The SMILES L2 Climatology

Yasuko Kasai
Remote Sensing Fundamentals Laboratory
NICT
Contact: smiles-data-access@ml.nict.go.jp
November 9, 2012

Contents

Introduction

This report summarizes the work and results for the SMILES L2 trace gas climatology. The period of work started at the first of October, 2011 and ended on 30th of November, 2012.

SMILES = [Superconducting SubMIllimeter-wave Limb-Emission Sounder] The SMILES sensor is mounted aboard the International Space Station (ISS) and was performing passive micro-wave measurements on various trace gas species in the strato- and mesosphere. The sophisticated radiometer produces highly accurate measurements with very low observation noise, in comparison to the conventional. Further more, being placed on the ISS with a non-sun-synchronous orbit, SMILES measurements comprise observations at all local solar times (LST) and thus solar zenith angles (SZA).

A climatology for SMILES data should make best use of the unique advantages of the data, but also has to consider problems like the irregular spatial and temporal distribution of the data and periods of lacking data due to the SMILES observation band cycle

The primary results of the climatology are a diurnal variation climatology and a zonally averaged climatology. Where zonal averages are calculated for geographical latitude and also for equivalent latitude. Additionally daily averaged products for long-lived species provide a seasonal overview.

Section 1 describes the SMILES data used for processing and the methods and techniques applied. Section 2 presents the current state of results and provides an overview of the climatology processing code.

1 Climatology processing

1.1 SMILES L2 data

Data Version: SMILES L2 version 2.1.5 was used for climatology processing. For ClO the updated/ reprocessed version 2.1.5 was used. (ClO had to be reprocessed after discovery of L2 calculation problem.)

Species: The following species have been processed so far.

Table 1: SMILES Climatology species. Clim. type abbreviations stand for: LAT - zonal averaged, EQL - equivalent latitude averaged, LST - local solar time averaged, SZA - solar zenith angle averaged.

Species	**Band**	**Effective altitude**	**Specie type**	**Clim. type**	**VMR limits**	χ^2
O_3	A & B	18 – 95 km	dynamical stratosphere, photochemical mesosphere	LAT, EQL, LST, SZA	-4e^{-6} / 19e^{-6}	0.8
$H^{37}Cl$	A	20 – 90 km	dynamical	LAT, EQL, SZA	-0.8e^{-9} / 4e^{-9}	1.2
$H^{35}Cl$	B	20 – 95 km	dynamical	LAT, EQL, SZA	-4e^{-9} / 10e^{-9}	0.8
HNO_3	C	20 – 80 km	dynamical	LAT, EQL, LST, SZA	-7.5e^{-9} / 25e^{-9}	1.2
ClO	C	20 – 80 km	photochemical	LAT, EQL, SZA	-1e^{-9} / 3e^{-9}	0.8
BrO	A & C	30 – 60 km	photochemical	LAT, EQL, SZA	-80e^{-12} / 100e^{-12}	1.5
HO_2	C	20 – 90 km	photochemical	LAT, EQL, SZA	-100e^{-9} / 100e^{-9}	0.65
HOCl	A	30 – 70 km	photochemical	LAT, EQL, SZA	-1e^{-9} / 10e^{-9}	1.0
Temp	A & B	18 – 60 km	derived from O_3	LAT, EQL, LST, SZA	50K / 400K	0.8

1.2 Data binning and averaging

1.2.1 General data processing

This is a brief summary and overview of data handling and processing steps in the climatology processing chain.

Note: Python files are denoted as `some_python_filename.py`, whereas python functions are denoted with the filename, containing the function, followed by the function name, as in `some_python_filename.py:some_py_function`.

1. The raw SMILES L2 data is read from the daily L2 HDF-files per species and month. The function `read_multiple_hdf.py:read_multiple_hdf` can for example read all HDF-files of one month of a certain species. This function makes use of the `readhdf` function (Sagawa-san) and the `h5py-module`.

   The return value is a list of days of SMILES L2 data, stored in a list of daily dictionaries.

2. The function `read_multiple_hdf.py:merge_l2_list` breaks the daily list apart and concatenates the SMILES profiles stored in the daily dictionaries to one dictionary containing the full (e.g. monthly) data set (these are profiles × altitude matrices). The profiles get sorted according to time. Additionally some meta data is read from the `coremetadata.0`.

3. `read_multiple_hdf.py:check_l2_data` performs species dependent sanity checks on the L2 data dictionary. Thereby several parameters are checked and the result of the checks is written to a so called quality flag matrix (`Quality_flg`: profiles × altitude levels), containing boolean values and stating if the individual measurement is reliable or not. See Section 1.2.4.

   Note: The L2 data itself is not getting manipulated by this step, only the quality flag matrix is generated.

4. `calc_climavg.py:calc_climavg` is the major function for data binning and averaging. Different optional keyword arguments enable different types of averaging. The averaging bins can be specified as: latitude (LAT), equivalent latitude (EQL), solar zenith angle (SZA) or local solar time (LST) (see function documentation). Mandatory input to this function is the L2 python dictionary, including the quality flag matrix. All other inputs are optional.

5. As a first step, in `calc_climavg.py:calc_climavg` the quality flag matrix is applied to filter the VMR values and the measurement response.

6. Next step is the vertical interpolation of the VMR, measurement response, L2Precision (error), altitudes, pressure and temperature to the vertical level vector. This can either be a pressure level vector, or a altitude level vector or potential temperature. The default is the SPARC CCMval pressure levels, specified from 300 hPa to 0.001 hPa (the vector was modified to fit the SMILES vertical range, several levels in the middle and upper mesosphere have been appended.)

7. By using the optional keyword arguments, `szalim, absszalim, lstlim` and `latlim`, pre-filtering options can be specified. E.g. `absszalim = [100, 180]`, selects only nighttime measurements of either negative or positive SZAs, between 100°and 180°. Technically all keywords can be used in the same call, but this usually doesn't make much sense, or will reduce the number of selected measurements dramatically.

8. The data binning is done according to the optional keyword `lat_type` (default `'lat'`, or `'eql'`, or `'sza'`, or `'lst'`). This means the data gets rather binned to

latitudinal bins, to equivalent latitudinal bins, to solar zenith angle bins or to local solar time bins, respectively. The data is binned in two dimensions. The variable `lat_type` determines the primary averaging bins. According to the primary bins the secondary bins get defined. The secondary bins can be thought of as the horizontal bins, while the primary are the vertical bins. This description fits well, if SZA or LST is the primary bin type and latitude is the secondary bin type. If latitude or equivalent latitude is the primary bin type, the secondary will be set to LST.

9. If there are at least 30 profiles in a primary averaging bin, the VMRs are filtered for outlier values being $3 \times$ MAD bigger or smaller than the median of VMRs in this primary bin.

