Datasheet for Hueholt et al. *submitted* "Assessing Outcomes in Stratospheric Aerosol Injection Scenarios Shortly After Deployment" [1]

Released: November 29, 2022
Last updated: January 17, 2022

Daniel M. Hueholt
Department of Atmospheric Science
Colorado State University
Fort Collins, CO
daniel.hueholt@colostate.edu

1. Purpose

A. For what purpose was the dataset created?

Motivation: Describe the reason for the creation of the dataset (e.g., to provide insight on a knowledge gap, or to carry out some specific task).

The dataset corresponding to this datasheet (henceforth [2]) was created to explore the climate response to two scenarios of stratospheric aerosol injection (SAI) climate intervention, with a focus on short, policy-relevant timescales after deployment. This dataset supports the goals of the U.S. National Academies of Sciences report [3] to "advance knowledge relevant to decision making" and "develop policy-relevant knowledge" about potential risks and benefits of SAI, particularly the "global- and regional-scale impacts," and the "distribution of impacts across different parts of the world." It is a subset of the pre-existing Geoengineering Large ENSemble (GLENS [4]), Assessing Responses and Impacts of Solar climate intervention on the Earth system with Stratospheric Aerosol Injection (ARISE-SAI-1.5 [5]), and CMIP6 CESM2(WACCM6) Historical [6] numerical modeling experiments as well as additional variables derived from these datasets.

B. Who created the dataset (e.g., which individual or research group), on behalf of which entity (e.g., institution or company), and under what funding (e.g., grantor[s] and grant number[s])?

Motivation: Provide clarity about the authorship and funding source of the given dataset.

The dataset [2] was created by Daniel M. Hueholt [Contact: daniel.hueholt@colostate.edu] as part of his M.S. thesis work at Colorado State University advised by Prof. Elizabeth A. Barnes and Prof. James W. Hurrell. Funding was provided by the LAD Climate Fund and the Defense Advanced Research Projects Agency (DARPA, grant no. HR00112290071). DMH was additionally funded by the National Science Foundation Graduate Research Fellowship Program.
The GLENS and ARISE-SAI-1.5 datasets, from which [2] was derived, were created by Tilmes et al. [4] with DARPA grant funding and Richter et al. [5] with funding from SilverLining through the Safe Climate Research Initiative. CESM2(WACCM6) Historical [6] was produced by the National Center for Atmospheric Research (NCAR). GLENS, ARISE-SAI-1.5, and the CESM2(WACCM6) Historical were produced and maintained at NCAR, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement no. 1852977. The CESM project is supported primarily by the National Science Foundation.

C. Was the author of the datasheet involved in creating the dataset? If so, how? If not, please describe your relation to this dataset.

Motivation: Document the authorship of the datasheet, which may be different than the creator of the dataset.

The writer of this datasheet (DMH) was the creator of [2]. The GLENS and ARISE-SAI-1.5 modeling experiments were conducted by Tilmes et al. [4] and Richter et al. [5], and CESM2(WACCM6) Historical by Danabasoglu et al. [6].

D. Any other comments?

Motivation: Space for any other relevant information about the creation of the dataset.

2. Composition

This section concerns technical aspects of the dataset. If any information is documented elsewhere you may simply provide a brief description and stable link (e.g., digital object identifier [DOI]) in the relevant question(s).

A. What type of data is contained in this dataset? (e.g., is it model output, observational data, reanalysis, etc.?)

Motivation: Basic information about the fundamental classification of your data.

This dataset contains output from modeling experiments using ensembles of the Community Earth System Model (CESM) with Whole Atmosphere Community Climate Model (WACCM). GLENS uses CESM1(WACCM5) while ARISE-SAI-1.5 and CESM2(WACCM6) Historical use CESM2(WACCM6). Both GLENS and ARISE-SAI-1.5 contain parallel simulations: one following a climate change trajectory, and one where SAI is also deployed. GLENS and ARISE-SAI-1.5 both inject sulfur dioxide at multiple latitudes and use a proportional-integral feedback-control algorithm to maintain three climate targets: global mean temperature, the pole-to-pole temperature gradient, and the pole-to-equator temperature gradient. For a more complete summary of both experiments, see [1]. For details of GLENS and ARISE-SAI-1.5 see [4] and [5], respectively.

B. What is the data? (e.g., file format, dimensionality, variables and metadata, spatiotemporal coverage)

Motivation: Provide format and characteristics of the data.

