CMIP Greenhouse Gas (GHG) Concentration Historical Dataset

Zebedee Nicholls, Florence Bockting, Mika Pflüger

Oct 29, 2025

CONTENTS

CHAPTER

ONE

DATA DESCRIPTION AND USER GUIDE

1.1 Overview

Here we provide a short description of the historical dataset and a guide for users. This is intended to provide a short introduction for users of the data: its construction, key features, metadata and relationship to CMIP6 forcing data. The full details of the dataset's construction and evaluation against other data sources will be provided in the full manuscript which is being prepared.

1.2 Dataset construction

The dataset is constructed following the methodology of Meinshausen *et al.* [1]. The methods are described in full in that paper and will be clarified and described again in the forthcoming manuscript describing this dataset's construction.

In brief, the dataset for each greenhouse gas is constructed via the following steps:

1. collect as many ground-based observations as possible
2. from ground-based networks such as the NOAA [2, 3] and AGAGE [4, 5, 6, 7] networks
    - these are only available over the last few decades at most (less for some greenhouse gases)
    - these are spatially sparse because sampling stations are discrete points and there are not an infinite number of stations (at most, usually around 30, often far fewer)
3. bin the ground-based observations in space and time, averaging over input stations and observations that fall in the same cell
4. interpolate the binned data in space, to derive a dataset with spatial coverage
5. use the interpolated, ground-based data to derive a statistical model for seasonal variation and latitudinal gradients specific to each greenhouse gas
    - the exact form of the statistical model varies by gas, but is generally driven by either concentrations of the gas itself, global-mean temperature or purely statistical regressions/extensions
6. use the models, plus ice core or other proxy records, to extend global-mean concentrations, seasonality and latitudinal gradients over the full time period of the dataset (i.e. back to year 1)
    - where ice cores or proxy records are not available, purely statistical extrapolations are used instead
    - the extension varies by gas, aiming to make use of as much information as is possible e.g. hemisphere specific ice core information and the latitudinal gradient over the period covered by ground-based observations
7. combine the extended global-mean, seasonality and latitudinal gradients to create a dataset that extends over the period year 1 to 2022 (the last year available for some observational networks at the time the data was compiled)

   - this dataset is on our binned grid, which we choose to be a grid comprised of latitudinal bins 15-degrees in size
   - it is not trivial to infer the global-means, seasonality and latitudinal gradient used to construct the dataset from the output dataset. For this reason, we include these components separately in the zenodo record [TODO better ref] that archives the output dataset, all its inputs and intermediate data prdoucts

8. calculate annual-, hemispheric- and global-means to produce our lower resolution data products
   - we can also produce higher spatial resolution data products, but have not done so at the moment to save processing and storage space given that there has been no demand for these products from modelling teams

The input datasets and associated references are documented in the `references*` attributes of each netCDF file. This documentation is limited, so cannot document how each input dataset is used (that is the role of the manuscript), but does provide machine-readable provenance information (which is used to support links between all the input data e.g. linking of the Zenodo archive underpinning this dataset).

1.3 Finding and accessing the data

1.3.1 ESGF

The **Earth System Grid Federation** (ESGF, REF-TODO) provides access to a range of climate data. The historical data of interest here, which is the data to be used for historical and piControl simulations within CMIP [TODO ref Dunne paper], can be found under the “source ID”, `CR-CMIP-1-0-0`. The concept of a “source ID” is a bit of a perculiar one to CMIP forcings data. In practice, it is simply a unique identifier for a collection of datasets (and it’s best not to read more than that into it).

It is possible to filter searches on ESGF via the user interface (see ESGF user guides[1]). Alternatively, searches can be encoded in URLs. However, a caveat with this approach is that URLs sometimes move, so we make no guarantee that this link will always be live. An example provides the following link:

> https://esgf-node.ornl.gov/search?project=input4MIPs&activeFacets=%7B%2ource_id%22%3A%22CR-CMIP-1-0-0%22%7D

To download the data, we recommend accessing it directly via the ESGF user interfaces via links like the one above. Alternately, there are tools dedicated to accessing ESGF data, with two prominent examples being **esgpull**[2] and **intake-esgf**[3]. Please refer to the tools’ docs for usage instructions.

1.3.2 Zenodo

While it aims to be, the ESGF is technically not a permanent archive and does not issue DOIs. In order to provide more reliable, citable access to the data, we also provide it on **Zenodo** (REF-TODO). The data, as well as all the source code and input data used to process it, can be found at https://doi.org/10.5281/zenodo.14892947.

[1] https://esgf.github.io/esgf-user-support/user_guide.html#data-search-and-download
[2] https://esgf.github.io/esgf-download
[3] https://intake-esgf.readthedocs.io

1.4 Data description

1.4.1 Format

The data is provided in **netCDF format** [TODO citation]. This self-describing format allows the data to be placed in the same file as metadata (in the so-called "file header"). To facilitate simpler use of the data, each dataset is split across multiple files. The advantage of this is that users do not need to load all years of data if they are only interested in data for a certain range, which can significantly improve data loading times. To get the complete dataset, the files can simply be concatenated in time.

1.4.2 Grids and frequencies provided

We provide five combinations of grids and time sampling (also referred to as frequency, although this is a bit of a misuse as the units of frequency are per time, which doesn't match the convention for these metadata values). The grid and frequency information for each file can be found in its netCDF header under the attributes `grid_label` (for grid) and `frequency` (for time sampling). The `grid_label` and `frequency` also appear in each file's name, which allows files to be filtered without needing to load them first.

The five combinations of grid and time sampling are:

1. global-, annual-mean (`grid_label="gm"`, `frequency="yr"`)
2. global-, monthly-mean (`grid_label="gm"`, `frequency="mon"`)
3. hemispheric-, annual-mean (`grid_label="gr1z"`, `frequency="yr"`)
4. hemispheric-, monthly-mean (`grid_label="gr1z"`, `frequency="mon"`)
5. 15-degree latitudinal, monthly-mean (`grid_label="gnz"`, `frequency="mon"`)

1.4.3 Species provided

We provide concentrations for 43 greenhouse gas concentrations and species, as well as three equivalent species. The species are:

- major greenhouse gases (3)
  - CH4, CO2, N2O
- ozone-depleting substances (17)
  - CFCs (5)
    - CFC-11, CFC-113, CFC-114, CFC-115, CFC-12
  - HCFCs (3)
    - HCFC-141b, HCFC-142b, HCFC-22
  - Halons (3)
    - Halon 1211, Halon 1301, Halon 2402
  - other ozone-depleting substances (6)
    - CCl4, CH2Cl2, CH3Br, CH3CCl3, CH3Cl, CHCl3
- ozone fluorinated compounds (23)
  - HFCs (11)

    * HFC-125, HFC-134a, HFC-143a, HFC-152a, HFC-227ea, HFC-23, HFC-236fa, HFC-245fa, HFC-32, HFC-365mfc, HFC-4310mee

- PFCs (9)

    * C2F6, C3F8, C4F10, C5F12, C6F14, C7F16, C8F18, CC4F8, CF4

- other (3)

    * NF3, SF6, SO2F2

Equivalent species

For most models, you will not use all 43 species. As a result, we provide equivalent species too. There are two options if you don't want to use all 43 species.

