CESM/CAM-chem-SE (MUSICAv0) Use Case

Compiled Apr. 2022 by Model Data RCN team

Summary
Weighted rubric score - 51
Category - Preserve selected simulation workflow outputs

- Use Case Description
    - High-level overview of the use case
        - Develop a new configuration of CESM with regional refinement and test with two chemical mechanisms of varying complexity to help describe ozone, which is an important air pollutant.
        - This project used an unreleased, 'Beta' version of CESM2.2, called MUSICAv0
        - The paper presents the first steps of development for MUSICA
        - It provides a reference for comparison against observations
    - Science goals and basic workflow
        - The data includes monthly averaged data across one year (focused on summer region)
            - It is a smaller scale collection from a larger group of simulations
        - Hourly data and 6 hour outputs also saved locally but not included in the data archive collection. These were used for diagnostics.
        - Two different chemical mechanisms were tested for comparison.
        - Two different grids were tested for comparison.
        - Outputs and models are expensive to run and it is useful to share these files among the community.
    - Published journal link:
        - https://doi.org/10.1029/2021MS002889
    - NCAR data repository link:
        - https://doi.org/10.5065/wtcc-at83
- What use-case specific additional materials were preserved and shared?
    - Data
        - Inputs to model
            - Two informational files for observational aircraft files are included in the archived data. These are needed within CESM in order to replicate with aircraft data. The aircraft files within the repository are direct raw output; unprocessed. They are preserved this way for consistency and for further evaluation by other researchers to use and recreate.
            - The namelist files are included within the article associated with this data.
        - Raw model output

- 50 TB in total. But only the relevant model data with aircraft observation data was included in the repository.
    - Data used in Schwantes et al., 2022 (https://doi.org/10.1029/2021MS002889) that would be most relevant to future analyses was archived, which included the model simulated data paired to the aircraft flight tracks and monthly averaged output of all saved model species. A dataset with finer temporal resolution from a newer simulation of MUSICAv0 for 2013 with code updates beyond that used in this work was also referenced in the publication: https://doi.org/10.5065/tgbj-yv18.
    - This research involved performing full runs of CAM-Chem, to make sure it was working correctly.
    - Hourly data for the full year for a small subset of chemical species was saved for a colleague, who may analyze it in the future, but was not included in this archive.

■ Processed model output
- Total data volume preserved in a repository by the PI
    - 354.72 GB Total across 53 file
- Four main simulations included in repository:
    - TS1 chemical mechanism, ne0conus30x8 grid, and 50 h specified dynamics relaxation time (default case with ne0conus30x8 grid)
        - 13 entries
            - 10.81 GB: One ‘aircraft all’ (include the closest 9 model grid cells horizontally and all model vertical levels along the 1 minute merge aircraft flight tracks)
            - 131.52 GB: Monthly averaged data for one year simulation, split by month
    - TS2.1 chemical mechanism, ne0conus30x8 grid, and 50 h specified dynamics relaxation time
        - 3 entries
            - 14.61 GB: One ‘aircraft all’
            - 17.07 GB: Monthly averaged data for only August 2013 (focus period of paper) split into two files.
    - TS1 chemical mechanism, ne30 grid, and 50 h specified dynamics relaxation time (default case with ne30 grid)
        - 13 entries
            - 10.81 GB: One ‘aircraft all’
            - 36.72 GB: Monthly averaged data for one year simulation, split by month

        - TS2.1 chemical mechanism, ne30 grid, and 50 h specified dynamics relaxation time
            - 3 entries
                - 14.61 GB: One ‘aircraft all’
                - 4.76 GB: Monthly averaged data for only August 2013 (focus period of paper) split into two files.
    - 6 sensitivity tests with different specified dynamics relaxation times, performed with the TS2.1 chemical mechanism and both grids (ne0conus30x8 and ne30) were performed and aircraft as well as monthly averaged August 2013 data was saved.
        - TS2.1 chemical mechanism, ne0conus30x8, and 12 h specified dynamics relaxation time
            - Three files total to 25.12 GB.
        - TS2.1 chemical mechanism, ne0conus30x8, and 6 h specified dynamics relaxation time
            - Three files total to 25.12 GB.
        - TS2.1 chemical mechanism, ne0conus30x8, and no nudging over CONUS and 50 h specified dynamics relaxation time everywhere else
            - Three files total to 25.12 GB.
        - TS2.1 chemical mechanism, ne30, and 12 h specified dynamics relaxation time
            - Three files total to 12.81 GB
        - TS2.1 chemical mechanism, ne30, and 6 h specified dynamics relaxation time
            - Three files total to 12.81 GB
        - TS2.1 chemical mechanism, ne30, and no nudging over CONUS and 50 h specified dynamics relaxation time everywhere else
            - Three files total to 12.81 GB
- Software
    - Model configuration
        - Discussed in the associated paper on project web pages
    - Preprocessing code
        - Available and described in MUSICA documentation online (https://wiki.ucar.edu/display/MUSICA/MUSICA+Home)
    - Model code
        - Released in CESM 2.2 version
    - Postprocessing code
        - Working on MELODIES MONET , which is a community tool for evaluating models against observations to share post-processing code more broadly in the community in a more consistent fashion. This has been a broad problem among the research community

where everyone has their own post-processing code and it would be hard for a new user unfamiliar with the code to use it.

- Other
  - Documentation
    - Includes README file.
    - MUSICA wiki documentation page has additional details.
  - Visualizations or images [products intended to be used visually, distinguished from processed output that exists as numerical data]
    - N/A

- Why were these things preserved and shared?
  - General
    - Size and potential usefulness are important, e.g. how likely is someone to download and use this?
    - Main motivation is that data deposit is required for the journal. It is required to have model data available publicly.
    - Atmospheric chemistry model simulations especially at finer horizontal resolution are computationally expensive to perform and create large datasets. If you can archive a small subset of the data that may have future uses such as model data paired to aircraft observations, this is useful as re-running these simulations is very computationally expensive.
  - 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?
        - No
      - Possible/aspirational users?
        - Yes, especially the model data paired to aircraft observations could be used to compare to other models or as an archive to compare to new versions of CESM.
    - Note any temporal considerations, such as particular products that become more/less useful over time
      - Possibly, depending on how CESM variations differ over time. These data may not be as useful as time goes on due to updates and new iterations of the models.

- Broader Impacts:
  - How will output from this project be used by stakeholders?
    - Not applicable, there are no stakeholders for this dataset specifically as the main purpose was to archive data used in a paper. However, a newer version of the model was performed and archived here: https://doi.org/10.5065/tgbj-yv18 and this dataset has stakeholders in the atmospheric chemistry research community.
  - How were stakeholders involved in the data curation decision-making?
    - Not applicable
  - How will stakeholders be compensated for their participation in the data curation decision-making process?

        - Not applicable
- 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?
    - The data is there for anybody to use for any purpose.