   Then statistical values are calculated for each grid box. This is a 3 dimensional tensor (primary bins $\times$ secondary bins $\times$ vertical levels). Further statistical values are calculated along the secondary bins, collapsing the 3d tensor to a 2d matrix, with primary bins $\times$ vertical levels. Several statistical estimators are provided. E.g. number of elements, median, median absolute deviation. See also Section 1.2.2.

   Example: Figure 1 shows an example of ozone data binned to latitude. $3 \times$ MAD outlier filtering is applied vertical level wise on the data of all profiles in the 15° N – 20° N latitude bin. Then the profiles will be binned to 1 hour local solar time bins. Data is first averaged (median) in each local solar time bin and then the median of all local solar time medians is taken bin. This secondary binning prohibits, that many measurements from a specific latitude region will dominate averaged value in the entire primary SZA bin.

10. Finally the averaged output is written to the python return dictionary, together with additional information about the filtering criteria.

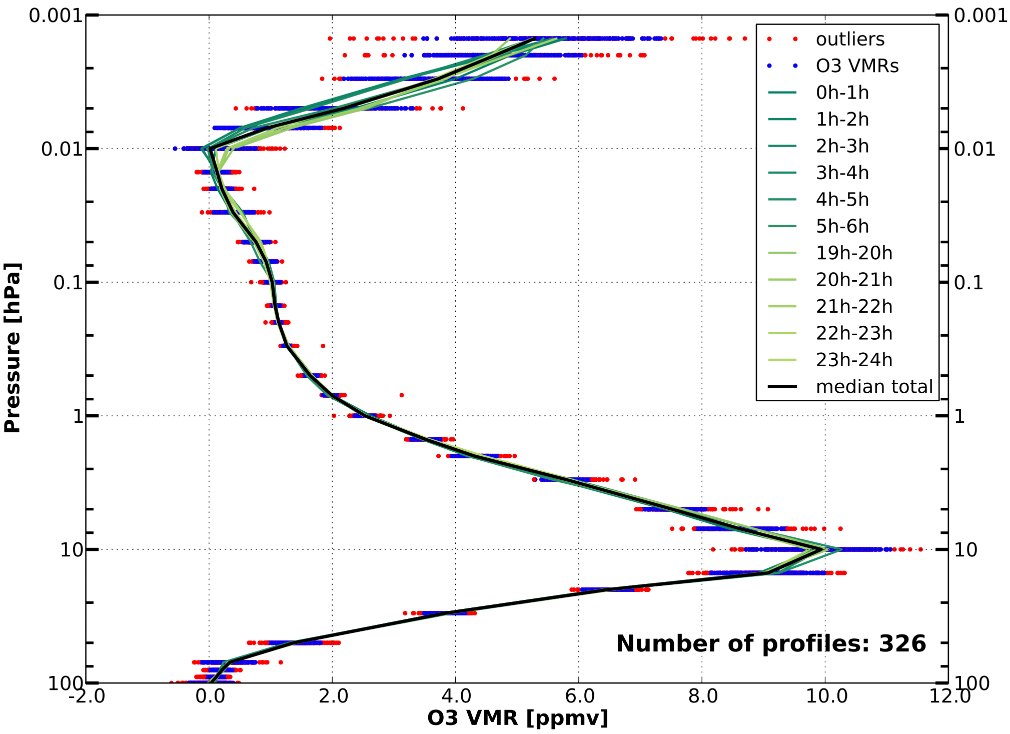


Figure 1: O_3 VMRs of band B in march 2010 for a 5°latitude bin (15° N–20° N), night time profiles only (SZA 100°–180°). Green to yellowish lines represent the median of the 1h LST bins. Thick black line is the median of the of the 1h medians. Please also refer to legend.

The functionality for calculating daily averages of data, basically applies the same steps as described above, but binning and averaging the data on a daily basis. See `calc_daymean.py:calc_daymean`.

1.2.2 Statistics

In order to average data sets, median statistics, instead of mean statistics were used. The benefit of the median is its robustness towards outliers. A statistical estimator, like the mean, is influenced by even a single arbitrary outlier value. Therefore the so called breakdown point of the mean-estimator is 0%. The Breakdown point is defined "as the proportion of incorrect observations (i.e. arbitrarily large observations) an estimator can handle before giving an arbitrarily large result." (`http://en.wikipedia.org/wiki/Breakdown_point#cite_note-2`). The median-estimator has a breakdown point of 50%, which corresponds to the possible maximum of a breakdown point value.

Regarding SMILES L2 data, extreme outlier values can be expected to be found in the data. However, a comparison of mean and median statistics, applied on monthly averaged SMILES L2 data proved to show less than 1% difference in average. Although maximum deviations in the example data (SMILES v2.1.5 O3 band B March) reach up to 2676%.

The median absolute deviation (MAD) is a robust measure of variability of a data distribution. The MAD value is comparable to the standard deviation, but using median-statistics instead of mean-statistics.

$$\mathrm{MAD} = \mathrm{median}_i\ (\ |X_i - \mathrm{median}_j(X_j)|\ ) \tag{1}$$

Where $\mathrm{median}_j(X_j)$ is the median of the entire data distribution and i is the index of each absolute difference of all data samples to the median. Finally the median of all absolute deviations yields the MAD. The MAD is used in `calc_climavg.py:calc_climavg` and `calc_daymean.py:calc_daymean` to determine outliers and remove them from the data. A sample value is considered an outlier, if its absolute difference to the median of the data is bigger than $3 \times \mathrm{MAD}$.

1.2.3 Data weighting

Data weighting has NOT been applied to the climatology processing, but its applicability has been tested. The data could be weighted by the L2Precision value in the SMILES L2 data, as follows:

$$\text{weighted-VMR} = \frac{\sum_{i=1}^{N_{alt}} \mathrm{VMR}_i * (\mathrm{L2Precision}_i)^{-1}}{\sum_{i=1}^{N_{alt}} (\mathrm{L2Precision}_i)^{-1}} \tag{2}$$

Where N_{alt} is the number of measurements in a certain altitude level and $(\mathrm{L2Precision}_i)^{-1}$ is the inverse of each L2 precision value. The general idea is to weight each VMR value by its precision value and thus giving less impact to potential outliers. (What exactly the precision value describes, I don't remember, ask Sagawa-san!). This approach to data weighting is done by the ACE-FTS community (ref. A.Jones).

The results of some brief comparisons between the median and the weighted data are displayed in Figure 2. As a conclusion of this investigation, I decided not to use data weighting since the $3\times$MAD filtering is already taking care of outliers and the median represents the average of the data more accurately than the weighted profile.

