Stratospheric aerosol data for use in CMIP6 models – data description

(for more details, see the paper by Thomason et al.)

Data sources

1979 – 2014: Based on satellite data assembled by Larry Thomason using SAGE, SAM, SAGE II, CALIPSO and OSIRIS. Furthermore, CLAES data were used for gap-filling of missing data for the period several months after the Pinatubo eruption.
1850 – 1978: For volcanically quiescent periods, the monthly resolved mean background aerosol measured by SAGE II during the volcanic quiescent time period (1996 -2005) is used. The volcanic contribution is calculated using the AER-2D model for the spatial and temporal evolution (depending on location of eruption and season), see Arfeuille et al. (2014). The stratospheric AOD is calibrated using photometer data when available. Elsewise, the best estimate of sulfur ejection is used for the volcanic contribution (estimated from proxies such as ice core data). All data points in the 1850-1978 period are flagged by a zero (see flag table below).

Data coverage and resolution

Data are three dimensional (time, altitude and latitude) arrays of single mode lognormal aerosol size distributions specified in terms of mean radius, surface area density, volume density and condensed phase H_2SO_4 molecule number densities. For each GCM or CCM, radiative properties of these aerosols are provided on the same grid, namely extinctions, single scattering albedos, and asymmetry factors for model-specific wavelength bands.

These data are provided

- as monthly means, starting with Jan 1850, ending with Dec 2014;
- pole to pole as zonal means, averaged in latitude bands of five degrees (centered at $\pm$2.5°, $\pm$7.5°, …, $\pm$87.5°), with data poleward of 80° being extrapolated (identical to 87.5°-data);
- in altitude bins of 0.5 km resolution from 5 km to 39.5 km, valid only for the stratosphere (and offered below the tropopause for purposes of matching with the tropospheric aerosol in the model).

The data are reliable strictly only above and at the tropopause. The tropospheric values are less reliable due to the cloud clearing process introducing noise. Furthermore, in the polar winter stratosphere, data may be contaminated by PSCs.
It is strongly recommended to use the data below the instantaneous local model tropopause only in order to establish a smooth transition between the stratospheric aerosol and the tropospheric aerosol used by the models. While the data at lower altitudes serve this purpose very well, they must not be abused as regular tropospheric aerosols. To serve this purpose the data have been extrapolated from SAGE data downward to lower altitudes (namely down to 5 km, filled up using the last measured value from SAGE). These values are flagged by "-1" in the flag column. See flag table below.

Procedure for dealing with the strat-trop interface

We suggest the following procedure to combine the stratospheric and tropospheric aerosol data sets:

(1) **In the model levels below the instantaneous local model tropopause**, use the tropospheric data for radiative forcing and heterogeneous chemistry calculations. Do not use the data of the current data set as regular tropospheric aerosols!
(2) **In the model layer holding the tropopause**, take the arithmetic means, i.e. 50% of stratospheric and tropospheric data each. Apply the arithmetic mean to extinction coefficients, single scattering albedos, asymmetry factors, surface area densities and mean radii. (If these quantities are not available for the troposphere, then use 50% of the extinction coefficient and surface area density, while keeping the other quantities at 100%).
(3) **In the model levels above the instantaneous local model tropopause**, use the stratospheric data provided here for radiative forcing and heterogeneous chemistry calculations.

Radiation data files for each individual model

CMIP_(YOUR MODEL)_radiation.nc (monthly zonal averaged values for the stratosphere, i.e. 1980 months).
CMIP_(YOUR MODEL)_radiation_average.nc (averaged values for all Januaries to Decembers in the period 1850 – 2014, i.e. 12 individual lat-alt data sets).
CMIP_(YOUR MODEL)_radiation_annual_average.nc (averaged values for the period 1850 - 2014 without seasonal variability, i.e. one single data lat-alt set).