All data files are in netCDF format. Files from each experiment are identified with the following tokens occurring at the start of each filename.

- **GLENS no-SAI (RCP8.5)** `control_*`
- **GLENS SAI** `feedback_*`
- **ARISE-SAI-1.5 no-SAI (SSP2-4.5)** `BWSSP245cmip6_*`
- **ARISE-SAI-1.5 SAI** `SSP245-TSMLT-GAUSS_*`
- **CESM2(WACCM6) Historical** `BWHIST_*`

The variables included and their basic properties are provided in the following list. Spatiotemporal coverage is described for the period 2010-2069 as this is the relevant range in [1].

- **Temperature [TREFHT]** 2m temperature from atmospheric model. Annual mean calculated from monthly data using cdo `yearmonmean`.
  - **Dimensions** time (variable) x lat (192) x lon (288)
  - **Units** Kelvin
  - **Spatiotemporal coverage through 2069**
    - **GLENS no-SAI (RCP8.5)** 21 members 2010-2030 with 4 extending through 2069, global
    - **GLENS SAI** 21 members 2020-2069, global
    - **ARISE-SAI-1.5 no-SAI (SSP2-4.5)** 10 members 2015-2069, global
    - **ARISE-SAI-1.5 SAI** 10 members 2035-2069, global
    - **CESM2(WACCM6) Historical** 3 members 2010-2014
- **Tropical nights [clxTR]** Annual tropical nights calculated from daily minimum 2m temperature using the Pyclimdex package [7].
  - **Dimensions** time (variable) x lat (192) x lon (288)
  - **Units** days/yr
  - **Spatiotemporal coverage through 2069**
    - **GLENS no-SAI (RCP8.5)** 20 members 2010-2030 with 3 extending through 2069, global
    - **GLENS SAI** 21 members 2020-2069, global
    - **ARISE-SAI-1.5 no-SAI (SSP2-4.5)** 5 members 2015-2069, global
    - **ARISE-SAI-1.5 SAI** 10 members 2035-2069, global
    - **CESM2(WACCM6) Historical** No data
- **Sea surface temperature [TEMP]** Annual sea surface temperature calculated by taking potential temperature from the top level of the ocean model. Annual mean calculated from monthly data using cdo `yearmonmean`. This data has been regridded from ocean grid to lat/lon grid using the Python implementation in `wrap_ocean_script`.
  - **Dimensions** time (variable) x lat (192) x lon (288)
  - **Units** degrees Celsius
  - **Spatiotemporal coverage through 2069**
    - **GLENS no-SAI (RCP8.5)** 21 members 2010-2030 with 4 extending through 2069, global
    - **GLENS SAI** 21 members 2020-2069, global
    - **ARISE-SAI-1.5 no-SAI (SSP2-4.5)** 10 members 2015-2069, global
    - **ARISE-SAI-1.5 SAI** 10 members 2035-2069, global
    - **CESM2(WACCM6) Historical** 3 members 2010-2014
- **Number of days with marine heatwaves [binary_mhw_pres]** Number of days with marine heatwaves at the point -30.628N, 112.5E. A binary timeseries of marine heatwave presence/absence is first calculated from daily sea surface temperature data using `marineHeatWaves` [8], implemented in accompanying Python code `fun_derive_data`. A 5-year left-aligned rolling sum of days is then applied to this data to smooth interannual variability.
  - **Dimensions** time (variable)
  - **Units** Days/yr
  - **Spatiotemporal coverage through 2069**
    - **GLENS no-SAI (RCP8.5)** 4 members 2010-2069, at point -30.628N, 112.5E

