CESM2 COVID and 2019/20 Australian Wildfire Simulations Use Case

Compiled Feb. 2022 by Model Data RCN team

Summary Weighted rubric score - 62 Category - Preserve selected simulation workflow outputs

- Use Case Description
 - High-level overview of the use case
 - This is a data set that underlies a published research article, titled "Coupled Climate Responses to Recent Australian Wildfire and COVID-19 Emissions Anomalies Estimated in CESM2"
 - Published journal link:
 - https://doi.org/10.1029/2021GL093841
 - NCAR data repository link:
 - <u>https://doi.org/10.5065/2sf9-mg51</u>
 - Science goals and basic workflow
 - Estimate climate response to Australian wildfires along with COVID related emission reduction.
 - This was a unique chance to see how wildfires impact the climate system. Effects of the fires were found to be similar to volcanic eruptions which was shocking to researchers and piqued interest to further research the topic.
 - Experimental Setup
 - Feeding model biomass emissions datasets into the Community Earth System Model Version 2 (CESM2), for consideration of COVID-19 and Australian wildfire impacts on climate.
- What use-case specific additional materials were preserved and shared?
 - Data
 - From NCAR repository:
 - "The simulations comprise three ensembles spanning 2019-2024. The ensembles include 50 members forced with 1) SSP245 climate forcings, 2) SSP245 combined with estimated COVID-related emissions reductions, and 3) SSP245 combined with estimated COVID-related emissions reductions and emissions from Australia's 2019-20 bushfire season."
 - The archived data is structured according to the output of the model.

- Inputs to model
 - COVID emissions inputs which were sourced from the greater climate research community.
 - Biomass data were self-created and are described in the methodology section of the paper and have potential to be reproducible. These are not available in the data repository.
 - COMMENTARY NOTE BY RCN TEAM: According to RCN project guidance the PI should archive and provide access to initialization datasets, or point to a public repository that provides access to the initialization datasets.
- Raw model output
 - Many ensemble members
 - The number of files is 3 x 50 x number of variables (11)
 - Estimated volume of original model output is 16 Tb
- Processed model output
 - Before archiving the data, they reformatted the data from time steps to time series
 - Just archived the variables that were used for this study, not all variables that the model generated.
 - This reduced the size by a factor of 2 or 3
 - Total data volume preserved in a repository by the PI
 - 15.71 GB, 1650 Files
 - Files within the repository are homogeneous in type (netCDF) and volume (around 10Mb per file).
- Software
 - Model configuration
 - No changes from default model configuration
 - Preprocessing code
 - Methodology for those datasets is described in the paper
 - Model code
 - CESM2 is a community model, available at https://doi.org/10.5065/d67h1h0v
 - Postprocessing code
 - Minimal post processing code. Only a basic concatenation of fifty variables into time series.
- \circ Other
 - Documentation
 - Description on the data set website and within the paper.
 - Visualizations or images [products intended to be used visually, distinguished from processed output that exists as numerical data]
 - Considered including this in the data archive, but did not.
- Why were these things preserved and shared?

- General
 - Main considerations were the journal requirements & the storage size required for the data.
 - The submission of this ensemble data set to the data repository was very easy due to the ease of use and for being located at NCAR. It was very convenient and easy to meet journal requirements by also submitting to the NCAR data repository.
- 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 specific users identified. Broad community use possible.
 - Possible/aspirational users?
 - Also possible by educators and students.
 - Note any temporal considerations, such as particular products that become more/less useful over time
 - No temporal considerations.
- Broader Impacts:
 - How will output from this project be used by stakeholders?
 - No direct stakeholders other than the PIs.
 - How were stakeholders involved in the data curation decision-making?
 - N/A
 - How will stakeholders be compensated for their participation in the data curation decision-making process?
 - N/A
- 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?
 - No. Researcher has not considered the possibility of data being misused. The researcher also mentions they haven't seen the issue of misuse in similar data sets.