03.01.2017

Filename conventions:

The following filename convention is used for the CMIP6 solar forcing netcdf files:

```
solarforcing-<scenario>-<frequency>_input4MIPs_solar_CMIP_SOLARIS-
HEPPA-<data_version>_gn_18500101-22991231.nc
```

```
scenario = picontrol (pre-industrial) / ref (CMIP6 reference)/
ext (extreme Maunder Minimum-like)
```

*The pre-industrial control forcing (**picontrol**) includes time-averaged historical data corresponding to 1850-1873 (SC9+SC10) mean conditions.*
*Both the reference scenario (**ref**) and the extreme solar minimum scenario (**ext**) are based on historical data before Dec 31, 2014 and hence are identical in this time period. For the future period, the solar forcing is derived from the heliospheric potential. We forecast the latter by using 9400 years of reconstruction of solar activity from cosmogenic isotopes [Steinhilber et al, PNAS 109 (2012), doi:10.1073/pnas.1118965109]. The reference scenario is the weighted average of 3 forecasts [Matthes et al., Geosc. Model. Developm. (2016)] obtained by: (i) analogue forecast (superposed epoch), (ii) deterministic forecast relying on periodicities in the observed heliospheric potential, and (iii) autoregressive model. The forecast skill does not exceed 70 years. Because we provide the most likely scenario, the forcing data does not decay towards a climatological mean, but keeps on varying. The extreme (low solar activity) scenario corresponds to the lower 5 percentile of the distribution of forecasts, and its level is equivalent to that of a Maunder minimum.*

```
frequency = day / mon /fx
```

*Transient forcing data are provided in daily (**day**) resolution. A reduced dataset (excluding particle-induced ionization data) is also available in monthly (**mon**) resolution. The standard PI control forcing (solarforcing_picontrol_fx_3.2) represents a scalar time average (**fx**). We also provide a variable PI control forcing (solarforcing_picontrol_day_3.2, time coverage 1 Jan 1850 – 9 Sep 2053), which includes an 11-year solar cycle but without longterm trend, for sensitivity studies. **Note that the latter is not officially part of a CMIP6 MIP proposal!** Please also note that all transient forcings (frequency = day, mon) should be kept constant within a time bin as defined by the time bounds (variable time_bnds).*

```
data_version = 3-2
```

The current data version is "final" and can be used in model simulations.

Dimensions:

```
time = unlimited (164359 for "day" / 5400 for "mon" / 1 for
       "fx" / 74397 for "picontrol_day")
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";

double calyear(time);
       :long_name = "year of Gregorian calendar";

double calmonth(time);
       :long_name = "month of Gregorian calendar";

double calday(time);
       :long_name = "day of Gregorian calendar";
```

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";
```

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 and extreme scenarios), not for average data in the standard pi-control forcing!

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

The transient forcing data sets cover cycle numbers 9-50. Future cycles are constructed from historical cycles by scaling (maintaining the shape of intra-cycle variability of the historical cycles). The following historical cycles are used for construction of future cycles (starting on 2015-01-01)

Current cycle	*Historical cycle*	*Start date current cycle*	*Start date historical cycle*
24	12	2015-01-01	1883-02-01
25	13	2020-02-02	1890-01-28
26	14	2031-12-18	1901-12-14
27	15	2043-06-19	1913-06-15
28	12	2053-09-10	1878-12-13
29	13	2064-10-26	1890-01-28
30	14	2076-09-10	1901-12-14
31	15	2088-03-12	1913-06-15
32	16	2098-06-04	1923-09-07
33	17	2108-07-05	1933-10-07
34	18	2118-11-21	1944-02-23
35	19	2129-01-16	1954-04-20
36	20	2139-07-02	1964-10-03
37	21	2150-12-04	1976-03-07
38	22	2161-04-21	1986-07-24
39	23	2171-05-21	1996-08-22
40	24 12	2183-08-19 2189-11-07	2008-11-20 1883-02-01
41	13	2194-10-31	1890-01-28
42	14	2206-09-16	1901-12-14
43	15	2218-03-18	1913-06-15
44	12	2228-06-09	1878-12-13
45	13	2239-07-26	1890-01-28
46	14	2251-06-10	1901-12-14
47	15	2262-12-10	1913-06-15
48	16	2273-03-03	1923-09-07
49	17	2283-04-03	1933-10-07
50	18	2293-08-19	1944-02-23

*The variable PI control dataset (**non-standard!**) covers the time period from 1.1.1850 until 9.9.2053 (end of solar cycle 27). This dataset can be extended to cover 1000 years by multiple repetition of the solar cycle sequence 12–27. The first 450 years of the resulting forcing time series are consistent in solar cycle phase and short-term fluctuations with the REF and EXT datasets.*

```
float scph(time);
        :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);
        :long_name = "Smoothed sunspot number";
        :units = "1";
```

Data are taken from the international sunspot number V1.0 (from http://www.sidc.be/silso/versionarchive), even though a newer version 2.0 recently came out (Clette et al., 2014). *For future scenarios (Jan 1, 2015 - Dec 31, 2299), multi-decadal variability is derived from the heliospheric potential. Sunspot number (SSN) variations on shorter time scales are taken from the corresponding past solar cycles and are scaled to a comparable cycle-average level. 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);
        :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";
```

