Filename conventions:

The following filename convention is used for the CMIP7 historical solar forcing netcdf files:

```
multiple_input4MIPs_solar_CMIP_SOLARIS-HEPPA-CMIP-
<data_version>_gn_<time_period>.nc
```

```
time_period = 18500101-20231231 (CMIP7 reference daily) / 185001-
202312 (CMIP7 reference monthly) / <empty> (picontrol = pre-industrial
fixed input)
```

The pre-industrial control forcing includes time-averaged historical data corresponding to 1850-1873 (SC9+SC10) mean conditions. The reference scenario is based on historical data until Dec 31, 2023.

Transient forcing data are provided in daily resolution, marked by the daily resolution of the time period. A reduced dataset (excluding particle-induced ionization data) is also available in monthly resolution, indicated by the monthly resolution of the time period. The standard PI control forcing (picontrol) represents a scalar time average. Note that all transient forcings (daily, monthly) should be kept constant within a time bin as defined by the time bounds (variable time_bnds).

Note that the alternative SSI dataset for sensitivity experiments (solar-sen-ssi) is provided only for the historical period (until Dec 31, 2023) and with daily resolution. This dataset includes only the TSI and SSI forcing terms.

```
data_version = 4-6  (daily, monthly, picontrol)
data_version = 4-4  (sen-ssi)
```

Dimensions:

```
time = unlimited (63552 for “day” / 2076 for “mon” / 1 for “fx”)
wlen = 3890
plev = 61  %%% not given in “monthly”
glat = 32  %%% not given in “monthly”
nbd  = 2
```

Variables:

Time describing variables

```
double time
        :long_name = "time";
        :standard_name = "time";
        :units = "days since 1850-01-01 00:00:00";
        :calendar = "gregorian";
```

Time corresponds to day 15 of each month in monthly data and to the central date of the 1850-01-01 - 1873-01-28 period in PI control average data.

```
double time_bnds(time, nbd);
        :long_name = "bounds of time bin";
        :units = "days since 1850-01-01 00:00:00";

Short  calyear(time);
        :_FillValue = -1s; // short
        :long_name = "year of Gregorian calendar";
```

```
        :units = "Gregorian_year";

short   calmonth(time);
        :_FillValue = -1s; // short
        :long_name = "month of Gregorian calendar";
        :units = "month";

short   calday(time);
        :_FillValue = -1s; // short
        :long_name = "day of Gregorian calendar";
        :units = "day";
```

Spectral bins (for SSI):

Spectral bins cover the wavelength range 10 – 100,000 nm. In the EUV (below 115 nm) the resolution is 1 nm. Above 115 nm, the spectral bins are adopted from the NRLSSI2 model (Coddington et al., 2015).

```
double wlen(wlen);
        :standard_name = "radiation_wavelength";
        :long_name = "bin center wavelength";
        :units = "nm";

double wlen_bnds(wlen, nbd);
        :long_name = "bounds of wavelength_bin";
        :units = "nm";

double wlenbinsize(wlen);
        :standard_name = "wavelength_interval";
        :long_name = "size of wavelength bin";
        :units = "nm";
```

Vertical grid (for ion-pair production rates due to solar protons, radiation belt electrons, and galactic cosmic rays):

Note that these variables are not provided in monthly data!

```
double plev(plev);
        :long_name = "Pressure level";
        :units = "hPa";
```

Pressure levels cover 1000 – 5.93e-06 hPa in ascending order (TOA to surface). Note that original data of individual sources might cover only a part of the pressure levels.

Geomagnetic latitude bins (for ion-pair production rates due to solar protons, radiation belt electrons, and galactic cosmic rays):

Note that these variables are not provided in monthly data!

```
double glat(glat);
        :long_name = "geomagnetic latitude";
        :units = "degrees_north";
```

```
double glat_bnds(glat, nbd);
        :long_name = "bounds of geomagnetic latitude bin";
        :units = "degrees_north";

float lshell_bnds(glat, nbd);
        :_FillValue = NaNf; // float
        :long_name = "McIlwain L shell";
        :units = "earth radii";
        :formula = "L = 1.018835 / cos(glat_bnds)^2";
```

Geomagnetic latitude bins have been selected to provide sufficient resolution for all particle-induced ionization data. Note that original data of individual sources might cover only a fraction of the bins.