1.2.4 Quality filtering

Quality filtering is applied with the function, `read_multiple_hdf.py:check_l2_data`. Filtering thresholds are species dependent and were determined by statistical analysis of monthly averaged sample data or from suggestions by the retrieval people. Refer to Figure 3 for an example of statistical determination of thresholds. The boolean `Quality_flg` matrix is generated by checking the following:

- Check for VMR limits. Maximum and minimum thresholds for unphysical VMR values are defined. VMR samples outside these boundaries are flagged `false`.

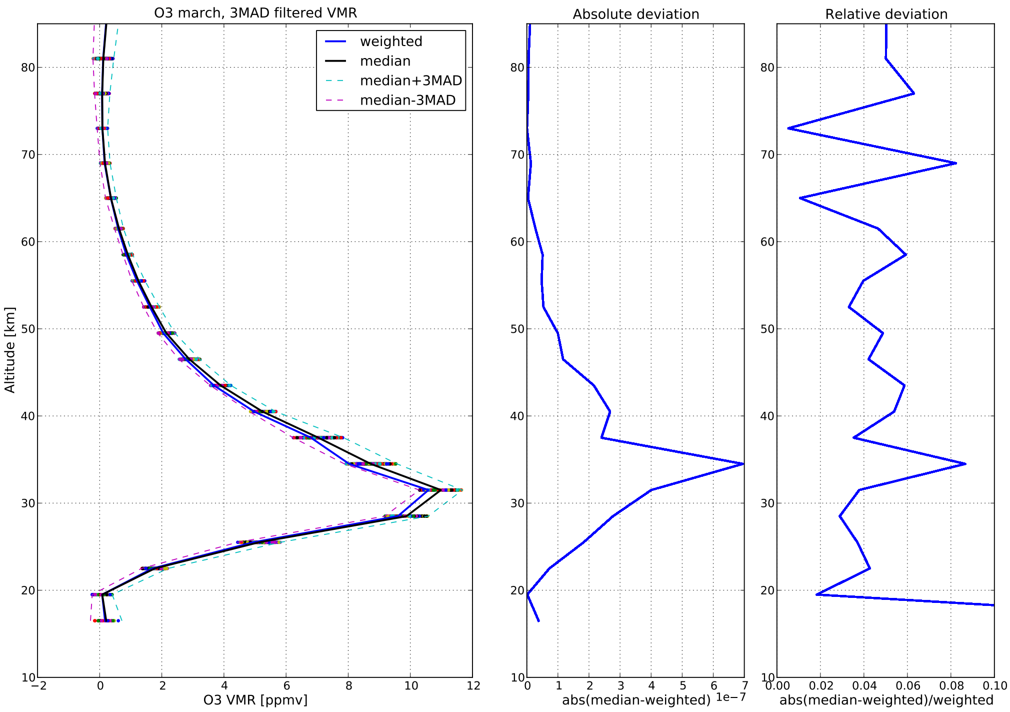


Figure 2: O3, Band A, March. Daytime data only, latitude bin = [2.5°, 7.5°]. Left plot shows quality filtered and 3MAD filtered VMR data (dots). The result of data weighting (blue curve) calculated according to Eq.2 and the median ± 3×MAD (see legend). Absolute and relative deviations of median to weighting, shown in the two right panels.

- Check for minimum and maximum measurement response. If measurement response is smaller/bigger than the threshold, the sample is flagged `false`. The maximum measurement response is set to 1.2 for all species. Minimum measurement response is generally 0.8, except for HOCl (0.7) and HNO3 (0.65).
- Check for retrieval algorithm convergence (χ^2). If χ^2 is bigger than the species dependent threshold, the whole profile gets set to `false`.
- Profile-wise check for minimum number of valid measurements per profile. The so far filtered profiles are now checked if they still contain a minimum number of valid measurements. Otherwise the profile is set to `false`.

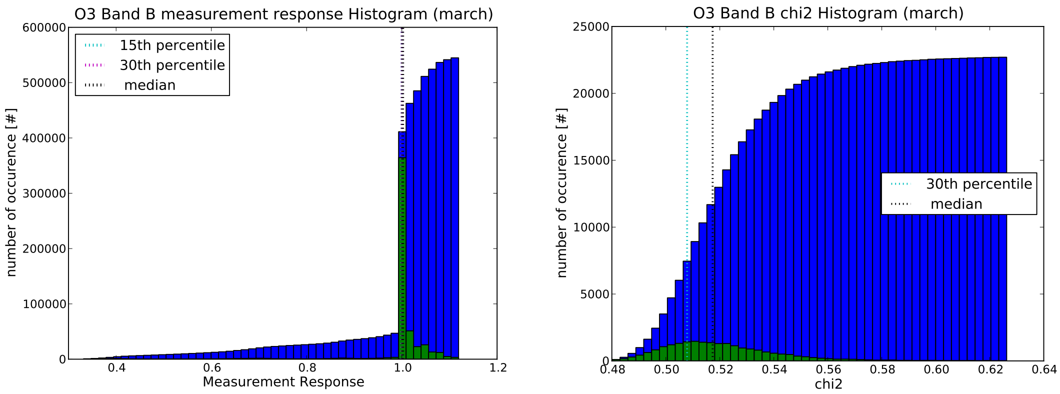


Figure 3: Histograms of measurement response (left) and χ^2 (right). Green shows normal histogram, while blue is the accumulated histogram. 15th, 30th and 50th percentile are given.

Apart from the `Quality_flg` matrix, an additional `Quality_info` dictionary entry is given. This lists the amount of filtered measurements and the ratio of filtered to total number of measurements. Depending on the species between 5% and 30% of the original data get filtered.

1.2.5 SMILES sampling issues

The sampling locations and times of SMILES, are not continuous and not evenly distributed in space. This causes problems and opportunities at the same time.

The possible dimensions of SMILES data can be summarized as: spatial (lat, lon, altitude), date, time of the day (LST or SZA), seasonal variations and the species. All of these need to be taken into account, when binning and averaging VMR values. Although, depending on the species some of these dimension are negligible.

Geo Location SMILES measurement locations are mostely between 38° South and 65° North (exceptions occur when the ISS is turned around the SMILES sensor is looking primarily towards the southern hemisphere [65° South and 38° North]). While in the equator region, the measurements are often quite sparse, the number of measurements accumulate at maximum and minimum latitude of ISS orbit. Other geo location issues, are the tangent point definition for a SMILES profile and the varying location with height.

SZA The SZA is strongly dependent on the latitude and the annual cycle. At middle and higher latitudes the SZA never reaches 0° or 180°. Further, for the winter hemisphere SZAs close to 0° are getting less, while those close to 180° are getting slightly more. Species with strong diurnal variations, often show steep VMR gradients around sunset and sunrise. Measurements taken from the gradient region, can strongly bias averages calculated for day and night time. Especially since the majority of SMILES measurements cluster around SZAs of ± 90°. The median statistics and the secondary binning according to LST help to prevent these biases

LST The LST is affected in the same way as the SZA, although to a lesser degree, since it is not latitude dependent. However LST is not a good coordinate for photochemical species, which concentrations are dependent on the fact if there is sun light or not.