The wave length range is given by:

- wl1_sun: lower boundary of wave length of short solar radiation
- wl2_sun: upper boundary of wave length of short solar radiation
- wl1_earth: lower boundary of wave length of terrestrial radiation
- wl2_earth: upper boundary of wave length of terrestrial radiation

The following quantities are provided as four dimensional (time, altitude and latitude, no of the band) arrays:

(1) ext_sun: extinction coefficient in (1/km) for solar radiation
(2) omega_sun: single scattering albedo for solar radiation
(3) g_sun: asymmetric factor for solar radiation
(4) ext_earth: extinction coefficient in (1/km) for terrestrial radiation
(5) omega_earth: single scattering albedo for terrestrial radiation
(6) g_earth: asymmetric factor for terrestrial radiation

In addition, information specifically for 550 nm is provided (but will not be needed by most modelers).[1]

SAD and mean radius data files

CMIP_sad_1850_2014_V2.nc (monthly zonal averaged values for the stratosphere, i.e. 1980 months).
CMIP_sad_1850_2014_average_V2.nc (averaged values for all Januaries to Decembers in the period 1850 – 2014, i.e. 12 individual lat-alt data sets).
CMIP_sad_1850_2014_annual_average_V2.nc (averaged values for the period 1850 - 2014 without seasonal variability, i.e. one single data lat-alt set).

The following quantities are provided in the data files above:

(1) sad: surface area density in um^2/cm^3, the SAD is corrected based on OPC data.
(2) rmean: mean radius in um (required for some heterogonous reactions).
(3) volume_density: aerosol volume density in um^3/cm^3.
(4) H2SO4_mass: H_2SO_4 density given as number of H_2SO_4 molecules in the aerosol phase per cm^3 of air.
(5) fill_flag: indicating the data source (see Table 1).

[1] Extinction coefficients at 550 nm can be found at ftp://iacftp.ethz.ch/pub_read/luo/CMIP6/IPSL-CM6/OLIVIER/. For models using stratospheric AOD at 550 nm to estimate the radiative forcing of stratospheric aerosol, calculate the stratospheric AOD from the provided extinction coefficients by integrating from the instantaneous local model tropopause to 39.5 km.

Table 1: Flag for the data sources (fill_flag)

-1	Value extrapolated from SAGE data downward to lower altitudes (namely down to 5 km, filled up using last measured value from SAGE).
0	Value based on AER-2D model calculation.
1	SAGE II and the angels sing…
2	CLAES, scaled in the data product using an empirical fit that is consistent with previous validations of this product. Cloud cleared as best as I can do it. Well behaved overall; I think this is a solid method for fixing the Pinatubo gaps.
3	HALOE empirically scaled to 1020 nm following the validation work
4	Equivalent Latitude reconstruction so SAGE II data only. It could possibly be used in fits for size distributions since it is not mixed with anything else or interpolated
5	Auxiliary data fill; the old lidar fill data for the Pinatubo period, it is used in part in the 6/91 to 9/91 period but to a far lesser extent than in previous version
6	Pinatubo June fix; Pinatubo makes a scant appearance in 6/91 from emission mostly before the large eruption. There is no longer any tropical Pinatubo in this month since it was strictly
7	525 estimates from valid 1020 nm data; uses the fits I had for the last go around.
8	CALIPSO converted to 525 nm extinction using a backscatter to extinction ratio of 53. This value is the median ratio of CALIPSO extinction and the scaled OSIRIS data and close to the nominal value we used before (50).
9	OSIRIS, Adam's 525 nm data set scaled by 0.8 in the final data record to pass the eye test.
11	Interpolated (secondary)
12	Interpolation (primary) the linear interpolation I have used for decades to fill holes that are no wider than 2 months.
13	1020 estimates from converted OSIRIS and/or CALIPSO 525 nm data
14	SAM II/SAGE data from 1/1979 through 12/1981
15	Replicated (same value) downward in Lidar period (1982-1984); mostly only below 10 km and at higher latitudes
16	Mostly SAM II and lidar data between 1/1982 and 10/84
17	Mean of OSIRIS and CALIPSO
20	High altitude climatology; average of data between 1984 and 1990 and 1995 and 2005. Try to avoid the random noise at high latitudes that makes Christoph Bruehl unhappy.
21	Patches by hand; it's an art as much as a science
22	Strays filled using neighbors; some holes in otherwise continuous data are patched using adjacent grid spots.
23	Replicated (same value) downward in early OSIRIS/CALIPSO era; mostly below 10 km or so and at higher latitudes
24	Fix some holes in the 525 data where 1020 exists in Pinatubo period
25	Fix some unsightly stuff in the early OSIRIS data in the tropics; it looked like crap so I smoothed it out.

NOTE: in addition, also ALL data during satellite time period at poleward of 80° (i.e. 82.5° and 87.5°) are extrapolated independent of the flag value.

Contact: beiping.luo@env.ethz.ch
Comments: The version before May 31, 2016 has a small bug (one data point has a small negative value in 2006 10). This does not affect the forcing, provided your code does not crash.