MERRA2 Global Forcing data for CESM2 Applications Use Case

Compiled May 2022 by Model Data RCN team

Summary
Weighted rubric score - N/A
Category - Preserve the majority of simulation workflow outputs

- Use Case Description
  - High-level overview of the use case
    - Dataset Description: "Atmospheric Forcing data, regridded from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2) data and are both horizontally and vertically interpolated to the CESM2 standard 32 vertical model levels."
    - MERRA data is originally downloaded from another provider, NASA. MERRA2 is a model itself. It produces a reanalysis dataset that reproduces the actual weather and meteorological phenomenon
    - The team takes the original MERRA2 data, and interpolates vertical and horizontal resolution for input to global models, specifically CESM2.
    - NCAR data repository link
      - https://doi.org/10.5065/H9NM-XC59
  - Science goals and basic workflow
    - The purpose of this data set is to provide an input data set for a CESM2 configuration that has a distinct resolution
      - The community asked for this data. Initially it was a community request from NASA, with modeling efficiency in mind.
    - The process is that the team downloads specific variables from the MERRA2 output. Then the team has a script that does statistical interpolation for input into CESM.
      - First they generate a version that uses the original MERRA2 resolution. Then the script is run to generate another version with the interpolated resolutions.
      - RDA does checks for data consistency and structure. But did very little for this data because it was very clean and well-structured.
    - Why would modelers use this data, and not use the NASA MERRA2 directly?
      - The original MERRA2 data is in single variable files, so it is very tedious to import into models.
    - Main use is for chemistry transport studies, for use by simulations that nudge from meteorological fields
    - There might be other options for people who would want to use the MERRA2 as input into CESM, , e.g. they could provide the script to the community for them to do themselves, or you could do it directly in CEMS. But using this data set is easier.

- What use-case specific additional materials were preserved and shared?
  - Data
    - Inputs
      - The original reanalysis is available from NASA [here](). The original data is much larger than the version provided for CESM2 applications.
    - Raw model output
      - N/A - This project is not generating new model output.
    - Processed model output
      - 3.89 TB total, Yearly, 3-hourly 0.9x1.25 degree 32L MERRA2 global atmospheric forcing files for 1990-2020
      - There is other data from the original MERRA2 reanalysis that were not included in this data set. Not all fields from MERRA were included.
  - Software
    - Model configuration
      - The Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) is a global atmospheric reanalysis produced by the NASA Global Modeling and Assimilation Office (GMAO). It spans the satellite observing era from 1980 to the present. The goals of MERRA-2 are to provide a regularly-gridded, homogeneous record of the global atmosphere, and to incorporate additional aspects of the climate system including trace gas constituents (stratospheric ozone), and improved land surface representation, and cryospheric processes.
      - Meridional wind, Surface pressure, Specific humidity, Temperature, Surface geopotential, Zonal wind
    - Preprocessing code
      - N/A
    - Model code
      - NASA MERRA2 model. The model and details about the configuration is available from NASA.
    - Postprocessing code
      - Script for doing the interpolation is part of CESM2 model library. Is available on Github.
      - Development scripts are in also on Github.
  - Other
    - Documentation
      - The data repository (NCAR RDA) creates field and file metadata
      - There is a wiki for the project that has some details.
      - People who run 32 level CESM2 would know what this is for, so would not need much documentation.
    - Visualizations or images
      - N/A

- Why were these things preserved and shared?
  - General
    - Generated for NASA and others to make it more efficient to run CESM2 with nudges by meteorological data.
    - Other people may not have access to the program that was used to generate the data.
    - Without this data, it would be very tedious and resource intensive to use MERRA2 with CESM.
  - Reasons why the things listed above are important
    - Expected/intended audience and what they expect/need
      - Are there specific people who will be using the data downstream?
        - NASA requested that this data be generated.
        - Associated data is used by quite a few users.
      - Possible/aspirational users?
        - To provide an input data set. If NCAR does the work to generate this data set, why not provide this to others to use also.
        - People may also be using it for analysis of meteorological phenomena, not even as input to a model. Because it has all of the information together.
    - Note any temporal considerations, such as particular products that become more/less useful over time
      - Probably not a long-term data set because it is associated with a specific configuration of the CESM2 model. So if the model changes, it may be less useful. But it could still be useful because it is a pre-processed data set. It would be faster to use this data than processing the original MERRA2 reanalysis in-line.
      - Future models may have more levels, but there are open questions about whether to regenerate a data set like this for different model configurations.
- Broader Impacts:
  - How will output from this project be used by stakeholders?
    - Can be used more broadly than just for the initial intended use. MERRA has a lot of emphasis on hydrologic cycles, so is useful for those as well.
  - How were stakeholders involved in the data curation decision-making?
    - Not discussed
  - How will stakeholders be compensated for their participation in the data curation decision-making process?
    - Not discussed
- Do you have any concerns about misuse of your data or software? If so, what concerns do you have, and what are the reasons for those concerns?
  - Not for this data set. This is taking what is already out there and changing the format.