Bin nb.	*Bin center (deg)*	*Lower boundary (deg)*	*Upper boundary (deg)*
1	-82.50	-90.0	-75.0
2	-72.9375	-75.0	-70.875
3	-70.1875	-70.875	-69.5
4	-68.50	-69.5	-67.5
5	-66.25	-67.5	-65.0
6	-63.75	-65.0	-62.5
7	-61.25	-62.5	-60.0
8	-58.75	-60.0	-57.5
9	-55.9375	-57.5	-54.375
10	-52.1875	-54.375	-50.0
11	-47.50	-50.0	-45.0
12	-42.50	-45.0	-40.0
13	-37.50	-40.0	-35.0
14	-32.50	-35.0	-30.0
15	-25.00	-30.0	-20.0
16	-10.00	-20.0	0.0
17	10.00	0.0	20.0
18	25.00	20.0	30.0
19	32.50	30.0	35.0
20	37.50	35.0	40.0
21	42.50	40.0	45.0
22	47.50	45.0	50.0
23	52.1875	50.0	54.375
24	55.9375	54.375	57.5
25	58.75	57.5	60.0
26	61.25	60.0	62.5
27	63.75	62.5	65.0
28	66.25	65.0	67.5
29	68.50	67.5	69.5
30	70.1875	69.5	70.875
31	72.9375	70.875	75.0
32	82.50	75.0	90.0

Solar cycle progression:

Note that these variables are only provided in transient (e.g., reference scenario), not for average data in the standard pi-control forcing!

```
int     scnum(time);
        :long_name = "solar cycle number";
        :units = "1";
```

```
float scph(time);
        :_FillValue = NaNf; // float
        :long_name = "solar cycle phase";
        :units = "radian/pi";
```

Solar cycle phase at time t is defined as 2 x (t - SC_starttime) / (SC_endtime-SC_starttime), hence representing the phase in multiples of π.

```
float ssn(time);
        :_FillValue = NaNf; // float
        :long_name = "Smoothed sunspot number";
        :units = "1";
```

Data are taken from the international sunspot number V2.0 after 1874. Before, an average of SILSO V2 sunspot number and scaled group number of Hoyt and Schatten (1998) is used. A modified Gaussian filter with a full width at half maximum (FWHM) of 365 days has been used for smoothing.

Solar irradiance data/proxies:

```
float ssi(time, wlen);
        :_FillValue = NaNf; // float
        :standard_name = "solar_irradiance_per_unit_wavelength";
        :long_name = "reconstructed spectral solar irradiance at 1 AU";
        :units = "W m^-2 nm^-1";
        :cell_methods = "time: mean wlen: mean";
```

Historical reference dataset (ref): *These SSI data are taken from a preliminary version of the new empirical NASA NOAA LASP (NNL) Solar Spectral Irradiance Version 1 model, NNLSSI1 [Coddington and Lean, 2024]. The NNL solar variability models were previously known as the Naval Research Laboratory (NRL) solar variability models. The model relies on facular brightening and sunspot darkening indices from which the corresponding bolometric changes due to faculae and sunspots are determined. For wavelengths above 115 nm, wavelength-dependent model coefficients are applied to these two bolometric functions to determine the incremental change in SSI due to faculae and sunspots. For wavelengths below 115 nm, where sunspot contribution to extreme ultraviolet change is negligible, the model relies on three inputs: the bolometric faculae function, its temporally smoothed value, and the F10.7 cm solar radio flux. The sunspot darkening time series for the NNLSSI1 models since March 2001 is the near-noon average intensity reduction of white light images made at 676.8 nm relative to the local emission reported by all Global Oscillation Network Group (GONG) stations. From 1976 to 2001, the sunspot darkening index is derived from sunspot areas and locations reported by the SOON network. Prior to 1976, the sunspot darkening index is 66% of the valued measured by the Royal Greenwich Observatory (RGO) to account for systematic differences between RGO and SOON measurements. The facular brightening time series that the NNLSSI1 model relies on after 2017 is the GOES MgII index, from 2000 to 2017 the University of Bremen MgII index scaled to the GOES level, and from 1996 to 2000 the chromospheric proxy that Lean et al. (2000) constructed from multiple MgII datasets cross-calibrated with the CaII index. From 1954 to 1976 the facular brightening time series is constructed using multiple linear regression of the F10.7 cm solar radio flux (smoothed and daily values) at lags of 0, 27 and 54 days. From 1874-1954, the facular brightening time series is constructed from the multiple linear regression of the SILSO V2 Sunspot number record (smoothed and daily values) at lags of 0, 27, and 54 days. The cycle minima in the facular index time series prior to 1954 are adjusted using the flux transport simulations of Wang & Lean App 2021 (based on an average of SILSO V2 sunspot number and scaled group number of Hoyt and Schatten, 1998 that was shown to agree with the cosmogenic isotope record of Usoskin (2017)). Since NNLSSI1 has only yearly averages before 1874 (based on an average of SILSO V2 sunspot number and scaled group number of Hoyt and Schatten, 1998), sub-yearly variations were extrapolated by using an autoregressive model.*