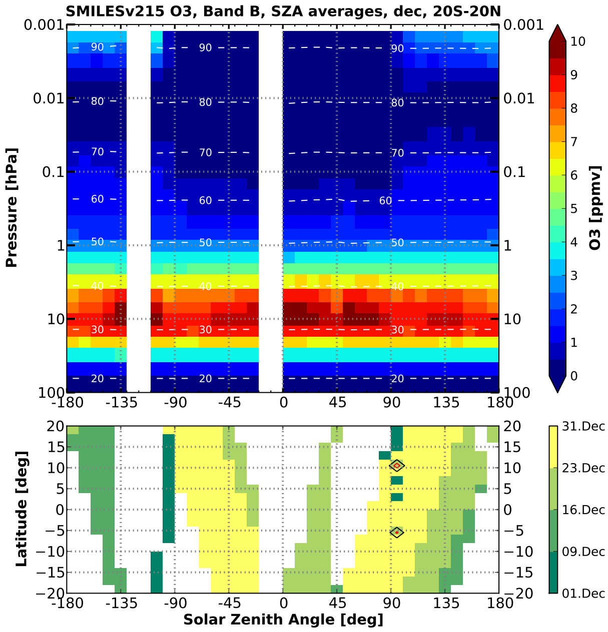
Example: The following example (Figure 4) briefly shows the problem of binning SMILES data. Figure 4a, shows band B ozone data of December 2009, binned to 10°SZA bins. The discontinuities in the stratospheric ozone layer are primarily caused by data originating from different latitudes and secondarily from different sampling times in one month. The subplot in Figure 4a depicts the latitudes from where the measurements in each SZA bin originate from. The peak O3 concentrations close to the two data gaps are caused by the data mostly coming from the southern most latitudes (20° South). In December the equatorial O3 maximum is shifted southwards, following the zenith of the sun.

On the other hand Figure 4b shows stratospheric O3 layer discontinuities caused by data originating from rather the beginning of the averaging period (1st Jan) or the end (28th Feb). Where ever the subplot indicates measurements taken in late February, the O3 concentrations are a little higher. This can be seen in SZA bins at -80°– -40°and 90°– 60°.

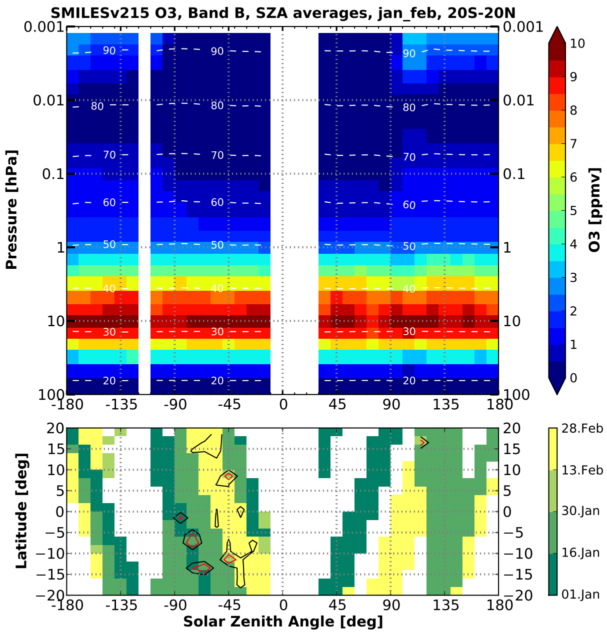
1.3 SMILES Derived Meteorological Product (DMP)

1.3.1 DMP

In the course of creating the SMILES climatology, we decided to generate equivalent latitude values for SMILES profiles. Therefore it was necessary to process extensive amounts of model data. Thus, apart from providing the equivalent latitude, we also wanted to add several other derived meteorological parameters. Three DMP subsets, one for each SMILES band, were created. Thereby the subsets of band A and band B were calculated along the L2 O3 profiles and band C was calculated along the L2 ClO profiles. These species were chosen, because they are the ones having the most retrieved L2 profiles in each band. In order to identify the DMP profile which corresponds to a



(a) LST averages



(b) SZA averages

Figure 4: The two figures show a prominent example of SMILES data binning problems. This causes features, that solely originate from data binning. See text for description.

L2 profiles, the `Time` variable can be used. This `Time` variable expresses the real time in seconds, since 0001-01-01 00:00:00. But also `TimeUTC` could be used for this purpose.

In the future it would be best and easiest to use the unique SMILES level 1b ID to avoid any possible ambiguities.

When adding the DMP information to any other L2 species, than O3 or ClO, it is very likely, that the DMP data set will contain profiles, which were not retrieved for this species. These exceptions need to be treated carefully, to make sure that the correct DMP profile gets associated to the L2 profile. Unfortunately it can also happen, that there is one or two single profiles in the L2, which can not be found in DMP data set. This also has to be handled.

Model data: GEOS5 data of the 3D dynamics fields, time-averaged on lcv coordinates (d520_fp.tavg3d_dyn_v) were used for deriving the meteorological product. The data can be found on the SMILES servers in the directory `/mnt/raid1/smilesdata/geos5/hdf`.

SMILES DMP The following atmospheric parameters were derived for the SMILES DMPs:

- Vertical coordinate parameters:
  - Altitude [m]
  - Potential temperature, [K]
  - Pressure [hPa]
  - Geopotential height [m]
- DMP values:
  - Potential vorticity [$K * m^2 * kg^{-1} * s^{-1}$]
  - Scaled potential vorticity [s^{-1}]
  - Equivalent Latitude [°]
  - Zonal wind [m/s]
  - Meridional wind [m/s]
  - Vertical pressure wind [Pa/s]

DMP data access The SMILES DMPs are stored in HDF 5 files and can be found on the SMILES servers: `/mnt/raid1/smilesdata/work/daniel/SMILESdmp/`.
The naming convention is:
`SMILESv215_L2DMP_GEOS-d520_Band_'band'_'yyyy'-'mm'-'dd'.he5`

1.3.2 Equivalent Latitude

Equivalent latitude is a useful measure to ensure, that binned and averaged data originates from the same air mass. The most critical boundary of air masses in the upper atmosphere is the polar vortex. Especially trace gas species tend to have steep gradients across the polar vortex boundary. Therefore it is a reasonable choice to calculate equivalent latitude and bin the SMILES profiles according to it.

General Concept: Global potential vorticity (PV) values on isentropic surfaces are getting associated with an eql. latitude, ranging from -90° to 90°. Thereby the areas with the lowest PV value on the southern hemisphere are getting the -90° eql. latitude value. Vice versa, the areas with highest PV values are associated with +90°. One can think of mapping the latitudes on PV values. Figure 5 shows an example of GEOS-5 PV data and the mapped eql. latitude.