    * **GLENS SAI** 21 members 2020-2069, at point -30.628N, 112.5E
    * **ARISE-SAI-1.5 no-SAI (SSP2-4.5)** 10 members 2015-2069, at point -30.628N, 112.5E
    * **ARISE-SAI-1.5 SAI** 10 members 2035-2069, at point -30.628N, 112.5E
    * **CESM2(WACCM6) Historical** 3 members 2010-2014
    * Note that the 3 Historical members and first 3 **ARISE-SAI-1.5 no-SAI (SSP2-4.5)** members have been merged into single files in order to simplify calculation code
- **Sea ice extent [ICEEXTENT]** Sea ice extent calculated by summing grid cell area with ice fraction greater than 0.15 in the atmospheric model. February and September data are selected from monthly data using cdo `selmon`.
  - **Dimensions** time (variable) x lat (192) x lon (288)
  - **Units** km^2
  - **Spatiotemporal coverage through 2069**
    * **GLENS no-SAI (RCP8.5)** 21 members Feb+Sept 2010-2030 with 4 extending through 2069, global
    * **GLENS SAI** 21 members Feb+Sept 2020-2069, global
    * **ARISE-SAI-1.5 no-SAI (SSP2-4.5)** 10 members Feb+Sept 2015-2069, global
    * **ARISE-SAI-1.5 SAI** 10 members Feb+Sept 2035-2069, global
    * **CESM2(WACCM6) Historical** 3 members Feb+Sept 2010-2014, global
- **Precipitation [PRECT]** Precipitation calculated by summing the convective and stratiform precipitation in GLENS, or taking the total precipitation variable available by default in ARISE-SAI-1.5 and CESM2(WACCM6) Historical. Annual mean data is calculated from monthly output using cdo `yearmonmean`. Monsoon average precipitation is Jun-Sept mean where months selected from monthly output then averaged, corresponding to the Northern Hemisphere monsoon season.
  - **Dimensions** time (variable) x lat (192) x lon (288)
  - **Units** m/s (precipitation rate)
  - **Spatiotemporal coverage through 2069**
    * **GLENS no-SAI (RCP8.5)** 21 members 2010-2030 with 4 extending through 2069, global
    * **GLENS SAI** 21 members 2020-2069, global
    * **ARISE-SAI-1.5 no-SAI (SSP2-4.5)** 10 members 2015-2069, global
    * **ARISE-SAI-1.5 SAI** 10 members 2035-2069, global
    * **CESM2(WACCM6) Historical** 3 members 2010-2014, global
- **Simple intensity index [sdii]** Simple intensity index calculated from daily precipitation using `Pyclimdex` [7].
  - **Dimensions** time (variable) x lat (192) x lon (288)
  - **Units** mm/day
  - **Spatiotemporal coverage through 2069**
    * **GLENS no-SAI (RCP8.5)** 20 members 2010-2030 with 3 extending through 2069, global
    * **GLENS SAI** 21 members 2020-2069, global
    * **ARISE-SAI-1.5 no-SAI (SSP2-4.5)** 10 members 2015-2069, global
    * **ARISE-SAI-1.5 SAI** 10 members 2035-2069, global
    * **CESM2(WACCM6) Historical** 3 members 2010-2014, global

The basic structure of each filename consists of the following pieces.

- **id** Identifies the experiment, described above
- **rlz** The ensemble member
- **var** The variable name
- **YYYYMM1** The starting year and month
- **YYYYMMn** The ending year and month
  - The YYYYMM are chained together from each individual file when they are merged as part of processing
  - An identifier RG may also be present, indicating the data has been regridded from an ocean grid to lat-lon coordinates using `wrap_ocean_script`
- **ext** Some extra information about time contents of file (e.g., annual implies annual, sept implies September)

A filename is given by:
`id_rlz_var_YYYYMM1[...]_YYYYMMn_ext.nc`

C. What processing has been applied to this data?

Motivation: Minimal description of the process to obtain the data described by this datasheet from its unprocessed form.

The processing for each variable is described in the answer to Question 2B.

D. Is the unprocessed data available in addition to the processed data? If so, please provide a stable link to the unprocessed data.

Motivation: Clarify the location of the unprocessed data to facilitate reproducibility or unforeseen future uses, if possible.

Data from GLENS, ARISE-SAI-1.5, and CESM2(WACCM6) Historical are found at the following archives.

- **GLENS:** doi.org/10.5065/D6JH3JXX, hosted by NCAR
- **ARISE-SAI-1.5:** NCAR Climate Data Gateway doi.org/10.5065/9kcn-9y79 (all SAI, 5 no-SAI members) and doi.org/10.26024/0cs0-ev98 (remaining 5 no-SAI members); or Amazon Web Services registry.opendata.aws/ncar-cesm2-arise/

- **CESM2(WACCM6) Historical:** in CMIP6 format from doi.org/10.22033/ESGF/CMIP6.11298. The raw version of CESM2(WACCM6) Historical directly used in this study is located on the NCAR GLADE file space

E. Is the code used to process the data available? If so, please provide a stable link or other access point.

Motivation: Share processing methodology to facilitate reproducibility, if possible.