Alternative SSI dataset for sensitivity experiments (sen-ssi): *This dataset was produced using the physics-based semi-empirical SATIRE (Spectral And Total Irradiance REconstruction) model. SATIRE divides the solar surface into 4 different components (bright faculae, dark sunspot umbrae and sunspot penumbrae, as well as the quiet Sun; Krivova et al. 2003 10.1051/0004-6361:20030029). The magnetic field emerging at the Sun's surface modifies its radiative properties (so sunspots are darker than the surrounding non-magnetic quiet surface while faculae are brighter). Thus, the continuously changing distribution of the magnetic features on the Sun leads to irradiance variability, and to reconstruct solar irradiance variability, knowledge about: (a) the distribution of the magnetic features on the Sun at a given time, and (b) the time-independent but wavelength and solar-disc-position-dependent brightness contrasts of these features is required. (a) The most reliable information on the positions and areas of the surface magnetic features is extracted from the full-disc magnetograms of the Sun, here used since 1976 (Yeo et al. 2014 10.1051/0004-6361/201423628; revised and updated by Chatzistergos et al. in preparation). Before 1976, sunspot observations are used as input into a physics-based model to reconstruct the evolution of the solar surface magnetic field, and thus the information on the surface coverage by faculae and sunspots (Wu et al. 2018 10.1051/0004-6361/201832956). Over the period 1874-1976, sunspot areas and positions (a cross-calibrated composite of RGO, Kislovodsk and SOON data by Balmaceda et al. 2009 10.1029/2009JA014299 going back to 1874) are used. Before that, SATIRE relies on group sunspot number (Chatzistergos et al. 2017 10.1051/0004-6361/201630045) to first construct a statistical synthetic record of sunspot areas and positions (Dasi-Espuig et al. 2016 10.1051/0004-6361/201527993), which is then eventually used to compute the facular and sunspot distributions. (b) The brightness contrasts of faculae and sunspots with respect to the quiet Sun are computed using a radiative transfer code from the corresponding semi-empirical solar model atmospheres (Unruh et al. 1999). They are time-independent but depend on the wavelengths. This allows both SSI and TSI to be self-consistently computed in one run (that is TSI is the spectrum-integrated SSI), without additional free parameters or tuning to SSI observations. Note that this dataset differs from the standard SATIRE reconstructions, which can be found in https://www2.mps.mpg.de/projects/sun-climate/data.html*

SSI data in the CMIP7 solar forcing files cover the wavelength range 10 – 100,000 nm *(consistent with the CMIP6 wavelength range and bins). Note that the wavelength range of the original NNLSSI data is 0 – 200,000 nm and that of the original SATIRE data is 115 – 162,500 nm.*

Note that SSI data represent wavelength bin averages. A MATLAB routine how to read and integrate the SSI data to the respective radiation bands in the respective CMIP model, is provided at solarisheppa.geomar.de/cmip6.

```
float tsi(time);
       :_FillValue = NaNf; // float
       :standard_name = "solar_irradiance";
       :long_name = "reconstructed total solar irradiance at 1 AU";
       :units = "W m^-2";
       :cell_methods = "time: mean";
```

TSI is calculated from the integral of SSI along wavelength in the 0 -100,000 nm range.

```
float f107(time);
       :_FillValue = NaNf; // float
       :long_name = "Adjusted F10.7 solar radio flux";
       :units = "10^-22 W m^-2 Hz^-1"; // (= "sfu")
       :cell_methods = "time: mean";
```

F10.7 data has been taken from NGDC adjusted values (ftp://ftp.ngdc.noaa.gov/STP/) and for the 2015-2023 period from GFZ/Potsdam. Missing (pre-1947) values are taken from the reconstruction incorporated in NNLSSI1 [Coddington and Lean, pers. comm.].