Option 1

Use CO2, CH4, N2O and CFC-12 directly. Use CFC-11 equivalent to capture the radiative effect of all other species.

Option 2

Use CO2, CH4 and N2O directly. Use CFC-12 equivalent to capture the radiative effect of all ozone depleting substances (ODSs) and HFC-134a equivalent to capture the radiative effect of all other fluorinated gases.

1.4.4 Uncertainty

At present, we provide no analysis of the uncertainty associated with these datasets. In radiative forcing terms, the uncertainty in these concentrations is very likely to be small compared to other uncertainties in the climate system, but this statement is not based on any robust analysis (rather it is based on expert judgement). It is also worth noting that the uncertainty increases as we go further back in time, particularly as we shift from using surface flasks to relying on ice cores instead.

1.4.5 Differences compared to CMIP6

At present, the changes from CMIP6 are minor, with the maximum difference in effective radiative forcing terms being 0.05 W / m2 (and generally much smaller than this, particularly after 1850). For more details, see the plots in the user guide below and the forthcoming manuscript.

1.5 User guide

Having downloaded the data, using it is quite straightforward.

1.5.1 Annual-, global-mean data

We start with the annual-, global-mean data. Like all our datasets, this is composed of three files, each covering a different time period:

1. year 1 to year 999
2. year 1000 to year 1749
3. year 1750 to year 2022

For yearly data, the time labels in the filename are years (for months, the month is included e.g. you will see `000101-09912` rather than `0001-0999` in the filename, the files also have different values for the `frequency` attribute). Global-mean data is identified by the 'grid label' `gm`, which appears in the filename. Below we show the filenames for the CO2 output.

```
- co2_input4MIPs_GHGConcentrations_CMIP_CR-CMIP-1-0-0_gm_1000-1749.nc
- co2_input4MIPs_GHGConcentrations_CMIP_CR-CMIP-1-0-0_gm_1750-2022.nc
- co2_input4MIPs_GHGConcentrations_CMIP_CR-CMIP-1-0-0_gm_0001-0999.nc
```

Output for other gases are named identically, with `co2` being replaced by the other gas name. For example, for methane the filenames are:

```
- ch4_input4MIPs_GHGConcentrations_CMIP_CR-CMIP-1-0-0_gm_1000-1749.nc
- ch4_input4MIPs_GHGConcentrations_CMIP_CR-CMIP-1-0-0_gm_1750-2022.nc
- ch4_input4MIPs_GHGConcentrations_CMIP_CR-CMIP-1-0-0_gm_0001-0999.nc
```

As described above, the data is netCDF files. This means that metadata can be trivially inspected using a tool like `ncdump`. As you can see, there is a lot of metadata included in these files. In general, you should not need to parse this metadata directly. However, if you have specific questions, please feel free to contact the emails given in the `contact` attribute.

Using a tool like xarray, loading and working with the data is trivial.

```python
import xarray as xr

time_coder = xr.coders.CFDatetimeCoder(use_cftime=True)
ds_co2_yearly_global = xr.open_mfdataset(co2_yearly_global_fps, decode_times=time_
↪coder)
```

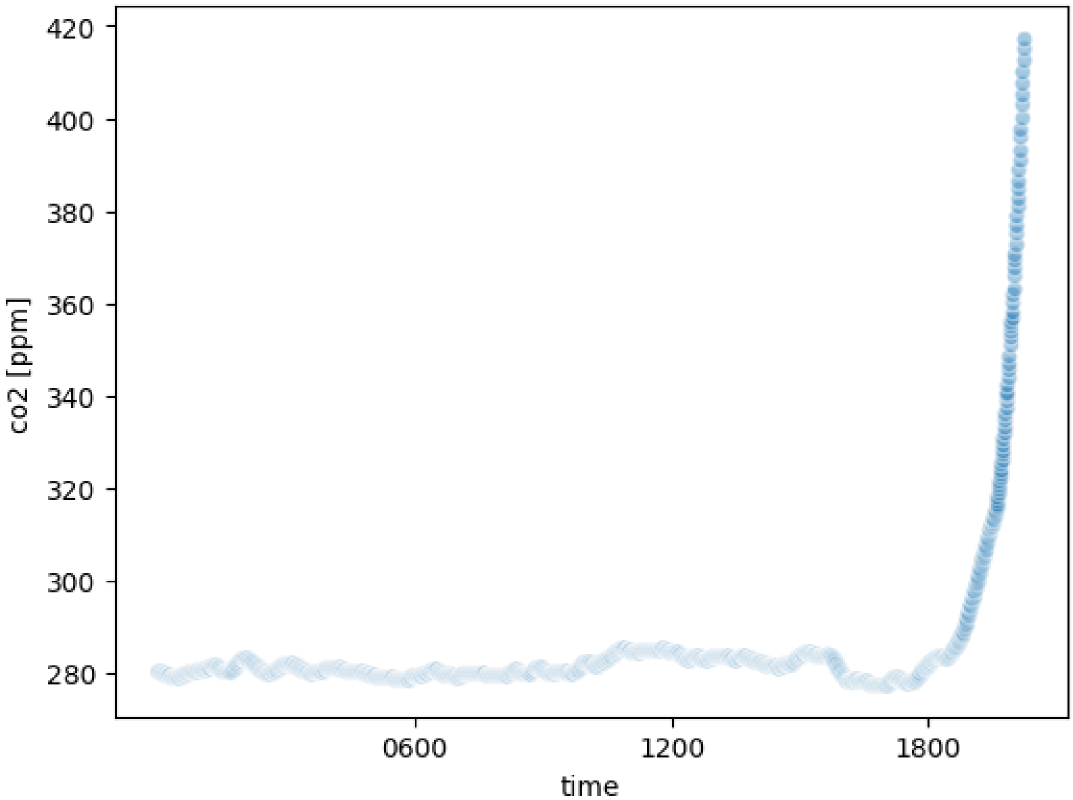
```python
ds_co2_yearly_global
```

```
<xarray.Dataset> Size: 57kB
Dimensions:     (time: 2022, bnds: 2)
Coordinates:
  * time        (time) object 16kB 0001-07-02 12:00:00 ... 2022-07-02 12:00:00
Dimensions without coordinates: bnds
Data variables:
    co2         (time) float32 8kB 280.5 280.4 280.4 280.3 ... 412.9 415.1 417.3
    time_bnds   (time, bnds) object 32kB 0001-01-01 00:00:00 ... 2023-01-01 00...
Attributes: (12/29)
    Conventions:             CF-1.7
    activity_id:             input4MIPs
    comment:                 Data compiled by Climate Resource, based on scie...
    contact:                 zebedee.nicholls@climate-resource.com;malte.mein...
    creation_date:           2025-02-28T18:23:33Z
    dataset_category:        GHGConcentrations
    ...                      ...
    references_urls:         https://scrippsco2.ucsd.edu/data/atmospheric_co2...
```