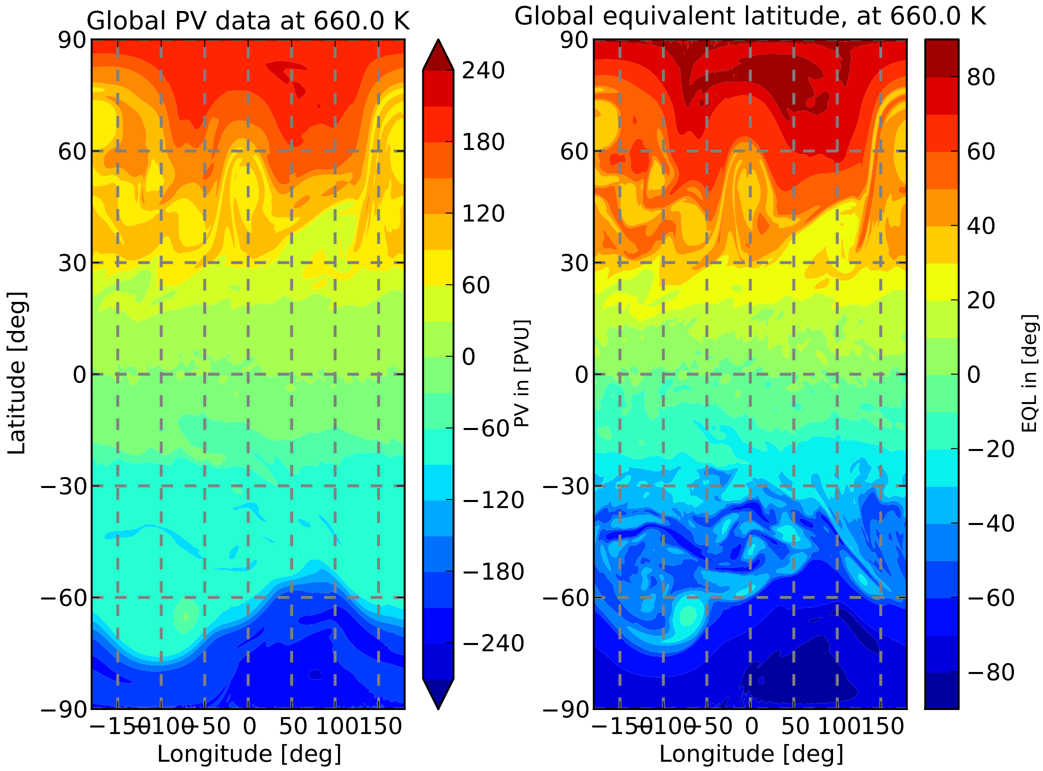


Figure 5: Potential vorticity from GEOS-5 data (left panel). Mapped eql. latitude (right panel). Isentropic level of 660K.

This eql. latitude data can now be added to the SMILES profiles by selecting the geo location and time from the profile and finding the corresponding eql. latitude for all altitudes.

2 Climatology Results

This section presents the results of the climatology. Example figures are shown as well as brief discussions about the processing. Further the directory path to the plots and data on the SMILES servers are given.

2.1 Naming conventions

Naming conventions for files or figures produced for the climatology are presented here.

Note: Not all attributes are needed or used for all files. It rather depends on the content of the file if additional specifications are needed. Therefore the attributes `daytime`, `latitude` and `altitude` are optional. They are only given if the information is necessary. That means, if a filename contains no `daytime` attribute, data of the entire day was used for climatology processing.

Table 2: Climatology attributes.

Attribute	Possible/Example values	Category
`data_version`	'2.1.5'	always
`band`	'A' | 'B' | 'C'	always
`species`	'O3' | 'H37Cl' | 'ClO' ...	always
`bin_type`	'lat' | 'eql' | 'sza' | 'lst' | 'days'	always
`month`	'oct_nov' | 'dec' | 'oct_apr' ...	always
`daytime`	'night' | 'day'	optional
`latitude`	'40S-20S' | '20N-50N' ...	optional

Table 3: Naming convention table. (ignore Hyphens at line break)

EPS/PDF/PNG figures:	`SMILESv'data_version'_'band'_'species'_'bin_type'-_'month'_'daytime'_'latitude'_'altitude'.eps/pdf/png`
HDF5 climatology data:	`res_SMILESv'data_version'_'band'_'species'_'bin_type'-_'month'_'daytime'_'latitude'_'altitude'.h5`
NetCDF minimum climatology data:	`res_SMILESv'data_version'_'band'_'species'_'bin_type'-_'month'_'daytime'_'latitude'_'altitude'.nc`

2.2 HDF5 file content:

The definitions of the fields given in the list below can also be found in the 'Definition' attribute of the HDF5 variables. HDF5 groups:

- Climatology_grid *[horizontal and vertical grid]*

  - `lat/eql/sza/lst+bins`: Boundaries between primary averaging bins.
  - `lat/lst+bins`: Boundaries between secondary averaging bins.
  - `levels`: Centers of vertical levels.

- 3D_statistics *[3 dimensional statistic values]*
  - `data_3d`: median VMR values on 3d climatology grid (primary bin × secondary bins × vertical levels), corrected by MAD outlier filtering.
  - `numel_3d`: Number of profiles per bin on 3d climatology grid.
  - `measresp_3d`: Median measurement response on 3d climatology grid.
  - `error_3d`: Median L2 retrieval error on 3d climatology grid.
  - `madvmr_3d`: Median absolut deviation of VMR values on 3d climatology grid.
  - `alt_3d`: Median altitude on 3d climatology grid.
  - `p_3d`: Median pressure on 3d climatology grid.
  - `T_3d`: Median model temperature on 3d climatology grid.

- 2D_statistics *[2 dimensional statistic values]*
  - `median_data`: Statistic median of VMR per primary averaging, corrected by MAD outlier filtering.
  - `mad_data`: Median absolut deviation of VMR per primary averaging bin, corrected by MAD outlier filtering.
  - `numel`: Total number of elements of measurements per primary averaging bin.
  - `median_measresp`: Statistic median of measurement response per primary averaging bin.
  - `median_error`: Statistic median of L2 retrieval error per primary averaging bin.
  - `median_alt`: Median altitude per primary averaging bin.
  - `median_p`: Median pressure per primary averaging bin.
  - `median_T`: Median model temperature per primary averaging bin.

- Auxiliaries *[some auxilliary data]*
  - `median_utc`: Median of UTC times per horizontal averaging box (primary × secondary dimension).
  - `mad_utc`: MAD of UTC times per horizontal averaging box (primary × secondary dimension).
  - `horizontal_numel`: Number of elements of measurements per horizontal averaging box (primary × secondary dimension).
  - **Optional values:** min, max and median of latitude, local solar time, solar zenith angle per horizontal averaging box (primary × secondary dimension).