Geomagnetic proxy data:

```
float ap(time);
        :_FillValue = NaNf; // float
        :long_name = "daily planetary Ap index";
        :units = "nT";
        :cell_methods = "time: mean";
```

Ap data for the period 1932-2014 have been taken from NGDC (ftp://ftp.ngdc.noaa.gov/STP/) and for the 2015-2023 period from GFZ/Potsdam. Ap data prior to 1932 have been constructed from aa (1878-1932) and Ak (1850-1877), provided by the International Service of Geomagnetic Indices (http://isgi.unistra.fr/), using a monthly piecewise polynomial fit. Ap data is required to generate the odd nitrogen upper boundary condition for chemistry climate models to account for the EPP indirect effect (polar winter descent of particle generated NOx into the model domain). A routine (IDL and MATLAB) for generation of the odd nitrogen upper boundary condition is provided as well at https://solarisheppa.kit.edu/75.php.

```
float kp(time);
        :_FillValue = NaNf; // float
        :long_name = "daily planetary Kp index";
        :units = "1";
        :cell_methods = "time: mean";
```

The daily Kp index for the period 1932-2014 has been taken from NGDC (http://isgi.unistra.fr/) and for the 2015-2023 period from GFZ/Potsdam. For 1868-1931, it was estimated by using monthly piecewise polynomial aa-Ap fits to estimate the 3-hourly ap-index values from the aa index values. These 3-hour ap values were then converted to the corresponding Kp indices, from which the daily mean was calculated. For the period 1850-1867, daily estimates of Ap, derived from Ak, are directly converted into daily Kp.

Particle-induced ionization data:

Note that these variables are not provided in monthly data! Ion pair production rates (IPR) are provided in units of ion pairs g-1 s-1. Conversion to ion pairs cm-3 s-1 (by multiplying with mass density) should be done within the atmospheric models ideally at each time step (but at least once per day).
Ionization by energetic particles (expressed by the ion pair production rates iprp, iprm, and iprg) leads to productions of reactive nitrogen (NOy) and odd hydrogen (HOx). As a basic approach, we recommend to consider these NOx and HOx productions by using the parameterizations provided by Porter et al. (J. Chem. Phys. 65, 154, 1976) and Solomon et al. (Planet. Space Sci. 8, 885, 1981), respectively.

Following Porter et al. (1976) it is assumed that ~1.25 N atoms are produced per ion pair. This study also further divided the proton impact of N atom production between the ground state N(4S) (~45% or ~0.55 per ion pair) and the excited state N(2D) (~55% or ~0.7 per ion pair). If a model does not include the excited state of atomic nitrogen in their computations, the NOy production from EPP can still be included by assuming that its production is instantaneously converted into NO, resulting in a N(4S) production of 0.55 per ion pair and a NO production of 0.7 per ion pair.

The production of HOx relies on complicated ion chemistry that takes place after the initial formation of ion pairs. Solomon et al. (1981) computed HOx production rates as a function of altitude and ion pair production. Each ion pair typically results in the production of around two HOx constituents in the upper stratosphere and lower mesosphere. In the middle and upper mesosphere, an ion pair is computed to produce less than two HOx constituents per ion pair.

If available, the use of more comprehensive parametrizations for productions of individual HOx (OH and H) and NOy (N^4S, N^2D, NO, NO_2, NO_3, N_2O_5, HNO_2, and HNO_3) compounds (e.g. Verronen and Lehmann, Ann. Geophys., 31, 909-956, 2013; Nieder et al., J. Geophys. Res. Space Physics, 119, 2014) is encouraged. Similarly, if atmospheric models include detailed cluster ion chemistry of the lower *ionosphere (D region), then the ionization rates should be used to drive the production rates of the primary ions (N_2^+, N^+, O_2^+, O^+) and neutrals (N, O) produced in particle impact ionization/dissociation [Sinnhuber et al., Surv. Geophys. 33, 1281, 2012]. In these cases, we encourage the modeling centers to carefully document the approaches they adopt.*

For the projection of IPR data (as function of geomagnetic latitude) onto geographic coordinates, we recommend to use a geomagnetic field model considering variations in the vertical and time domains, in particular:

- *1850-1900: gufm1 (https://geomag.colorado.edu/gufm1.html). Reference: Jackson et al., Phil. Trans. R. Soc. Lond. A 358, 957, 2000.*
- *1900-2020: IGRF13 (http://www.ngdc.noaa.gov/IAGA/vmod/igrf.html). Reference: Alken et al., Earth, Planets and Space 2021, 73:49, 2021.*
- *2020-2023: linear extrapolation of IGRF13.*