(continues on next page)

(continued from previous page)

```
    source_id:                   CR-CMIP-1-0-0
    source_version:              1.0.0
    target_mip:                  CMIP
    tracking_id:                 hdl:21.14100/d3e2e715-f5a7-4d0e-973c-1fb77263f9b7
    variable_id:                 co2
```



1.5.2 Space- and time-average nature of the data

All of our data represents the mean over each cell. This is indicated by the `cell_methods` attribute of all of our output variables.

```
ds_co2_yearly_global["co2"].attrs["cell_methods"]
```

```
'area: time: mean'
```

This mean is both in space and time. The time bounds covered by each step are specified by the `time_bnds` variable (when there is spatial information, equivalent `lat_bnds` and `lon_bnds` information is also included). This variable specifies the start (inclusive) and end (exclusive) of the time period covered by each data point.

```
ds_co2_yearly_global["time_bnds"]
```

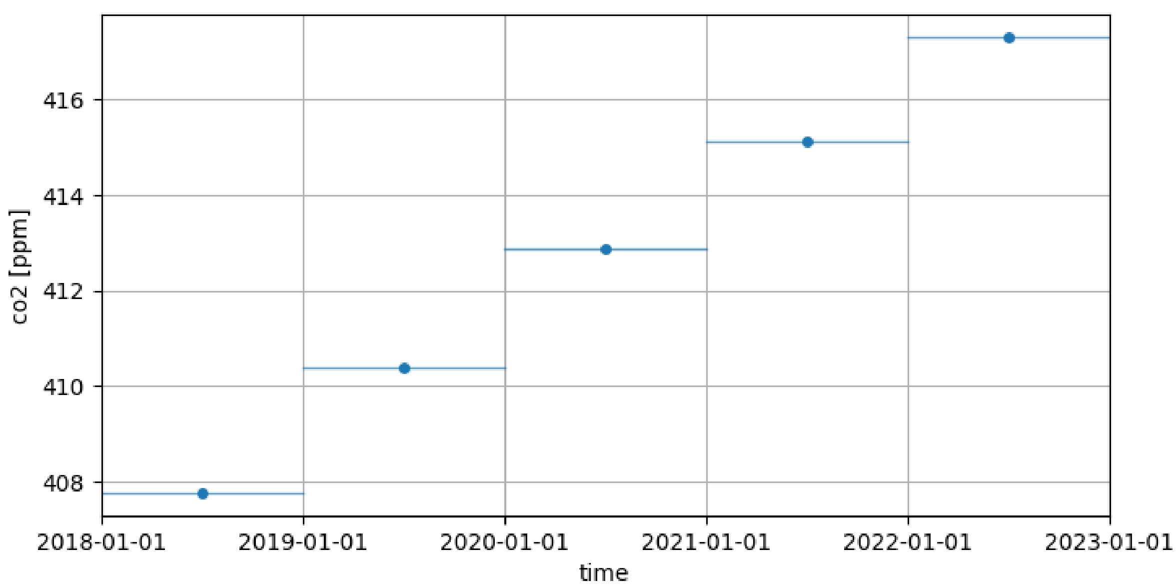
```
<xarray.DataArray 'time_bnds' (time: 2022, bnds: 2)> Size: 32kB
array([[cftime.DatetimeProlepticGregorian(1, 1, 1, 0, 0, 0, 0, has_year_zero=True),
        cftime.DatetimeProlepticGregorian(2, 1, 1, 0, 0, 0, 0, has_year_
↪zero=True)],
```

(continues on next page)

(continued from previous page)

```
       [cftime.DatetimeProlepticGregorian(2, 1, 1, 0, 0, 0, 0, has_year_zero=True),
        cftime.DatetimeProlepticGregorian(3, 1, 1, 0, 0, 0, 0, has_year_
↪zero=True)],
       [cftime.DatetimeProlepticGregorian(3, 1, 1, 0, 0, 0, 0, has_year_zero=True),
        cftime.DatetimeProlepticGregorian(4, 1, 1, 0, 0, 0, 0, has_year_
↪zero=True)],
       ...,
       [cftime.DatetimeProlepticGregorian(2020, 1, 1, 0, 0, 0, 0, has_year_
↪zero=True),
        cftime.DatetimeProlepticGregorian(2021, 1, 1, 0, 0, 0, 0, has_year_
↪zero=True)],
       [cftime.DatetimeProlepticGregorian(2021, 1, 1, 0, 0, 0, 0, has_year_
↪zero=True),
        cftime.DatetimeProlepticGregorian(2022, 1, 1, 0, 0, 0, 0, has_year_
↪zero=True)],
       [cftime.DatetimeProlepticGregorian(2022, 1, 1, 0, 0, 0, 0, has_year_
↪zero=True),
        cftime.DatetimeProlepticGregorian(2023, 1, 1, 0, 0, 0, 0, has_year_
↪zero=True)]],
      shape=(2022, 2), dtype=object)
Coordinates:
  * time     (time) object 16kB 0001-07-02 12:00:00 ... 2022-07-02 12:00:00
Dimensions without coordinates: bnds
```

As a result of the time average that the data represents, it is inappropriate to plot this data using a line plot (the mean of the lines joining the points is not the same as the data given in the files). Instead, the data should be plotted (and used) as a scatter or a step plot, as shown below. (The same logic applies to any spatial plots which could be created from our datasets that include spatial dimensions).



1.5.3 Monthly-, global-mean data

If you want to have information at a finer level of temporal detail, we also provide monthly files. Like the global datasets, these come in three files.

For monthly data, the time labels in the filename are months. Below we show the filenames for the CO2 output.

```
- co2_input4MIPs_GHGConcentrations_CMIP_CR-CMIP-1-0-0_gm_000101-099912.nc
- co2_input4MIPs_GHGConcentrations_CMIP_CR-CMIP-1-0-0_gm_100001-174912.nc
- co2_input4MIPs_GHGConcentrations_CMIP_CR-CMIP-1-0-0_gm_175001-202212.nc
```

Again, the data can be trivially loaded with xarray.

```
ds_co2_monthly_global = xr.open_mfdataset(
    co2_monthly_global_fps, decode_times=time_coder
)
```

```
ds_co2_monthly_global
```

```
<xarray.Dataset> Size: 679kB
Dimensions:     (time: 24264, bnds: 2)
Coordinates:
  * time        (time) object 194kB 0001-01-15 00:00:00 ... 2022-12-15 00:00:00
Dimensions without coordinates: bnds
Data variables:
    co2         (time) float32 97kB 281.1 281.6 281.9 282.1 ... 416.7 418.2 419.3
    time_bnds   (time, bnds) object 388kB 0001-01-01 00:00:00 ... 2023-01-01 0...
Attributes: (12/29)
    Conventions:             CF-1.7
    activity_id:             input4MIPs
    comment:                 Data compiled by Climate Resource, based on scie...
    contact:                 zebedee.nicholls@climate-resource.com;malte.mein...
    creation_date:           2025-02-28T18:23:29Z
    dataset_category:        GHGConcentrations
    ...                      ...
    references_urls:         https://scrippsco2.ucsd.edu/data/atmospheric_co2...
    source_id:               CR-CMIP-1-0-0
    source_version:          1.0.0
    target_mip:              CMIP
    tracking_id:             hdl:21.14100/1b88c4f6-477d-46b3-8fcb-0f21170ea735
    variable_id:             co2
```