- Info
  - `Band'`: SMILES band.
  - `Species`: Chemical species name.
  - `Version1b`: Version number of L1 data.
  - `Versionl2`: Version number of L2 data.
  - `L2coremetadata.0`: Some meta data of L2.
  - `L2StructMetadata.0`: Some structure meta data of L2 data.
  - `Primary_bin_type`: One of the following, Latitude, Equivalent Latitude, LocalSolarTime, SolarZenithAngle.
  - `Secondary_bin_type`: One of the following, Latitude, LocalSolarTime.
  - `Vertical_level_type`: Pressure.

2.3 NetCDF file content:

The NetCDF files contain the minimum data of the monthly climatologies. I.e. no extra information or auxiliary data is provided.

- `VMR`: Volume mixing ratio of the species, the variable name corresponds to the species chemical name.
- `mad`: Median absolute deviation, as measure of variability.
- `lat/eql/sza/lst+bins`: The X-axis, here the values define the center of the bin. (This change was necessary due to NetCDF file type specifications.)
- `plvl`: The vertical dimension (Y-axis). Also defining the center of the vertical bin.
- `starttime`: The first day of measurement in this climatology set.
- `endtime`: The last day of measurement in this climatology set.

2.4 Latitude Climatologies

Data access: The climatology data and figures are stored on the SMILES server at `raid1`.

- EPS and PDF figures:
  `smiles-p1:/mnt/raid1/smilesdata/work/daniel/clim/latclim/'species'/`
- HDF and NetCDF files:
  `smiles-p1:/mnt/raid1/smilesdata/work/daniel/clim/latclim/'species'/result/`

Example: As an example see Figure 6.

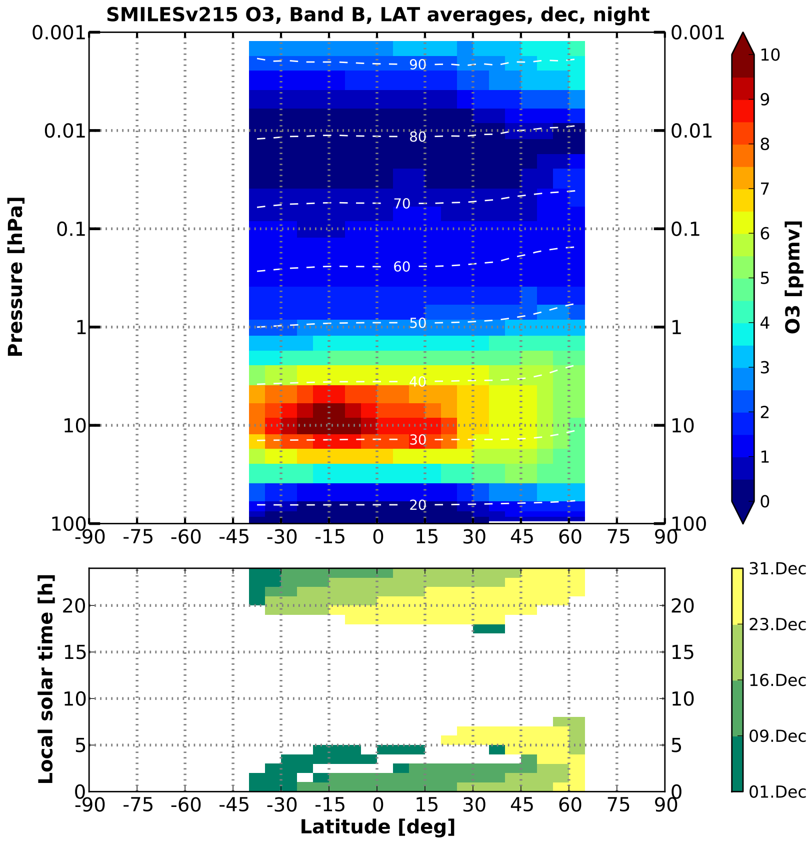


Figure 6: Zonal averaged O3 VMR, night time measurements only, for the month december.

Overview charts : Overview charts are uploaded to the Redmine -> Issues -> Climatology section.

- `https://smiles-p6.nict.go.jp/redmine/attachments/download/3668/lat_eql_climatologies.tar.gz`

The zonal climatologies take SMILES profiles from a single month and is binning the profiles into bins of 5° latitude. The bins are defined between the latitudes, meaning one bin goes from 10°to 15°. Horizontally the profiles get binned to one hour LST bins. Again the bins are defined between two full hours. Vertically the data is interpolated onto the modified SPARC CCMval pressure levels grid, as described in Section 1.2.1. Since the latitude bins start at -90°and go to +90°, many entries outside the SMILES latitude range are NaNs. Depending on the species, diurnal variations need to be taken into account or not. Data of diurnally varying species is split into daytime and nighttime data, while species unaffected by photochemical reactions are averaged over the full day.

2.5 Equivalent Latitude Climatologies

Data access: The climatology data and figures are stored on the SMILES server at `raid1`.

- EPS and PDF figures:
  `smiles-p1:/mnt/raid1/smilesdata/work/daniel/clim/eqlclim/'species'/`
- HDF and NetCDF files:
  `smiles-p1:/mnt/raid1/smilesdata/work/daniel/clim/eqlclim/'species'/result/`

Example: As an example see Figure 7.

Overview charts : Overview charts are uploaded to the Redmine -> Issues -> Climatology section.

- `https://smiles-p6.nict.go.jp/redmine/attachments/download/3668/lat_eql_climatologies.tar.gz`

The equivalent latitude climatologies take SMILES measurements from a single month and is binning the profiles into bins of 5° equivalent latitude. The bins are defined between the latitudes, meaning one bin goes from 10°to 15°. Note, that equivalent latitudes range from -90°to +90°. Horizontally the measurements get binned to one hour LST bins. Again the bins are defined between two full hours.