All code is available in the Python package included with the dataset. This version of record corresponds to the code used in [1]. An up-to-date version is available on GitHub at github.com/dmhuehol/SAI-CESM.

F. Is this dataset derived from another dataset? If so, how?

Motivation: Describe whether a dataset is drawn or derived from another preexisting dataset (e.g., field campaign, model intercomparison).

This data is derived from a subset of the GLENS [4], ARISE-SAI-1.5 [5], and CESM2(WACCM6) Historical [6] modeling experiments. The subset presented here was selected to focus on the impacts of SAI on familiar climate variables on policy-relevant timescales after deployment. See [1] for more discussion of the specific variables and times of interest.

G. Is any relevant information known to be missing from the dataset? If so, please provide an explanation.

Motivation: Describe missing data and be transparent about causes of missing data within the dataset.

See Section 2B for a description of data available for each variable. Not every ensemble member generated the same data for the same span of time due to limitations of computing availability. (This is not truly “missing” data, but should be noted when assessing the available data described here.)
Note that there may be missing data elsewhere in the GLENS, ARISE-SAI-1.5, or CESM2(WACCM6) Historical datasets. The answer here corresponds only to the specific subset included [2].

H. Are there any sources of noise, redundancies, or errors in the dataset? If so, please provide a description.

Motivation: Provide information about relevant known technical issues that affect all or portions of the dataset.

Daily minimum temperature data is only available for five members of the ARISE-SAI-1.5 no-SAI runs. In the other five, this data was unintentionally overwritten with daily mean temperature data. Thus, annual tropical nights can only be calculated for five ARISE-SAI no-SAI members.
Note that errors may exist elsewhere in the GLENS, ARISE-SAI-1.5, or CESM2(WACCM6) Historical datasets. The datasheet only provides information corresponding directly to [2].

I. Is the dataset self-contained, or does it rely on external resources? Please provide descriptions of external resources and any associated restrictions, as well as relevant links or other access points.

Motivation: Explicitly track external dependencies that may otherwise go unacknowledged.

This data [2] is self-contained in its ability to reproduce results from [1]. The Pyclimdex [7] and marineHeatWaves [8] Python packages used to derive data as described in Question 2B are not maintained by the author, and the author makes no claim to their long-term consistency. The greater GLENS, ARISE-SAI-1.5, and CESM2(WACCM6) Historical datasets are archived as described in Question 2D.

J. Any other comments?

Motivation: Space for any other relevant information about the composition of the dataset.

3. Uses

A. What tasks has the dataset been used for? Please provide a description and/or citation(s); if there is a repository that archives uses of the dataset, provide the stable link here.

Motivation: Document use cases of the dataset within the scope of this datasheet.

This specific dataset has been used for [1]. This datasheet does not archive uses of the the greater GLENS, ARISE-SAI-1.5, and CESM2(WACCM6) Historical datasets.

B. Is there anything about the construction of the dataset that might impact future uses?

Motivation: Be transparent about how the composition or processing of the dataset could affect future uses.

The variables selected here are a subset of the greater GLENS, ARISE-SAI-1.5, and CESM2(WACCM6) Historical experiments which were specifically relevant to the questions explored in [1]. The data [2] is provided for reproducibility and independent assessment of this work. There is a huge breadth of data available in the greater modeling experiments; future users are encouraged to build off of that raw data for original new research.
As a minor note relevant to modifying the data in [2]: Due to the amount of information contained in the filenames, the

filenames are near the 256-character maximum length imposed by most file systems.

C. Are there specific tasks for which the dataset should not be used? If so, please provide a description.

Motivation: Address relevant gaps or inadequacies of the data for specific use cases.

GLENS was run without interactive tropospheric trace gas chemistry, or ocean biogeochemistry. Thus, GLENS data cannot be used to explore these processes.

GLENS and ARISE-SAI-1.5 differed in aspects of their experimental design. We summarize these differences and their implications in more detail in [1] while [9] is specifically devoted to untangling these issues. In brief, differences include the version of CESM(WACCM) used, the underlying greenhouse gas forcing scenario, the temperature target and SAI deployment year, method of causing ensemble spread, and the injection height of the sulfur dioxide. GLENS and ARISE-SAI-1.5 provide high-fidelity depictions of two scenarios of SAI intervention, but the results are specific to these scenarios and should not be assumed to be true of any general SAI scenario.