For this data, the time bounds show that each point is the average a month, not a year.

```
ds_co2_monthly_global["time_bnds"]
```

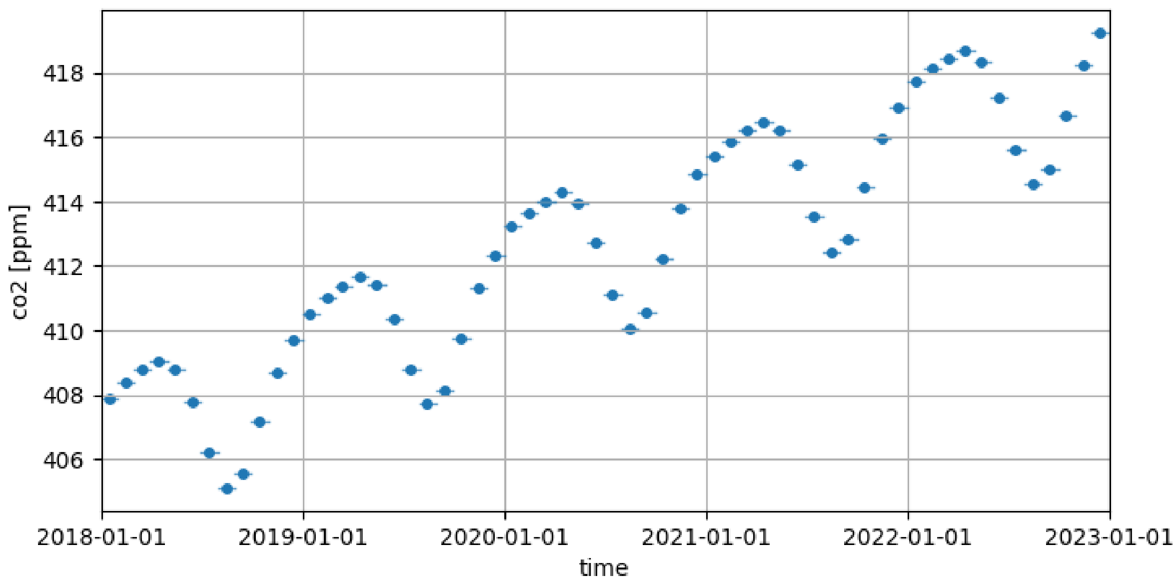
```
<xarray.DataArray 'time_bnds' (time: 24264, bnds: 2)> Size: 388kB
array([[cftime.DatetimeProlepticGregorian(1, 1, 1, 0, 0, 0, 0, has_year_zero=True),
        cftime.DatetimeProlepticGregorian(1, 2, 1, 0, 0, 0, 0, has_year_
 ↪zero=True)],
       [cftime.DatetimeProlepticGregorian(1, 2, 1, 0, 0, 0, 0, has_year_zero=True),
        cftime.DatetimeProlepticGregorian(1, 3, 1, 0, 0, 0, 0, has_year_
 ↪zero=True)],
       [cftime.DatetimeProlepticGregorian(1, 3, 1, 0, 0, 0, 0, has_year_zero=True),
        cftime.DatetimeProlepticGregorian(1, 4, 1, 0, 0, 0, 0, has_year_
 ↪zero=True)],
       ...,
```

(continues on next page)

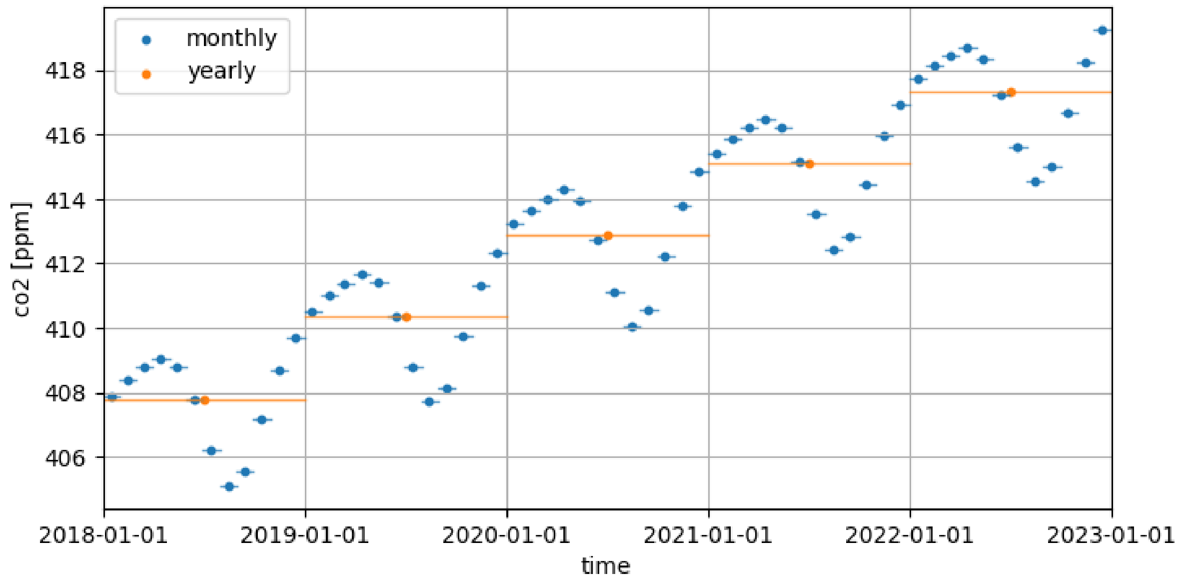
(continued from previous page)

```
       [cftime.DatetimeProlepticGregorian(2022, 10, 1, 0, 0, 0, 0, has_year_
↪zero=True),
        cftime.DatetimeProlepticGregorian(2022, 11, 1, 0, 0, 0, 0, has_year_
↪zero=True)],
       [cftime.DatetimeProlepticGregorian(2022, 11, 1, 0, 0, 0, 0, has_year_
↪zero=True),
        cftime.DatetimeProlepticGregorian(2022, 12, 1, 0, 0, 0, 0, has_year_
↪zero=True)],
       [cftime.DatetimeProlepticGregorian(2022, 12, 1, 0, 0, 0, 0, has_year_
↪zero=True),
        cftime.DatetimeProlepticGregorian(2023, 1, 1, 0, 0, 0, 0, has_year_
↪zero=True)]],
      shape=(24264, 2), dtype=object)
Coordinates:
  * time     (time) object 194kB 0001-01-15 00:00:00 ... 2022-12-15 00:00:00
Dimensions without coordinates: bnds
```

As above, as a result of the time average that the data represents, it is inappropriate to plot this data using a line plot. Scatter or step plots should be used instead.