A MATLAB routine for projection of geomagnetic latitudes onto geographic coordinates, following these recommendations, will be made available soon at https://solarisheppa.kit.edu/75.php.

```
float iprp(time, plev, glat);
       :_FillValue = NaNf; // float
       :long_name = "Ion pair production rate by solar protons";
       :units = "g^-1 s^-1";
       :cell_methods = "time: mean";
```

Interplanetary Monitoring Platform (IMP) series of satellites provided measurements of proton fluxes from 1963-1993. IMPs 1-7 were used for the fluxes from 1963-1973 (Jackman et al., 1990) and IMP 8 was used for the fluxes from 1974-1993 (Vitt and Jackman, 1996). The National Oceanic and Atmospheric Administration (NOAA) Geostationary Operational Environmental Satellites (GOES) were used for proton fluxes from 1994 onwards (see, e.g., Jackman et al., 2005a, 2014). An artificial proton forcing was created for the 1850-1962 and future periods by randomly projecting sequences of individual solar cycles covered by the GOES/IMP-derived dataset. The valid range of proton IPR data is 60-90 deg geomagnetic latitude (both hemispheres) and 7.26e-5 – 1000 hPa. Outside this latitude range proton IPR data are set to zero. Standard PI control data (frequency = "fx") represent the mean values of the 1850-1873 period.

```
float iprg(time, plev, glat);
       :_FillValue = NaNf; // float
       :long_name = "Ion pair production rate by cosmic rays";
       :units = "g^-1 s-1";
       :cell_methods = "time: mean";
```

See full description of galactic cosmic ray (GCR) induced ionization in (Usoskin and Kovaltsov, J. Geophys. Res., 111, D21206, 2006) and (Usoskin et al., J. Geophys. Res., 115, D10302, 2010; and Usoskin at al., JSWSC 2024). GCR IPR was calculated was calculated by the CRAC:CRII model_version 3. The valid pressure range of GCR IPR data is 0.01046 hPa – 1000.0 hPa. For smaller pressure levels, GCR IPR data are set to zero. Transient data (monthly and daily) have been generated by interpolating the original annual data to the respective time resolution. Standard PI control data (frequency = "fx") represent the mean values of the 1850-1873 period.

```
float iprm(time, plev, glat);
       :_FillValue = NaNf; // float
       :long_name = "Ion pair production rate by MEE";
       :units = "g^-1 s^-1";
       :cell_methods = "time: mean";
```

The mid-energy electron IPR data set has been calculated using an electron flux model (van de Kamp et al., pers. comm.) and an atmospheric ionization parameterization (Fang et al., Geophys. Res. Lett., 37, *L22106, doi:10.1029/2010GL045406, 2010). The electron flux model is fit to equivalent isotropic precipitating electron fluxes estimated from observations of the MEPED detectors onboard POES NOAA-6, 8, 10, 12 and 15 satellites during 1979 to 2023 (Asikainen, 2024 pers. comm.). These data have been instrumentally calibrated and corrected for proton contamination (Asikainen and Mursula, (2013), J. Geophys. Res, 118, https://doi.org/10.1002/jgra.50584) as well as temporally and spatially corrected for background noise, changing satellite orbits and differences in satellite instrumentation Asikainen and Ruopsa (2019), J. Geophys. Res. Space Phys., 124, https://doi.org/10.1029/2018JA026214 and Asikainen (2019), J. Geophys. Res. Space Phys., 124, https://doi.org/10.1029/2019JA026699). The model can be seen as an updated version of the model which was used in CMIP6 and described by van de Kamp et al. (J. Geophys. Res. Atmos., 121, doi:10.1002/2015JD024212, 2016). The valid range of MEE IPR data is 45-90 deg geomagnetic latitude (both hemispheres) and 5.96e-6 - 1 hPa; outside this range MEE IPR data are set to zero. The flux model depends on the geomagnetic Ap index. The CMIP7 reconstruction of Ap has been used to cover the whole period between 1850 and 2023. Standard PI control data (frequency = "fx") represent the median values of the 1850-1873 period.*

Contact:

Please notify us if you encounter any problems with the data and/or their description.

Bernd Funke:	bernd[at]iaa.es
Thierry Dudok De Wit:	ddwit[at]cnrs-orleans.fr
Margit Haberreiter:	Margit.Haberreiter[at]pmodwrc.ch
Daniel Marsh:	D.Marsh[at]leeds.ac.uk