson.

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. https://doi.org/10.1016/j.icarus.2023.115542

Bertrand, T., Wilson, R. J., Kahre, M. A., Urata, R., & Kling, A. (2020). Simulation of the 2018 Global Dust Storm on Mars Using the NASA Ames Mars GCM: A Multitracer Approach. *Journal of Geophysical Research: Planets*, *125*(7), e2019JE006122. https://doi.org/10.1029/2019JE006122

Center for Ocean-Land-Atmosphere Studies. (2026). *GrADS: The grid analysis and display system*. George Mason University. https://github.com/j-m-adams/GrADS

Dask Development Team. (2016). *Dask: Library for dynamic task scheduling*. https://dask.org

Forget, F., Hourdin, F., Fournier, R., Hourdin, C., Talagrand, O., Collins, M., Lewis, S. R., Read, P. L., & Huot, J.-P. (1999). Improved general circulation models of the Martian atmosphere from the surface to above 80 km. *Journal of Geophysical Research: Planets*, *104*(E10), 24155–24175. https://doi.org/10.1029/1999JE001025

Greybush, S. J., Kalnay, E., Wilson, R. J., Hoffman, R. N., Nehrkorn, T., Leidner, M., Eluszkiewicz, J., Gillespie, H. E., Wespetal, M., Zhao, Y., Hoffman, M., Dudas, P., McConnochie, T., Kleinböhl, A., Kass, D., McCleese, D., & Miyoshi, T. (2019). The Ensemble Mars Atmosphere Reanalysis System (EMARS) version 1.0. *Geoscience Data Journal*, *6*(2), 137–150. https://doi.org/10.1002/gdj3.77

Haberle, R. M., Kahre, M. A., Hollingsworth, J. L., Montmessin, F., Wilson, R. J., Urata, R. A., Brecht, A. S., Wolff, M. J., Kling, A. M., & Schaeffer, J. R. (2019). Documentation of the NASA/Ames Legacy Mars Global Climate Model: Simulations of the present seasonal water cycle. *Icarus*, *333*, 130–164. https://doi.org/10.1016/j.icarus.2019.03.026

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hartogh, P., Medvedev, A. S., Kuroda, T., Saito, R., Villanueva, G., & others. (2005). Description and climatology of a new general circulation model of the Martian atmosphere. *Journal of Geophysical Research: Planets*, *110*(E11), E11008. https://doi.org/10.1029/2005JE002498

Hartwick, V. L., Haberle, R. M., & Kahre, M. A. (2023). Stabilization of tenuous atmospheres by atmospheric dust in the HZ of M-dwarf stars. *AGU Fall Meeting Abstracts*, P23A–06.

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 the impact of radiatively active dust on land planet climate. *The Astrophysical Journal*, *941*(1), 54. 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. *Tenth International Conference on Mars*, *3007*.

Hartwick, V. L., Toon, O. B., Lundquist, J. K., Pierpaoli, O. A., & Kahre, M. A. (2023). Assessment of wind energy resource potential for future human missions to Mars. *Nature Astronomy*, *7*(3), 298–308. https://doi.org/10.1038/s41550-022-01851-4

Holmes, J. A., Lewis, S. R., & Patel, M. R. (2020). OpenMARS: A global record of martian weather from 1999 to 2015. *Planetary and Space Science*, *188*, 104962. https://doi.org/10.1016/j.pss.2020.104962

Hoyer, S., & Hamman, J. (2017). xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, *5*(1), 10. https://doi.org/10.5334/jors.148

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Iris contributors. (n.d.). *Iris*. Zenodo. https://doi.org/10.5281/zenodo.595182

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, June). 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, December). The NASA Ames Mars Global Climate Model: Benchmarking Publicly Released Source Code and Model Output. *AGU Fall Meeting Abstracts*.