The ocean state in all members of GLENS is branched off of the first member of the CESM Large Ensemble in which the Atlantic Meridional Overturning Circulation is strengthening. Until the memory of these initial conditions dissipates, GLENS does not disperse ocean internal variability.

D. What are the potential impacts of this dataset on humans? Please provide a description as well as a stable link to any supporting documentation.

Motivation: Reflect on the potential impacts (direct or downstream) of the dataset on human systems.

This dataset was intended to explore questions related to the global and regional climate responses to stratospheric aerosol injection climate intervention, consistent with the goals of [3]. We explicitly distinguish these goals from research on deployment technologies, which are discouraged due to ethical and governance concerns. This data is intended to support informed and transparent public discussion of climate intervention.
A large body of literature discussing the ethics of climate intervention exists; the author suggests e.g. [10], [11], [12], [3], [13] as a sample of resources for readers interested in engaging with some of the different perspectives on this topic in depth.

E. Any other comments?

Motivation: Space for any other relevant information about uses of the dataset.

4. Distribution and Maintenance

A. How will the dataset be distributed (e.g., FTP server, Earth System Grid, Amazon Web Services, etc.)? Is there a DOI or other stable link?

Motivation: Document stable access to the dataset.

[2] is archived at the Open Science Foundation doi.org/10.17605/OSF.IO/5A2ZF. GLENS, ARISE-SAI-1.5, and CESM2(WACCM6) Historical are archived at the locations described in Question 2D.

B. Who is/are the point(s) of contact for this dataset?

Motivation: Provide information about who is responsible for responding to inquiries about this dataset.

The point of contact for this dataset [2] is Daniel M. Hueholt [Contact: daniel.hueholt@colostate.edu].
As of the time of this writing (December 2022), the current points of contact for GLENS were provided at cesm.ucar.edu/community-projects/glens and ARISE-SAI at cesm.ucar.edu/community-projects/arise-sai. Note these links are NOT stable and may break without warning. The current point of contact for CESM2(WACCM6) Historical is unclear.

C. Is the dataset complete or will it be updated in the future?

Motivation: Clarify whether this version of the data is final.

This data [2] accompanies the specific work [1] and is not intended to be updated following publication.

D. Is the dataset receiving ongoing maintenance? If so, please provide one or more point(s) of contact and describe the method (if any) by which updates would be communicated to users.

Motivation: Provide information about whether the dataset is receiving ongoing support.

[2] is stably hosted on the Open Science Foundation doi.org/10.17605/OSF.IO/5A2ZF. Ongoing maintenance for the dataset is not anticipated. Any changes that do occur (i.e. in the case of an unforeseen error) will be logged in the datasheet and in the changelog of the OSF repository.

E. What license or other terms of use is the dataset distributed under? Please link to any relevant licensing terms or terms of use (if in the public domain, simply state this).

Motivation: Provide information about what future uses of the data are permitted.

This dataset is a work in the public domain. The code repository that accompanies this dataset is hosted under the GNU General Public License 3.0, included as LICENSE.txt.

F. Is there an erratum? If so, please provide a link or other access point.

Motivation: Document any corrections to the dataset.

No erratum currently exists. Any corrections will be recorded under this question in future updates to the datasheet, and in the changelog of the OSF repository.

G. Will older versions of the dataset continue to be available? If so, please describe where.

Motivation: Describe whether any specific version of the dataset will always be accessible.

Any previous versions of the dataset are accessible through the changelog of the OSF repository.

H. Who is hosting the datasheet? Is the datasheet receiving ongoing maintenance?

Motivation: Clarify stable access to the datasheet and whether it will be updated.

This datasheet is stably hosted on the Open Science Foundation in the same archive as the data (doi.org/10.17605/OSF.IO/5A2ZF). The point of contact is Daniel M. Hueholt [Contact: daniel.hueholt@colostate.edu]. Any changes to the datasheet will be recorded in the changelog of the OSF repository.

I. Any other comments?

Motivation: Space for any other relevant information about data distribution and maintenance.

5. Data-dependent questions

Responses in this section will be dependent on the type(s) of data contained in the dataset. Questions that do not apply can be left blank.

A. How was the data generated or collected? (e.g., a model used to produce output, reanalysis estimation of conditions, observations using remote sensing methods or in situ sensors) Please provide relevant citation(s); if none exist, describe why.