The monthly data includes seasonality. Plotting the monthly and yearly data on the same axes makes particularly clear why a line plot is inappropriate.



At present, we do not provide data at a higher temporal resolution than monthly. In theory, such a dataset is possible to compile, however this requires careful consideration of daily and potentially sub-daily trends (e.g. the diurnal cycle).

1.5.4 Monthly-, latitudinally-resolved data

We also provide data with spatial, specifically latituindal, resolution. This data comes on a 15-degree latituindal grid (see below for details of the grid and latitudinal bounds). These files are identified by the grid label `gnz`. We only provide these files with monthly resolution.

For completeness, we note that we also provide hemispheric means. These are not shown here, but are identified by the grid label `gr1z`.

Below we show the filenames for the latitudinally-resolved data for CO2

```
- co2_input4MIPs_GHGConcentrations_CMIP_CR-CMIP-1-0-0_gnz_175001-202212.nc
- co2_input4MIPs_GHGConcentrations_CMIP_CR-CMIP-1-0-0_gnz_000101-099912.nc
- co2_input4MIPs_GHGConcentrations_CMIP_CR-CMIP-1-0-0_gnz_100001-174912.nc
```

Again, the data can be trivially loaded with xarray.

```
ds_co2_monthly_lat = xr.open_mfdataset(
    co2_monthly_lat_fps, decode_times=time_coder, data_vars=None, compat="no_conflicts
↪"
)
```

```
ds_co2_monthly_lat
```

```
<xarray.Dataset> Size: 2MB
Dimensions:     (time: 24264, lat: 12, bnds: 2)
Coordinates:
  * time        (time) object 194kB 0001-01-15 00:00:00 ... 2022-12-15 00:00:00
  * lat         (lat) float64 96B -82.5 -67.5 -52.5 -37.5 ... 37.5 52.5 67.5 82.5
Dimensions without coordinates: bnds
Data variables:
    co2         (time, lat) float32 1MB 280.9 280.8 280.8 ... 425.3 424.6 424.1
```

(continues on next page)

(continued from previous page)

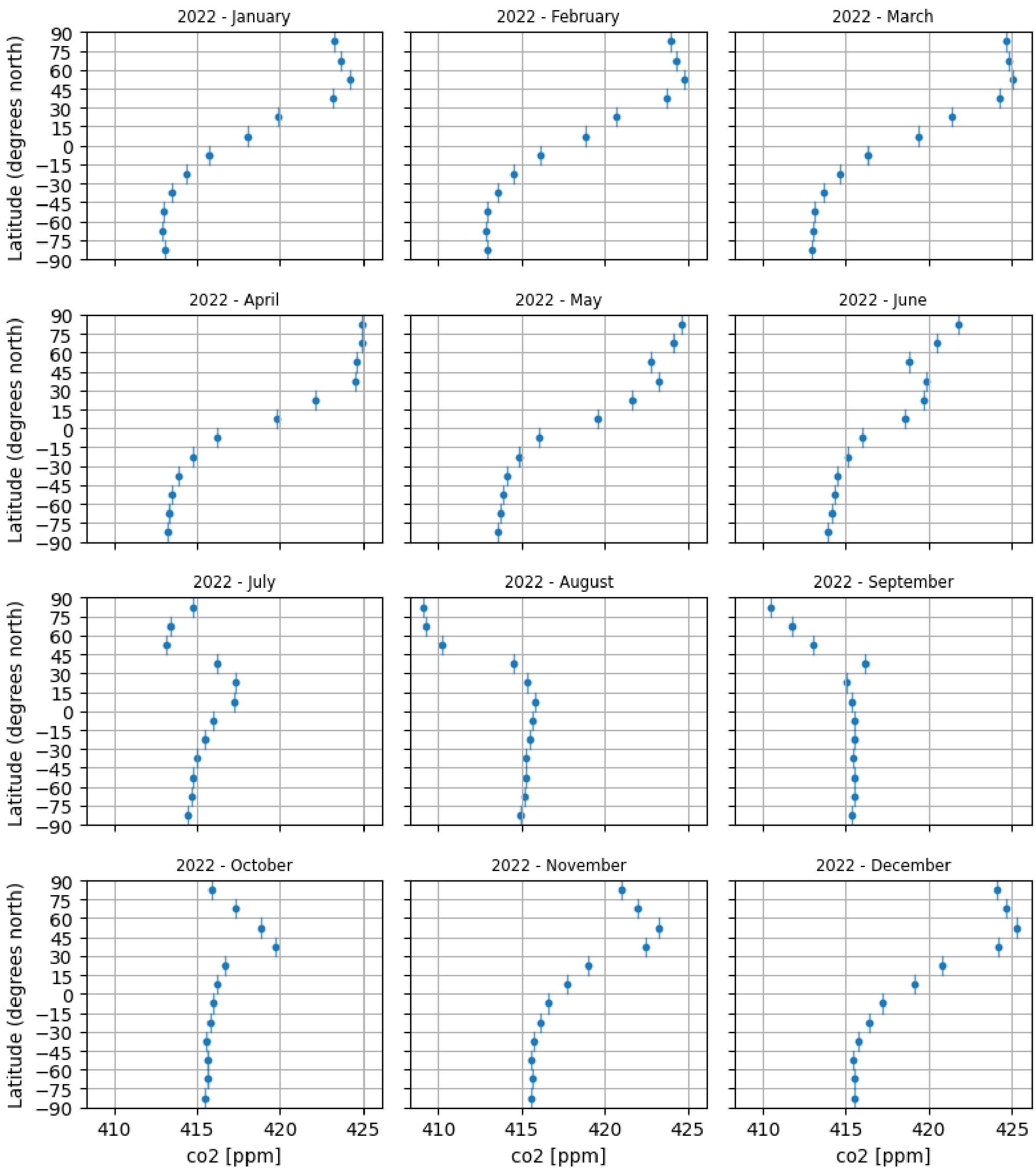
```
    time_bnds  (time, bnds) object 388kB 0001-01-01 00:00:00 ... 2023-01-01 0...
    lat_bnds   (lat, bnds) float64 192B -90.0 -75.0 -75.0 ... 75.0 75.0 90.0
Attributes: (12/29)
    Conventions:             CF-1.7
    activity_id:             input4MIPs
    comment:                 Data compiled by Climate Resource, based on scie...
    contact:                 zebedee.nicholls@climate-resource.com;malte.mein...
    creation_date:           2025-02-28T18:23:23Z
    dataset_category:        GHGConcentrations
    ...                      ...
    references_urls:         https://scrippsco2.ucsd.edu/data/atmospheric_co2...
    source_id:               CR-CMIP-1-0-0
    source_version:          1.0.0
    target_mip:              CMIP
    tracking_id:             hdl:21.14100/f6635404-8a1a-4aa9-918d-3792e8321f04
    variable_id:             co2
```

For this data, the latitudinal bounds show the area over which each point is the average.