In this case the binning of the data does not happen profile wise, since equivalent latitude does vary with the vertical dimension of each profile. Hence each measurement in the vertical direction of each profile needs to be binned individually. Vertically the data is interpolated onto the modified SPARC CCMval pressure levels grid, as described in Section 1.2.1. Depending on the species, diurnal variations need to be taken into

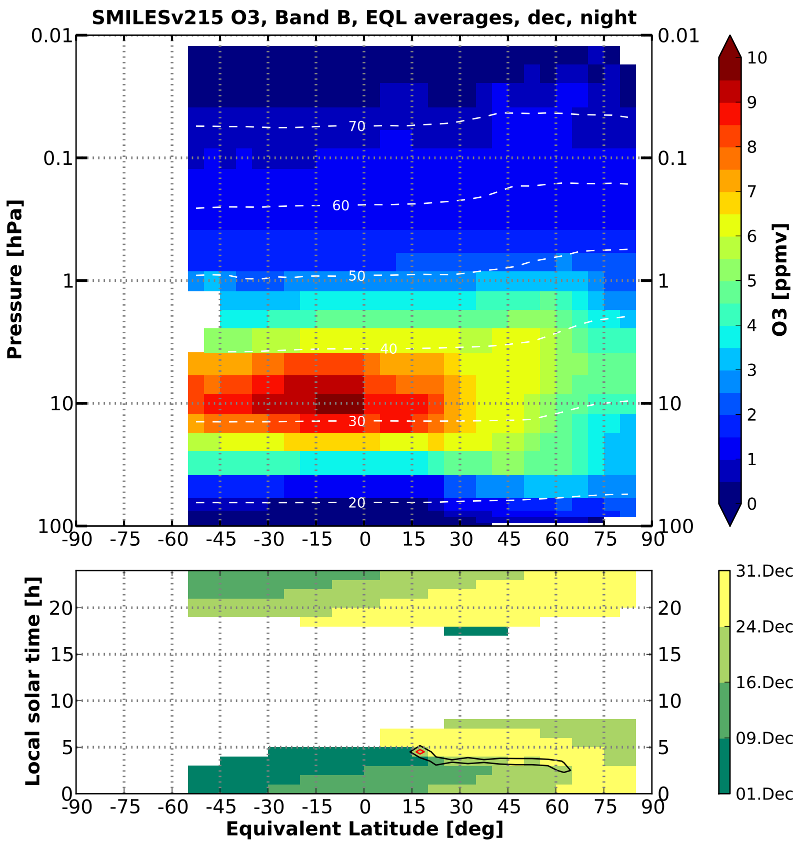


Figure 7: EQL averaged O3 VMR, night time measurements only, for the month december.

account or not. Data of diurnally varying species is split into daytime and nighttime data, while species unaffected by photochemical reactions are averaged over the full day.

2.6 Solar Zenith Angle Climatologies

Data access: The climatology data and figures are stored on the SMILES server at `raid1`.

- EPS, PDF and PNG figures:
  `smiles-p1:/mnt/raid1/smilesdata/work/daniel/clim/szaclim/'species'/`
- HDF and NetCDF files:
  `smiles-p1:/mnt/raid1/smilesdata/work/daniel/clim/SZAclim/'species'/result/`

Example: As an example see Figure 4.

Overview charts : Overview charts are uploaded to the Redmine -> Issues -> Climatology section.

- `https://smiles-p6.nict.go.jp/redmine/attachments/download/3667/sza_climatologies.tar.gz`

The solar zenith angle climatologies take SMILES profiles from 2 months and is binning the profiles according to their SZA into bins of 10° size. The bins are defined between the SZAs, meaning one bin goes from 10°to 20°. Horizontally the profiles get binned to 2°latitude bins. Again the bins are defined between two full latitude degrees. Vertically the data is interpolated onto the modified SPAR CCMval pressure levels grid, as described in Section 1.2.1. Different latitude ranges have been chosen to limit the profile selection to specific latitudinal regions. These are southern mid-latitudes (40°S–20°N), equator (20°S–20°N), northern mid-latitudes (20°N–50°N) and northern sub-polar latitudes (50°N–65°N). Also consult Section 1.2.5 for SMILES sampling issues.

2.7 Local Solar Time Climatologies

Data access: The climatology data and figures are stored on the SMILES server at `raid1`.

- EPS, PDF and PNG figures:
  `smiles-p1:/mnt/raid1/smilesdata/work/daniel/clim/lstclim/'species'/`
- HDF and NetCDF files:
  `smiles-p1:/mnt/raid1/smilesdata/work/daniel/clim/lstclim/'species'/result/`

Overview charts : Overview charts are uploaded to the Redmine -> Issues -> Climatology section.

- `https://smiles-p6.nict.go.jp/redmine/attachments/download/3666/lst_climatologies.tar.gz`

The Local solar time climatologies take SMILES profiles from 2 months and is binning the profiles according to their LST into bins of 1 hour. The bins are defined between the hours. Horizontally the profiles get binned to 2°latitude bins. Again the bins are defined between two full latitude degrees. Vertically the data is interpolated onto the modified SPAR CCMval pressure levels grid, as described in Section 1.2.1. Different latitude ranges have been chosen to limit the profile selection to specific latitudinal regions. These are southern mid-latitudes (40°S–20°N), equator (20°S–20°N), northern mid-latitudes (20°N–50°N) and northern sub-polar latitudes (50°N–65°N). Also consult Section 1.2.5 for SMILES sampling issues.

2.8 Daily averaged Climatologies

The time series climatologies are not an official product of the climatology, but were created together with it anyway. Some details in the processing, were handled differently, so the above mentioned specifications do generally not apply.

Data access: The climatology data and figures are stored on the SMILES server at `raid1`.

- EPS, PDF and PNG figures:
  `smiles-p1:/mnt/raid1/smilesdata/work/daniel/clim/dayclim/'species'/`
- HDF files:
  `smiles-p1:/mnt/raid1/smilesdata/work/daniel/clim/dayclim/'species'/result/`

Example: As an example see Figure 8.

Daily average climatologies show the concentration variability versus time. These products might give an insight into seasonal variations during northern hemisphere winter. For vertical interpolation and latitudinal limits, the same applies as stated in Section 2.6.

2.9 Climatology code

Code access: The climatology processing code is stored on the SMILES server at `raid1`.

- `smiles-p1:/mnt/raid1/smilesdata/work/daniel/clim/tools/`

The climatology processing code is written in Python v2.7. The code version on the server, had to be adapted for Python v2.6. Unfortunately some deprecation warnings remain.