Motivation: Establish fundamental information about the methods used to generate or collect data in the dataset.

This dataset is a subset of the greater GLENS, ARISE-SAI-1.5, and CESM2(WACCM6) Historical modeling experiments. GLENS utilized CESM1(WACCM5) [14], [15]. ARISE-SAI-1.5 and CESM2(WACCM6) Historical utilized CESM2(WACCM6) [6], [16]. Processing to obtain this data [2] from the original datasets is described in Section 2.

B. If the data has been evaluated against some baseline(s) (e.g., an observational product or fundamental physical laws), please describe its evaluation against that baseline(s). If available, simply provide the relevant citation.

Motivation: Document adequacy of the method (e.g., model, remote sensing retrieval) within the scope of this datasheet.

WACCM evaluates well against observations of stratospheric conditions in the mean state and in anomalous conditions of volcanic aerosol loading [15], [17], i.e. WACCM has very good adequacy for the purpose of simulating the climate response to stratospheric aerosol injection. The performance of CESM1(WACCM5) and CESM2(WACCM6) has been evaluated for many products in the citations provided in Question 5A. Consult these works for information about the general performance of CESM(WACCM).

C. Please provide relevant known biases in the generation or collection method of this data and citations as available. This list does not need to be exhaustive, but should include any known biases relevant to the scope of the project the data was created for.

Motivation: Document known biases that pertain specifically to the scope of the project at hand.

Broadly, Earth system model output is most reliable and verifiable where observational datasets are easily available to test, evaluate, and constrain the model. Appropriate long-term observational datasets are sparse over much of the world, particularly in the Global South. It is important to evaluate the model's performance before making specific quantitative claims in these areas.

D. Please note configurations or modifications made to any model used to complete runs in this dataset (e.g. changes to seasonality, changes to coupling, nudging), or provide relevant startup files.

Motivation: Be transparent about the exact setup of the model to create the data at hand.

The model was not altered as part of the project at hand. No original model runs were completed in [1].

E. If this data is restricted to a single point or region, why was this location or region chosen? What are some potential implications of this choice of location on the interpretation of the data?

Motivation: Describe the reasoning for and any relevant impacts of the selection of this location.

Marine heatwaves were calculated for a single grid point off the coast of Western Australia (30.63°S, 112.5°E). This location was chosen as it frequently experiences marine heatwaves with substantial negative impacts to local ecology (e.g. [18], [19]). All other data in this dataset has global coverage.

F. Describe relevant uncertainties associated with this data or provide relevant citation(s). If no formal analysis of uncertainties has been completed, then please state this here.

Motivation: Provide information about known uncertainties within the scope of the project.

As GLENS, ARISE-SAI, and CESM2(WACCM6) Historical are single-model initial condition large ensembles, the spread of the members represents internal climate variability. Uncertainty quantification within CESM is discussed in [14] for CESM1 and [6] for CESM2.

G. Did the method of generation or collection of the data change within the extent of the dataset?

Motivation: Be transparent about important changes to instruments or methodology within the dataset.

No known changes to the method of generation of the data within the extent of the dataset.

H. Are there any relevant unexplained but important numerical values ("magic numbers") that go into the generation, collection, or processing of this data? (e.g., model tuning values, calibration constants, machine learning hyperparameters)

Motivation: Define unique numerical values that exist within or impact this data, but may not be documented elsewhere.

The following magic numbers may appear in the accompanying Python code. These should be documented with inline comments in the code itself, but are compiled here for completeness.

- Standard date fill values 7,15,12,0,0,0 (month, day, hour, minute, second, microsecond) or 1,1,0,0 (month, day, hour, minute, second) where arbitrary times must be inserted for annual-mean data
- A 5-year left-aligned rolling sum of days applied to the marine heatwaves data to smooth interannual variability
- The exact latitude size of cells in the model is 0.94240838°
- Robustness thresholds of 15 SAI members outside of 11 no-SAI members for GLENS and 7 SAI members outside of 6 no-SAI members in ARISE-SAI. These correspond to values outside 90% of a randomly-generated distribution. For a description of robustness, see the Supplement to [1].

I. Is this dataset an ensemble? If so, how many members are there? Describe how the ensemble is perturbed, and whether there are relevant forms of variability that are not dispersed. Are there differences in coverage between the ensemble members?