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

Kobayashi, M., Kuroda, T., Forget, F., Kamada, A., Kurokawa, H., Aoki, S., Kazama, A., Nakagawa, H., & Terada, N. (2026). The roles of slope winds, regolith, and clouds in shaping the Martian water vapor column unveiled with Mars global climate model simulations. *Journal of Geophysical Research: Planets*, *131*(7), e2026JE009758. https://doi.org/10.1029/2026JE009758

Kuroda, T., Hashimoto, N., Sakai, D., & Takahashi, M. (2005). Simulation of the Martian atmosphere using a CCSR/NIES AGCM. *Journal of the Meteorological Society of Japan. Ser. II*, *83*(1), 1–19. https://doi.org/10.2151/jmsj.83.1

Ladwig, W. (2017). *wrf-python*. NSF National Center for Atmospheric Research. https://doi.org/10.5065/D6W094P1

Neary, L., & Daerden, F. (2018). The GEM-Mars general circulation model for Mars: Description and evaluation. *Icarus*, *300*, 458–476. https://doi.org/10.1016/j.icarus.2017.09.028

Newman, C. E., Kahanpää, H., Richardson, M. I., Martínez, G. M., Vicente-Retortillo, A., & Lemmon, M. T. (2019). MarsWRF convective vortex and dust devil predictions for Gale Crater over 3 Mars years and comparison with MSL-REMS observations. *Journal of Geophysical Research: Planets*, *124*, 3442–3468. https://doi.org/10.1029/2019JE006082

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

Richardson, M. I., Toigo, A. D., & Newman, C. E. (2007). PlanetWRF: A general purpose, local to global numerical model for planetary atmospheric and climate dynamics. *Journal of Geophysical Research: Planets*, *112*(E9), E09001. https://doi.org/10.1029/2006JE002825

Righi, M., Andela, B., Eyring, V., Lauer, A., & others. (2020). Earth system model evaluation tool (ESMValTool) v2.0 – technical overview. *Geoscientific Model Development*, *13*(3), 1179–1199. https://doi.org/10.5194/gmd-13-1179-2020

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/

Sergeev, D. E., & Zamyatina, M. (2021). *Aeolus: A Python library for the analysis and visualisation of climate model output*. Zenodo. https://doi.org/10.5281/zenodo.5145604

Spiga, A. (n.d.). *planetoplot*. https://github.com/aymeric-spiga/planetoplot

Steakley, K. E., Hartwick, V., Kahre, M. A., & Haberle, R. M. (2023). Cloud condensation nuclei in the early Martian atmosphere. *AGU Fall Meeting Abstracts*, *2023*, P41D–07. https://ui.adsabs.harvard.edu/abs/2023AGUFM.P41D..07S

Steakley, K. E., Kahre, M. A., Haberle, R. M., & Lee, R. (2024). Sensitivities of the water cycle in a post-impact H_2-rich early Mars environment. *Tenth International Conference on Mars*, *3007*, 3442. https://ui.adsabs.harvard.edu/abs/2024LPICo3007.3442S

Unidata. (2024). *netCDF4*. UCAR/Unidata Program Center. https://doi.org/10.5065/D6H70CW6

Urata, R. A., Bertrand, T., Kahre, M. A., Wilson, R. J., Kling, A. M., & Wolff, M. J. (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

Urata, R. A., Kahre, M., Wilson, J., Brecht, A., Harman, S., Steakley, K., Kling, A., & Batterson, C. (2023). *NASA Ames Mars Global Climate Model* (Version 3.1). National Aeronautics and Space Administration. https://doi.org/10.25966/ty4s-w176

Visualization & Analysis Systems Technologies. (2020). *Geoscience community analysis toolkit (GeoCAT)*. NSF NCAR Computational and Information Systems Laboratory. https://doi.org/10.5065/A8PP-4358

Way, M. J., Aleinov, I., Amundsen, D. S., Chandler, M. A., Clune, T. L., & others. (2017). Resolving orbital and climate keys of Earth and extraterrestrial environments with dynamics (ROCKE-3D) 1.0: A general circulation model for simulating the climates of rocky planets. *The Astrophysical Journal Supplement Series*, *231*(1), 12. https://doi.org/10.3847/1538-4365/aa7a06

Wieczorek, M. A., & Meschede, M. (2018). SHTools: Tools for working with spherical harmonics. *Geochemistry, Geophysics, Geosystems*, *19*, 2574–2592. https://doi.org/10.1029/2018GC007529

Zhao, M., Golaz, J.-C., Held, I. M., Guo, H., Balaji, V., Benson, R., & others. (2018). The GFDL global atmosphere and land model AM4.0/LM4.0: 1. Simulation characteristics with prescribed SSTs. *Journal of Advances in Modeling Earth Systems*, *10*(3), 691–734. https://doi.org/10.1002/2017MS001208