




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

Alexandre M. Kling^{1*}, Courtney M. L. Batterson^{1*}, Richard A. Urata^{1*}, Victoria L. Hartwick^{3*}, and Melinda A. Kahre^{2*}

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

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: 

Submitted: 19 May 2025

Published: unpublished

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 large datasets created by global climate models (GCMs). CAP consists of a suite of tools that manipulate NetCDF files in order to produce secondary datasets and figures useful for science and engineering applications. CAP also facilitates inter-model and model-observation comparisons, and it is the first software of its kind to standardize these comparisons. The goal is to enable users with varying levels of programming experience to work with complex data products from a variety of GCMs and thereby lower the barrier to entry for 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 therein), calculate radiative transfer within those mediums, and use a computational fluid dynamics (CFD) solver (the “dynamical core”) to predict the transport of heat and momentum within the atmosphere. Typical GCM products 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 currently compatible with 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. Of these, CAP is compatible with four models so far: the NASA Ames MGCM, PCM, OpenMars, and MarsWRF.

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

- Files tend to be fairly complex in structure, with output fields represented by multiple variables (e.g., air vs surface temperature), varying units (e.g., Kelvin), complex dimensional structures (e.g., 2–5 dimensions), and a variety of sampling frequencies (e.g., temporally averaged or instantaneous) on different horizontal and vertical grids.
- File sizes typically range from ~10 Gb–10 Tb for simulations describing the Martian

climate over a full orbit around the Sun (depending on the number of atmospheric fields being analyzed, time sampling, and the horizontal and vertical resolutions of the run). Large files require curated processing pipelines in order to manage memory storage. This 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. 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.

CAP is the first software package to provide data visualization, file manipulation, variable derivation, and inter-model or model-observation comparison features in one software suite. Existing tools perform a subset of the functions that CAP offers, but none of them provide both complex data analysis tools and visualization, nor are they specifically designed 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 and some provide minimal flexibility for user-defined computations. However, CAP is the only software package with an open-source Python framework for analyzing and plotting climate data and performing inter-model and model-observation comparisons.

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)).

Functionality

CAP consists of six command-line executables that can be used sequentially or individually to derive secondary data products, thus offering a high level of flexibility. A configuration text file is provided so that users can define the input file structure (e.g., variable names, longitudinal structure, and interpolation levels) and preferred plotting style (e.g., time axis units) for their analysis. The six executables in CAP are described below:

MarsPull

MarsPull is a data pipeline utility for downloading MGCM data products from the NAS Data Portal (<https://data.nas.nasa.gov/>). Recognizing that each member within the science and engineering community has their own requirements for hosting proprietary Mars climate datasets (e.g., institutional servers, Zenodo, GitHub, etc.), MarsPull is intended to be a mechanism for interfacing those datasets. MarsPull enables users to query data meeting a specific criteria, such as a date range (e.g., solar longitude), which allows users to parse repositories first and download only the necessary data, thus avoiding downloading entire repositories which can be large (»15Gb). A typical application of MarsPull is:

```
MarsPull directory_name -f MGCM_file1.nc MGCM_file2.nc
```

86 MarsFormat

87 MarsFormat is a utility for converting non-NASA Ames MGCM products into NASA Ames-
88 like MGCM products for compatibility with CAP. MarsFormat reorders dimensions, adds
89 standardized coordinates that are expected by other executables for various computations (e.g.,
90 pressure interpolation), converts variable units to conform to the International System of Units
91 (e.g., Pa for pressure), and reorganizes coordinate values as needed (e.g., reversing the vertical
92 pressure array for plotting). Additional, model-specific operations are performed as necessary.
93 For example, MarsWRF data requires un-staggering latitude-longitude grids and calculating
94 absolute fields from perturbation fields. A typical application of MarsFormat is:

```
95 MarsFormat MGCM_file.nc -gcm model_name
```

96 MarsFiles

97 MarsFiles provides several tools for file manipulation such as file size reduction, temporal and
98 spatial filtering, and splitting or concatenating data along specified dimensions. Operations
99 performed by MarsFiles are applied to entire NetCDF files producing new data structures with
100 amended file names. A typical application of MarsFiles is:

```
101 MarsFiles MGCM_file.nc -[flags]
```

102 MarsVars

103 MarsVars performs variable operations such as adding, removing, and editing variables and
104 computing column integrations. It is standard practice within the modeling community to
105 avoid outputting variables that can be derived outside of the MGCM in order to minimize file
106 size. For example, atmospheric density (ρ) is easily derived from temperature and pressure
107 and therefore typically not included in output files. MarsVars derives ρ from temperature and
108 pressure and adds it to the file with a single command line argument. A typical application of
109 MarsVars is:

```
110 MarsVars MGCM_file.nc -add rho
```

111 MarsInterp

112 MarsInterp interpolates the vertical coordinate to a standard grid: pressure, altitude, or altitude
113 above ground level. Vertical grids vary considerably from model to model. Most MGCMs use
114 a pressure or hybrid pressure vertical coordinate (e.g., terrain-following, pure pressure levels, or
115 sigma levels) in which the geometric heights and mid-layer pressures of the atmospheric layers
116 vary in latitude and longitude. It is therefore necessary to interpolate to a standard vertical
117 grid in order to do any rigorous spatial averaging or inter-model or observation-to-model
118 comparisons. A typical application of MarsInterp is:

```
119 MarsInterp MGCM_file.nc -t pstd
```

120 MarsPlot

121 MarsPlot is the plotting utility for CAP. It accepts a modifiable text template containing a list
122 of plots to generate (Custom.in) as input and outputs graphics to PDF or PNG. It supports
123 multiple types of 1-D or 2-D plots, color schemes, map projections, and can customize axes
124 range, plot titles, or contour intervals. It also supports some simple math functions to derive
125 secondary fields not supported by MarsVars. A typical application of MarsPlot is:

```
126 MarsPlot Custom.in
```

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., & Lundquist, 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

DRAFT