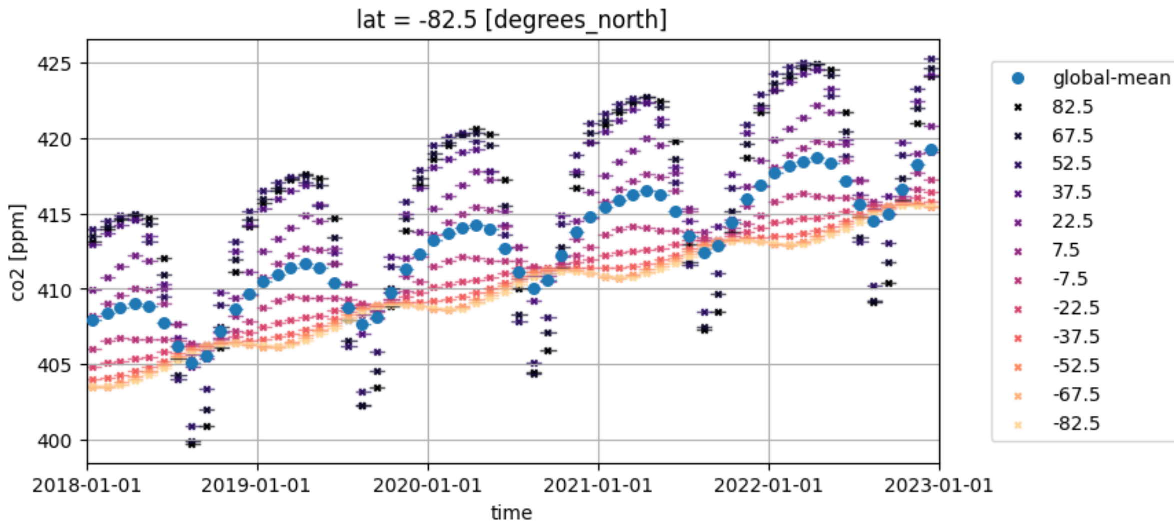
```
ds_co2_monthly_lat["lat_bnds"]
```

```
<xarray.DataArray 'lat_bnds' (lat: 12, bnds: 2)> Size: 192B
array([[-90., -75.],
       [-75., -60.],
       [-60., -45.],
       [-45., -30.],
       [-30., -15.],
       [-15.,   0.],
       [  0.,  15.],
       [ 15.,  30.],
       [ 30.,  45.],
       [ 45.,  60.],
       [ 60.,  75.],
       [ 75.,  90.]])
Coordinates:
  * lat      (lat) float64 96B -82.5 -67.5 -52.5 -37.5 ... 37.5 52.5 67.5 82.5
Dimensions without coordinates: bnds
```

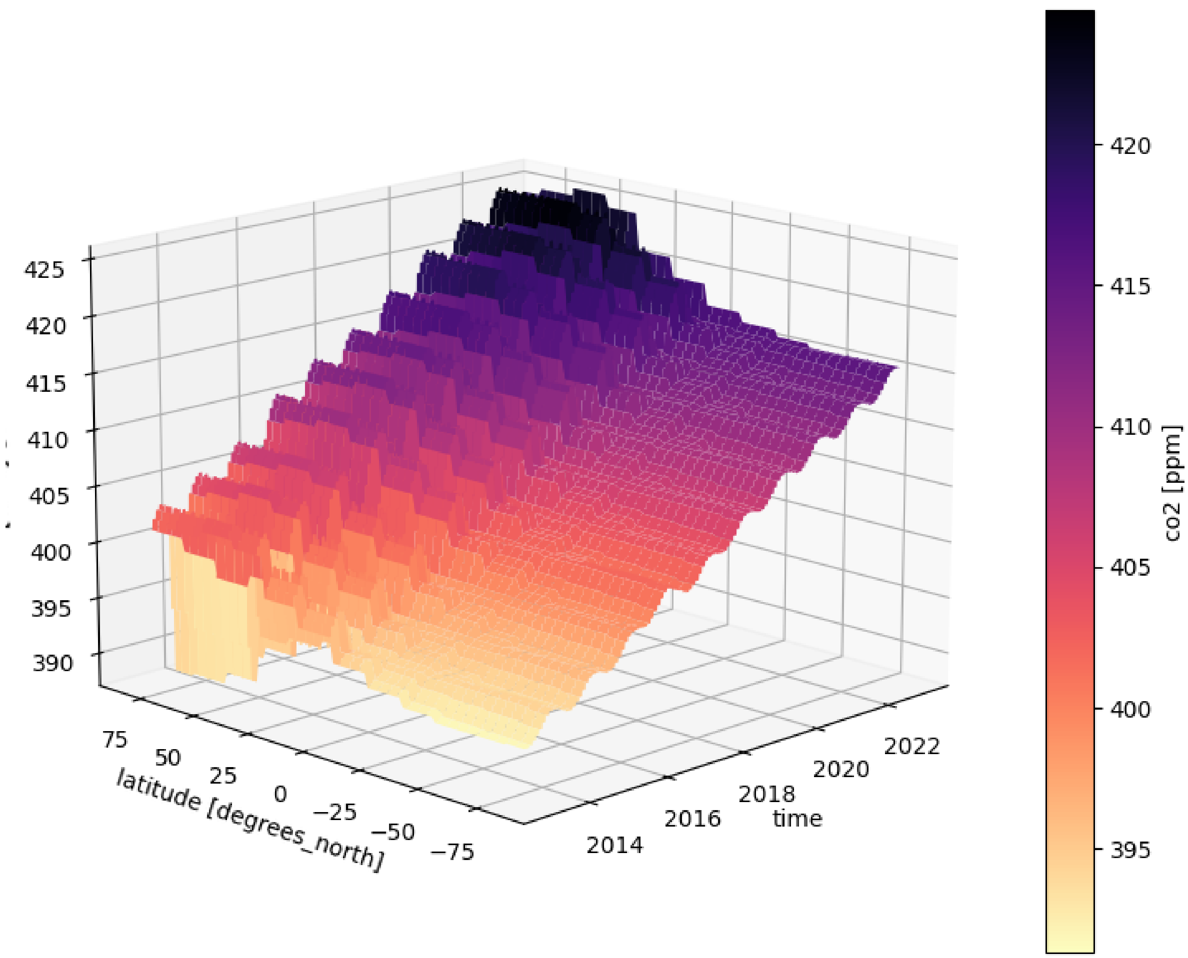
As above, but this time for the spatial axis, it is inappropriate to plot this data using a line plot. Scatter or step plots should be used instead.



We can compare the global-mean data to the data at each latitude. The strength of the latitudinal gradient varies also by gas (not shown).



The data can also be plotted in a so-called "magic carpet" to see the variation in space and time simultaneously.