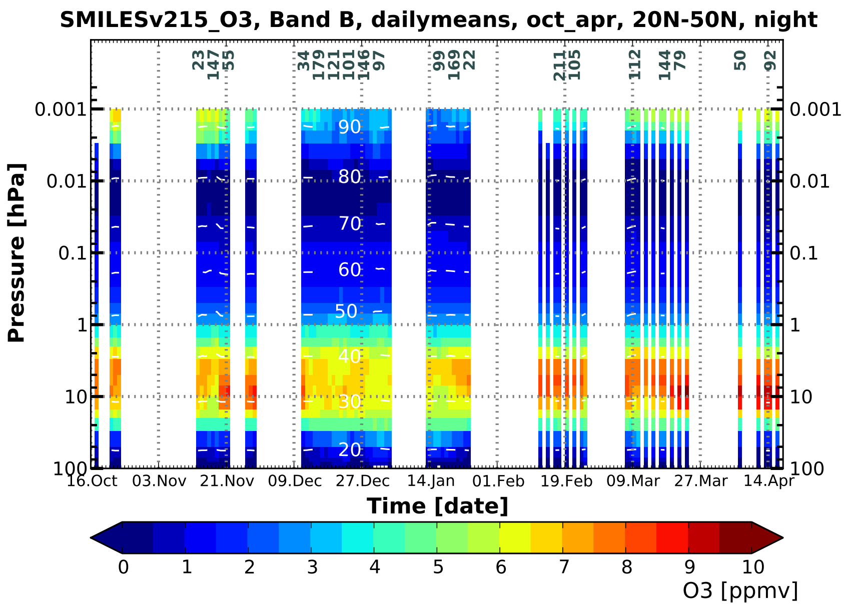


Figure 8: Daily averaged O3 VMR from 20°N–50°N, night time measurements only, for the month October to April. **Note**, due to limited space, not all number of measurements per day can be shown, only every 4th–5th number is shown.

Errors/Warnings: The execution of `calc_climavg.py:calc_climavg` creates several warnings. These warnings are caused by different statistical functions, like `nanmedian`, `nanmean`, or `nanstd`, when encountering a division by '0' or a negative argument to a square root. The values in these bins are set to '0' and NaN, respectively. The warnings are not specially treated so far, since they cause no disadvantage in the calculations.

Overview: The code and all functions contain a header and in-line documentation.

- `read_multiple_hdf.py`
  - `read_multiple_hdf(filename)`
    "'Read multiple daily L2 hdf files. Returns a list including dictionaries with profiles per day. "'
  - `merge_l2_list(l2list)`
    "'Merges the l2 data list of several days to a single dictionary."'
  - `check_l2_data(l2)`
    "'Performs species dependent sanity checks on L2 data dictionary."'
  - `merge_dmp2l2(dmplist, l2)`
    "' Copies the derived meteorological product data to the l2 dictionary. The profiles in the DMPs and the L2 are checked for having the same Time stamp. "'

- `calc_daymean.py`
  - `calc_daymean(l2, **kwargs)`
    "'Main function to calculate daily averaged climatology values."'
- `calc_climavg.py`
  - `calc_climavg(l2, **kwargs)`
    "'Main function to calculate climatological average values. The full vertical data range is taken into account."'
- `funcs.py`
  Compendium of small helper functions.
  - `mad(a, axis=None)`
    "' "'Compute median absolute deviation over axis 0 ignoring nans. Only axis 0 is possible!"'
  - `Odintime2mpltime(tarr)` "'Convert odin time (days since 1950-01-01 00:00:00) to MATLAB date number."'
  - `lst2hours(lst)`
    "'Compute the decimal hour (13:30 -> 13.5) from local solar time."'
  - `mydate2num(tarr)`
    "'Convert datetime string/object array to MATLAB date number."'
  - `theta(p,T)`
    "' Calculate potential temperature theta for pressure (Pa) and temperature (K)"'
  - `USSA()`
    "'Calculate the US standard atmosphere (1976 version)"'
  - `pressure(ps,dp)`
    "' pressure(ps,dp) -> returns layer-mean pressure profile from surface [Pa]"'
- `daymean_plot.py`
  - `daymean_plot(result, **kwargs)`
    "'Creates a contourf plot for daily averaged climatology data. The full vertical data range is taken into account."'
  - `daymean_part(result, **kwargs)`
    "' Creates a contourf plot for daily avergaed climatology data. Only a specified vertical data range is taken into account. (e.g. mesosphere only, or stratosphere only)"'
- `szamean_plot.py`
  - `szamean_plot(result, **kwargs)`
    "' Creates a pcolor plot for climatology data binned/averaged to SZA. The full vertical data range is taken into account."'

  - `szamean_split(result, **kwargs)`
    "' Creates a pcolor plot for climatology data binned/averaged to SZA. MESOSPHERE AND STRATOSPHERE are split into two subplots."'

- `zonalmean_plot.py`
  - `zonalmean_plot(result, **kwargs)`
    "' Creates a contourf plot for climatology data binned/averaged to latitude. The full vertical data range is taken into account, although vertical limitations can be defined."'
  - `zonalmean_split(result, **kwargs)`
    "' Creates a pcolor plot for climatology data binned/averaged to latitude. MESOSPHERE AND STRATOSPHERE are split into two subplots."'

- `eqlmean_plot.py`
  - `eqlmean_plot(result, **kwargs)`
    "' Creates a contourf plot for climatology data binned/averaged to equivalent latitude. The full vertical data range is taken into account, although vertical limitations can be defined."'

- `lstmean_plot.py`
  - `lstmean_plot(result, **kwargs)`
    "' Creates a contourf plot for climatology data binned/averaged to LST. The full vertical data range is taken into account, although vertical limitations can be defined."'

- `write_clim.py`
  - `write_res(res, **kwargs)`
    "'Function to write the results of climatology average to HDF5 files."'

- `NC_converter.py`
  - `write_res(res, **kwargs)`
    ""'Function to write the essential results of climatology HDF files to netCDF files."'

- `'species'_sza_clim.py`
  These are python scripts for executing the climatology code for a specific species.

- `'species'_lat_clim.py`
  These are python scripts for executing the climatology code for a specific species.

- `'species'_eql_clim.py`
  These are python scripts for executing the climatology code for a specific species.

- `'species'_lst_clim.py`
  These are python scripts for executing the climatology code for a specific species.
- `'species'_day_clim.py` scripts only exist for long-lived species, like O3.
- `dmp2scan.py`
  Function for creating SMILES DMPs:
  - `emGEOS_converter(path, filename)`
    "'Extracting and reshaping GEOS5 data to prepare for DMP processing. Also calculating DMP values not available in GEOS5 data."'
  - `dmp2scan(l2)`
    "'Interpolate the GEOS5 data to the L2 profiles, band-wise."'
- `calc_eql_serv.py`
  - `calc_eql(pv)`
    "'calculating equivalent latitude from GEOS-5 potential vorticity (PV) on isentropic surfaces."'
  - `dmp_interp(dmp)`
    "'Vertical interpolation from native GEOS5 grid to potential temperature levels."'
- `write_dmp.py`
  - `write_dmp(dmp, **kwargs)`
    "'Function to write the DMPs to HDF5 files."'
- `DMP_'band'script.py`
  Script to execute DMP processing of the specified 'band'.

Acknowledgments

This work has been strongly supported by the entire NICT SMILES group. Thank you for viable discussions and comments to Yasuko Kasai, Hideo Sagawa, Tomohiro Sato, Kodai Suzuki. Also thanks for computational and technical assistance to Philippe Baron, Joe Urban, Joakim Moeller, and SEC company. And finally thanks to Ashley Jones for climatological key issue discussions and a lot of expertise.