These SSI data are the arithmetic mean of two model datasets: (i) the semi-empirical model SATIRE-T/S [Yeo et al., Astron. & Astroph., 570 (2014), doi:10.1051/0004-6361/201423628], and (ii), the empirical model NRLSSI2 [Coddington et al., Bulletin of the Am. Met. Soc., (2015), doi:10.1175/BAMS-D-14-00265.1]. For historical data (Jan 1, 1850 - Dec 31, 2014) both models rely on either or several of: international sunspot number V1.0, sunspot area distribution (after 1882), solar photospheric magnetic field (after 1974), and the MgII index (after 1978). For future scenarios (Jan 1, 2015 - Dec 31, 2299) both models rely on SSN only, whose multi-decadal variability is derived from the heliospheric potential. SSI variations on shorter time scales are taken from the corresponding past solar cycles and are scaled to a comparable cycle-average level of activity by means of a dedicated scaling procedure [Matthes et al., Geosc. Model. Developm. (2016)]. EUV channels (10-115 nm) have been added to the SSI model data. Since NRLSSI2 has yearly averages only before 1882, sub-yearly variations were derived from the sunspot number by using an autoregressive model. In both SATIRE-T/S and NRLSSI2, the SSI in spectral bins from 10.5-114.5 nm were derived by nonlinear regression from the SSI in the 115.5-188.5 nm band, based on TIMED-SEE data.
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 as well at solarisheppa.geomar.de/cmip6.

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

TSI is calculated from the integral of SSI along wavelength.

```
float f107(time);
        :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/). Missing (pre-1947) and future data have been constructed by regression to SSI principal components with non-linear corrections. See Matthes et al., Geosc. Model. Developm. (2016) for more details.

Geomagnetic proxy data:

```
float ap(time);
        :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/). 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. For future scenarios (Jan 1, 2015 - Dec 31, 2299), Ap data is constructed from historical data by means of a dedicated scaling procedure based on sunspot number. See Matthes et al., Geosc. Model. Developm. (2016) for more details. Ap data is used to generate a 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 solarisheppa.geomar.de/cmip6.

```
float kp(time);
        :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/). 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. Future Kp data (Jan 1, 2015 - Dec 31, 2299) is constructed historical data by means of a dedicated scaling procedure based on Ap. See Matthes et al., Geosc. Model. Developm. (2016) for more details.

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 (protons), iprm (mid-energy electrons), and iprg (galactic cosmic rays) 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, NO2, NO3, N2O5, HNO2, and HNO3) 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 (N2+, N+, O2+, 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 (http://jupiter.ethz.ch/~cfinlay/gufm1.html). Reference: Jackson et al., Phil. Trans. R. Soc. Lond. A 358, 957, 2000.*
- *1900-2015: IGRF12 (http://www.ngdc.noaa.gov/IAGA/vmod/igrf.html). Reference: Thébault et al., Earth, Planets and Space 2015, 67:79 , 2015.*
- *2015-2300: geodynamo model forecast by Julien Aubert (http://www.ipgp.fr/~aubert/Julien_Aubert,_Geodynamo,_IPG_Paris/Research/Entrees/2016/1/28_Operational_modelling_of_the_geodynamo.html) until 2115. Afterwards, the secular variation values for year 2115 should be used. Reference: Aubert, J, Geophysical Journal International 203 (3), 1738–1751, doi:10. 1093/gji/ggv394, 2015.*

A MATLAB routine for projection of geomagnetic latitudes onto geographic coordinates, following these recommendations, is provided as well at solarisheppa.geomar.de/cmip6.

```
float iprp(time, plev, glat);
       :standard_name = "prot_ion_pair_production_rate";
       :long_name = "Ion pair production rate by solar protons";
       :units = "ion pairs 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 1994-2014 (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. Proton IPR data in the reference and extreme scenario, as well as in the non-standard variable PI control forcing, are identical. The valid range of proton IPR data is 60-90 deg geomagnetic latitude (both hemispheres) and 7.26e-5 – 63.1 hPa. Outside this range zero fill values have been applied and should be used in calculations. Standard PI control data (frequency = "fx") represent the median values of the 1850-1873 period.

```
float iprg(time, plev, glat);
       :standard_name = "gcr_ion_pair_production_rate";
       :long_name = "Ion pair production rate by cosmic rays";
       :units = "ion pairs 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 etal., J. Geophys. Res., 115, D10302, 2010.). GCR IPR was calculated by applying the phi values as provided by the CMIP6 scenarios. The valid pressure range of GCR IPR data is 0.01046 – 1000.0 hPa. Outside this range zero fill values have been applied and should be used in calculations. 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);

        :standard_name = "mee_ion_pair_production_rate";
        :long_name = "Ion pair production rate by MEE";
        :units = "ion pairs 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., submitted, 2015JD024212, 2015) 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 observations of the MEPED detectors onboard the POES satellites, between 2002 and 2012, following the concepts outlined in Whittaker et al. (J. Geophys. Res. Space Physics, 119, 8784-8800, doi:10.1002/2014JA020446, 2014). The flux model depends on the geomagnetic Ap index. The CMIP6 reconstruction of Ap has been used to cover the whole period between 1850 and 2299. The valid range of MEE IPR data is 44.5-71.5 deg geomagnetic latitude (both hemispheres) and 7.26e-5 – 10 hPa. Outside this range, zero fill values have been applied and should be used in calculations. 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.

Katja Matthes: kmatthes[at]geomar.de
Bernd Funke: bernd[at]iaa.es