1.5.5 Differences from CMIP6

File formats and naming

The file formats are generally close to CMIP6. There are three key changes:

1. we have split the global-mean and hemispheric-mean data into separate files. In CMIP6, this data was in the same file (with a grid label of `GMNHSH`). We have split this for two reasons: a) `GMNHSH` is not a grid label recognised in the CMIP CVs [REF-TODO] and b) having global-mean and hemispheric-mean data in the same file required us to introduce a 'sector' coordinate, which was confusing and does not follow the CF-conventions.
2. we have split the files into different time components. One file goes from year 1 to year 999 (inclusive). The next file goes from year 1000 to year 1749 (inclusive). The last file goes from year 1750 to year 2022 (inclusive). This simplifies handling and allows groups to avoid loading data they are not interested in (for CMIP, this generally means data pre-1750).
3. we have simplified the names of all the variables. They are now simply the names of the gases, for example we now use "co2" rather than "mole_fraction_of_carbon_dioxide". A full mapping is provided below.

There is one more minor change. The data now starts in year one, rather than year zero. We do this because year zero doesn't exist in most calendars (and we want to avoid users of the data having to hack around this when using standard data analysis tools).

Variable name mapping

```
CMIP6_TO_CMIP7_VARIABLE_MAP = {
    # name in CMIP6: name in CMIP7
    "mole_fraction_of_carbon_dioxide_in_air": "co2",
    "mole_fraction_of_methane_in_air": "ch4",
    "mole_fraction_of_nitrous_oxide_in_air": "n2o",
    "mole_fraction_of_c2f6_in_air": "c2f6",
    "mole_fraction_of_c3f8_in_air": "c3f8",
    "mole_fraction_of_c4f10_in_air": "c4f10",
    "mole_fraction_of_c5f12_in_air": "c5f12",
    "mole_fraction_of_c6f14_in_air": "c6f14",
    "mole_fraction_of_c7f16_in_air": "c7f16",
    "mole_fraction_of_c8f18_in_air": "c8f18",
    "mole_fraction_of_c_c4f8_in_air": "cc4f8",
    "mole_fraction_of_carbon_tetrachloride_in_air": "ccl4",
    "mole_fraction_of_cf4_in_air": "cf4",
    "mole_fraction_of_cfc11_in_air": "cfc11",
    "mole_fraction_of_cfc113_in_air": "cfc113",
    "mole_fraction_of_cfc114_in_air": "cfc114",
    "mole_fraction_of_cfc115_in_air": "cfc115",
    "mole_fraction_of_cfc12_in_air": "cfc12",
    "mole_fraction_of_ch2cl2_in_air": "ch2cl2",
    "mole_fraction_of_methyl_bromide_in_air": "ch3br",
    "mole_fraction_of_ch3ccl3_in_air": "ch3ccl3",
    "mole_fraction_of_methyl_chloride_in_air": "ch3cl",
    "mole_fraction_of_chcl3_in_air": "chcl3",
    "mole_fraction_of_halon1211_in_air": "halon1211",
    "mole_fraction_of_halon1301_in_air": "halon1301",
    "mole_fraction_of_halon2402_in_air": "halon2402",
    "mole_fraction_of_hcfc141b_in_air": "hcfc141b",
    "mole_fraction_of_hcfc142b_in_air": "hcfc142b",
    "mole_fraction_of_hcfc22_in_air": "hcfc22",
```

(continues on next page)

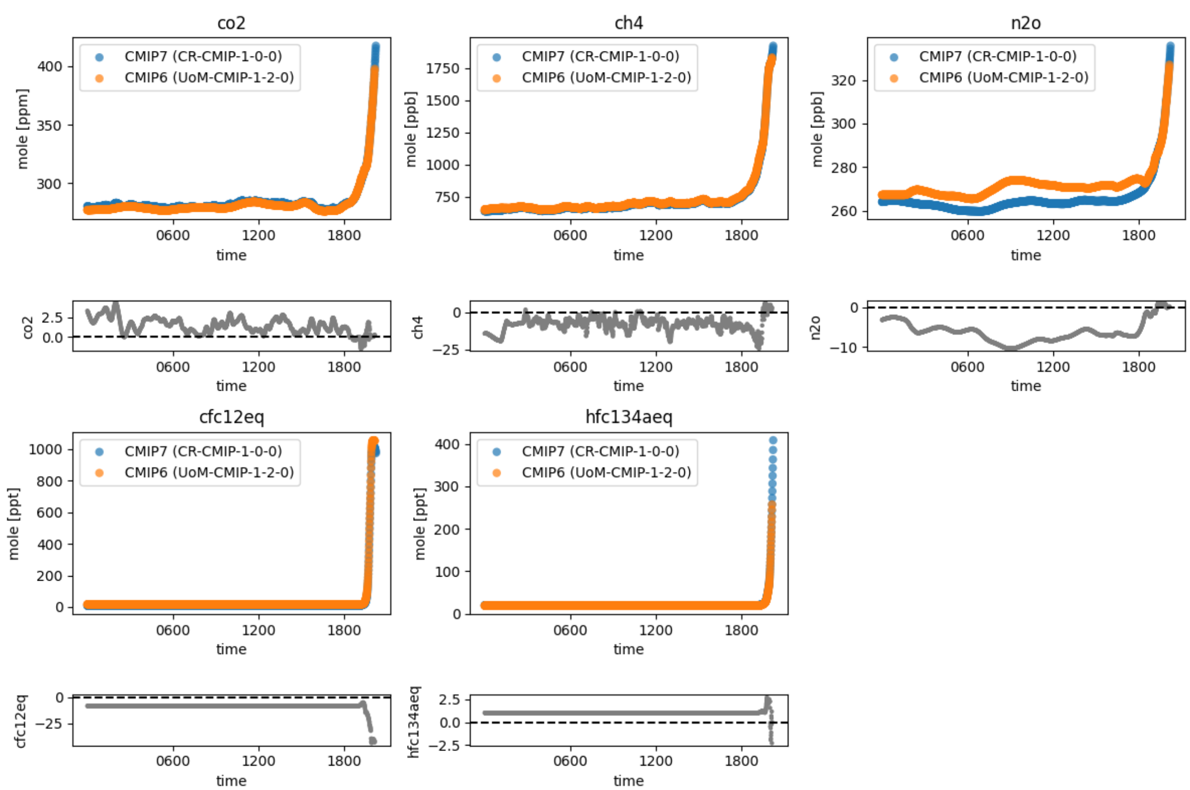
(continued from previous page)

```
    "mole_fraction_of_hfc125_in_air": "hfc125",
    "mole_fraction_of_hfc134a_in_air": "hfc134a",
    "mole_fraction_of_hfc143a_in_air": "hfc143a",
    "mole_fraction_of_hfc152a_in_air": "hfc152a",
    "mole_fraction_of_hfc227ea_in_air": "hfc227ea",
    "mole_fraction_of_hfc23_in_air": "hfc23",
    "mole_fraction_of_hfc236fa_in_air": "hfc236fa",
    "mole_fraction_of_hfc245fa_in_air": "hfc245fa",
    "mole_fraction_of_hfc32_in_air": "hfc32",
    "mole_fraction_of_hfc365mfc_in_air": "hfc365mfc",
    "mole_fraction_of_hfc4310mee_in_air": "hfc4310mee",
    "mole_fraction_of_nf3_in_air": "nf3",
    "mole_fraction_of_sf6_in_air": "sf6",
    "mole_fraction_of_so2f2_in_air": "so2f2",
    "mole_fraction_of_cfc11eq_in_air": "cfc11eq",
    "mole_fraction_of_cfc12eq_in_air": "cfc12eq",
    "mole_fraction_of_hfc134aeq_in_air": "hfc134aeq",
}
```