Motivation: Describe the sampling, construction, and any important limitations of the ensemble.

GLENS, ARISE-SAI-1.5, and CESM2(WACCM6) Historical are all single-model initial condition ensembles. GLENS does not disperse ocean internal variability in the initial conditions (see Question 3C). Consult Question 2B for detailed descriptions of coverage for each variable.

- **GLENS:** 21 ensemble members. GLENS was run with 20 members in [4]; the 21st member was added later.
- **ARISE-SAI-1.5:** 10 ensemble members. The no-SAI simulations were run in two sets; five of the members are simply the SSP2-4.5 CESM2(WACCM6) simulations completed as part of CMIP6. Five of the members were run later to provide extra information for ARISE-SAI-1.5.
- **CESM2(WACCM6) Historical**: 3 ensemble members

J. Are there relevant categories, groupings, or labels within the data? If so, how are these determined?

Motivation: Be transparent about the processes used to define groups within the data.

Beyond the separation of the parallel SAI and no-SAI simulations, there are no noteworthy categories, groupings or labels within the data.

K. Can users contribute to this dataset? If so, please describe the process. Will these contributions be evaluated or verified? If so, please describe how. If not, why not?

Motivation: Describe if user contributions make up part of the dataset (e.g., citizen science or human labeling).

Users cannot contribute directly to this dataset.

L. Any other comments? Are there any other citations necessary to document some important aspect of the data? If so, provide the citation(s) and describe their purpose.

Motivation: Space for any other relevant information about the data. Can include specific useful citations that do not fall naturally into any other question.

References

[1] Daniel M. Hueholt, Elizabeth A. Barnes, James W. Hurrell, Jadwiga H. Richter, and Lantao Sun. Assessing Outcomes in Stratospheric Aerosol Injection Scenarios Shortly After Deployment *(submitted to Earth's Future)*, 2022.

[2] Daniel M. Hueholt. Dataset accompanying "Assessing Outcomes in Stratospheric Aerosol Injection Scenarios Shortly After Deployment" *(in prep)*, 2022.

[3] NASEM. Reflecting Sunlight: Recommendations for Solar Geoengineering Research and Research Governance. Technical report, National Academies of Science, Engineering, and Medicine, 2021.

[4] Simone Tilmes, Jadwiga H. Richter, Ben Kravitz, Douglas G. MacMartin, Michael J. Mills, Isla R. Simpson, Anne S. Glanville, John T. Fasullo, Adam S. Phillips, Jean-Francois Lamarque, Joseph Tribbia, Jim Edwards, Sheri Mickelson, and Siddhartha Ghosh. CESM1(WACCM) Stratospheric Aerosol Geoengineering Large Ensemble Project. *Bulletin of the American Meteorological Society*, 99(11):2361–2371, November 2018. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

[5] Jadwiga H. Richter, Daniele Visioni, Douglas G. MacMartin, David A. Bailey, Nan Rosenbloom, Brian Dobbins, Walker R. Lee, Mari Tye, and Jean-Francois Lamarque. Assessing Responses and Impacts of Solar climate intervention on the Earth system with stratospheric aerosol injection (ARISE-SAI): protocol and initial results from the first simulations. *Geoscientific Model Development*, 15(22):8221–8243, November 2022. Publisher: Copernicus GmbH.

[6] G. Danabasoglu, J.-F. Lamarque, J. Bacmeister, D. A. Bailey, A. K. DuVivier, J. Edwards, L. K. Emmons, J. Fasullo, R. Garcia, A. Gettelman, C. Hannay, M. M. Holland, W. G. Large, P. H. Lauritzen, D. M. Lawrence, J. T. M. Lenaerts, K. Lindsay, W. H. Lipscomb, M. J. Mills, R. Neale, K. W. Oleson, B. Otto-Bliesner, A. S. Phillips, W. Sacks, S. Tilmes, L. van Kampenhout, M. Vertenstein, A. Bertini, J. Dennis, C. Deser, C. Fischer, B. Fox-Kemper, J. E. Kay, D. Kinnison, P. J. Kushner, V. E. Larson, M. C. Long, S. Mickelson, J. K. Moore, E. Nienhouse, L. Polvani, P. J. Rasch, and W. G. Strand. The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12(2):e2019MS001916, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS001916.

[7] Brian Groenke. pyclimdex, March 2022. original-date: 2020-01-23T06:37:54Z.

[8] Eric Oliver. Marine Heatwaves detection code, May 2022. original-date: 2015-04-28T07:33:04Z.

[9] John T. Fasullo and Jadwiga H. Richter. Scenario and Model Dependence of Strategic Solar Climate Intervention in CESM, April 2022. Archive Location: world Publisher: Earth and Space Science Open Archive Section: Climatology (Global Change).

[10] Kyle Powys Whyte. Now This! Indigenous Sovereignty, Political Obliviousness and Governance Models for SRM Research. *Ethics, Policy & Environment*, 15(2):172–187, June 2012. Publisher: Routledge _eprint: https://doi.org/10.1080/21550085.2012.685570.

[11] Holly Jean Buck, Andrea R. Gammon, and Christopher J. Preston. Gender and Geoengineering. *Hypatia*, 29(3):651–669, 2014. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/hypa.12083.

[12] Elizabeth T. Burns, Jane A. Flegal, David W. Keith, Aseem Mahajan, Dustin Tingley, and Gernot Wagner. What do people think when they think about solar geoengineering? A review of empirical social science literature, and prospects for future research. *Earth's Future*, 4(11):536–542, November 2016. Publisher: John Wiley & Sons, Ltd.

[13] Olúfemi O. Táíwò and Shuchi Talati. Who Are the Engineers? Solar Geoengineering Research and Justice. *Global Environmental Politics*, pages 1–7, July 2021.

[14] James W. Hurrell, M. M. Holland, P. R. Gent, S. Ghan, Jennifer E. Kay, P. J. Kushner, J.-F. Lamarque, W. G. Large, D. Lawrence, K. Lindsay, W. H. Lipscomb, M. C. Long, N. Mahowald, D. R. Marsh, R. B. Neale, P. Rasch, S. Vavrus, M. Vertenstein, D. Bader, W. D. Collins, J. J. Hack, J. Kiehl, and S. Marshall. The Community Earth System Model: A Framework for Collaborative Research. *Bulletin of the American Meteorological Society*, 94(9):1339–1360, September 2013. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

[15] Michael J. Mills, Jadwiga H. Richter, Simone Tilmes, Ben Kravitz, Douglas G. MacMartin, Anne A. Glanville, Joseph J. Tribbia, Jean-François Lamarque, Francis Vitt, Anja Schmidt, Andrew Gettelman, Cecile Hannay, Julio T. Bacmeister, and Douglas E. Kinnison. Radiative and Chemical Response to Interactive Stratospheric Sulfate Aerosols in Fully Coupled CESM1(WACCM). *Journal of Geophysical Research: Atmospheres*, 122(23):13,061–13,078, 2017. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2017JD027006.

[16] A. Gettelman, M. J. Mills, D. E. Kinnison, R. R. Garcia, A. K. Smith, D. R. Marsh, S. Tilmes, F. Vitt, C. G. Bardeen, J. McInerny, H.-L. Liu, S. C. Solomon, L. M. Polvani, L. K. Emmons, J.-F. Lamarque, J. H. Richter, A. S. Glanville, J. T. Bacmeister, A. S. Phillips, R. B. Neale, I. R. Simpson, A. K. DuVivier, A. Hodzic, and W. J. Randel. The Whole Atmosphere Community Climate Model Version 6 (WACCM6). *Journal of Geophysical Research: Atmospheres*, 124(23):12380–12403, 2019. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019JD030943.

[17] Jadwiga H. Richter, Simone Tilmes, Michael J. Mills, Joseph J. Tribbia, Ben Kravitz, Douglas G. MacMartin, Francis Vitt, and Jean-Francois Lamarque. Stratospheric Dynamical Response and Ozone Feedbacks in the Presence of SO2 Injections. *Journal of Geophysical Research: Atmospheres*, 122(23):12,557–12,573, 2017. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2017JD026912.

[18] Arani Chandrapavan, Nick Caputi, and Mervi I. Kangas. The Decline and Recovery of a Crab Population From an Extreme Marine Heatwave and a Changing Climate. *Frontiers in Marine Science*, 6, 2019.

[19] Neil Holbrook, Alex Gupta, Eric Oliver, Alistair Hobday, Jessica Benthuysen, Hillary Scannell, Dan Smale, and Thomas Wernberg. Keeping pace with marine heatwaves. *Nature Reviews Earth & Environment*, 1, July 2020.