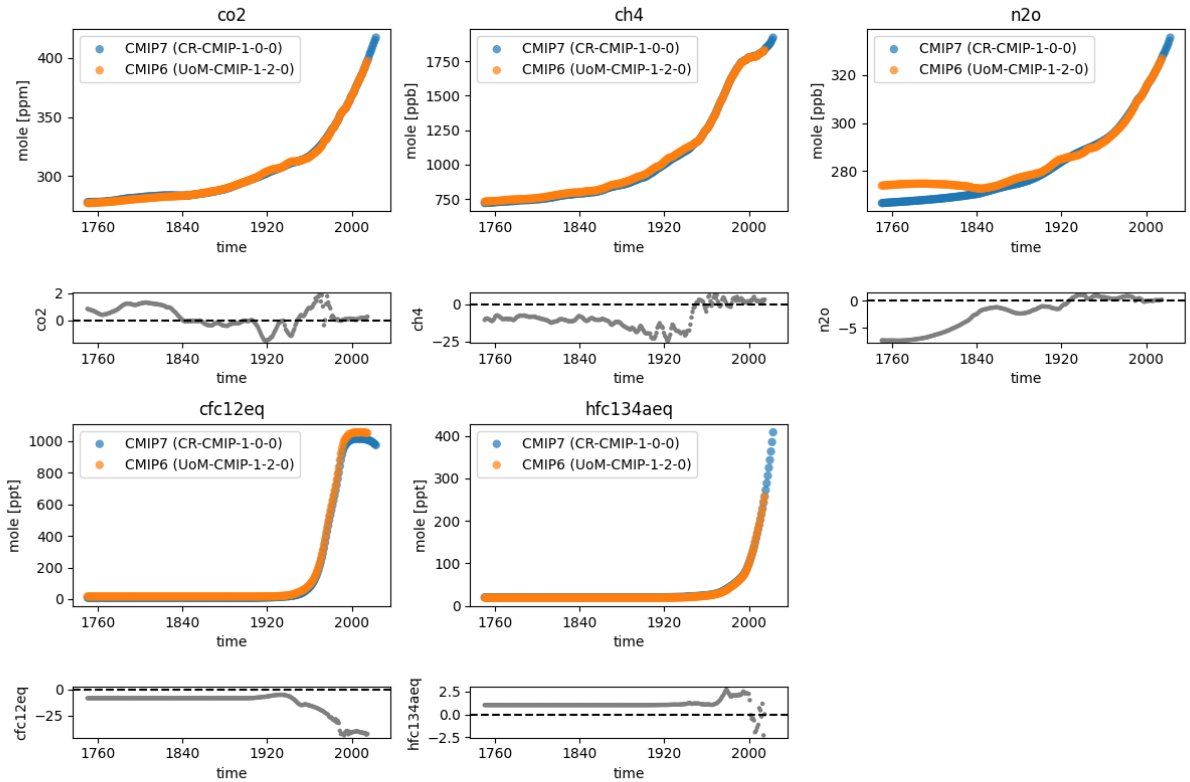
Data comparisons

Comparing the data from CMIP6 and CMIP7 shows minor changes (although doing this comparison requires a bit of care because of the changes in file formats).

Atmospheric concentrations: Year 1 - 2022

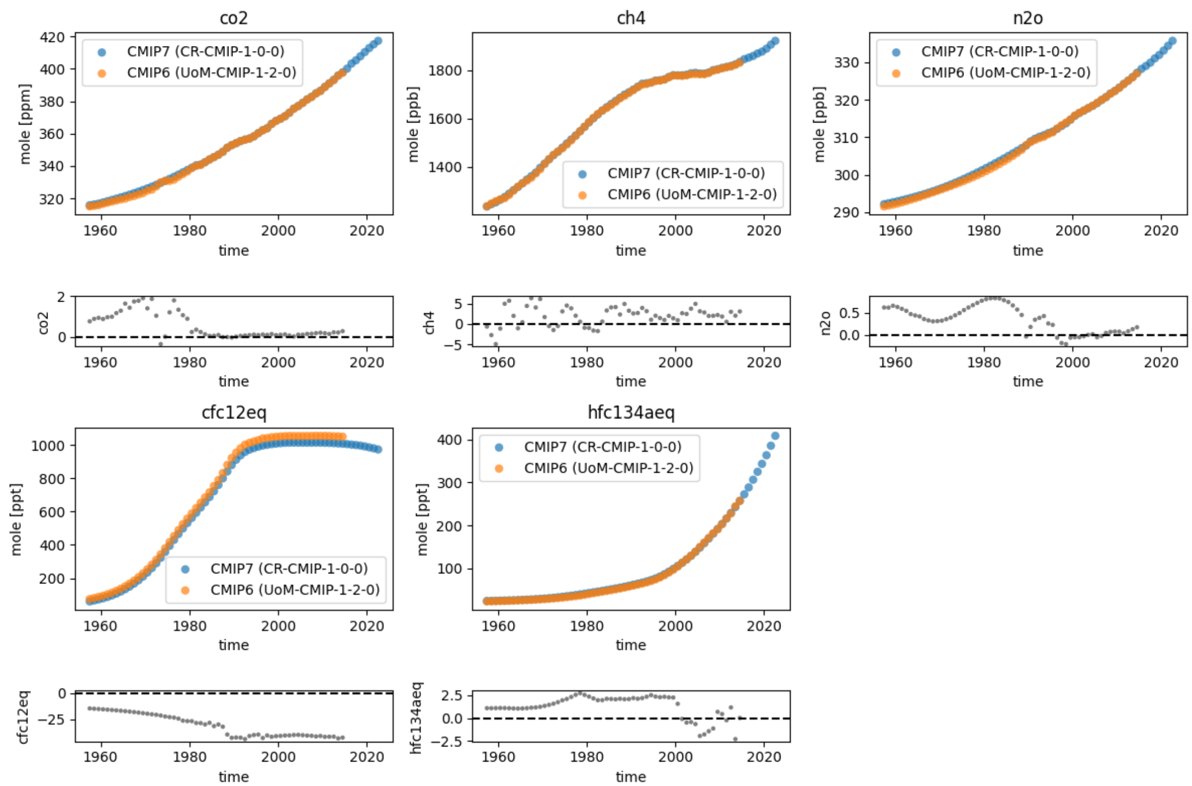


Atmospheric concentrations: Year 1750 - 2022



Atmospheric concentrations: Year 1957 - 2022

1957 is the start of the Scripps ground-based record. Before this, data is based on ice cores alone.



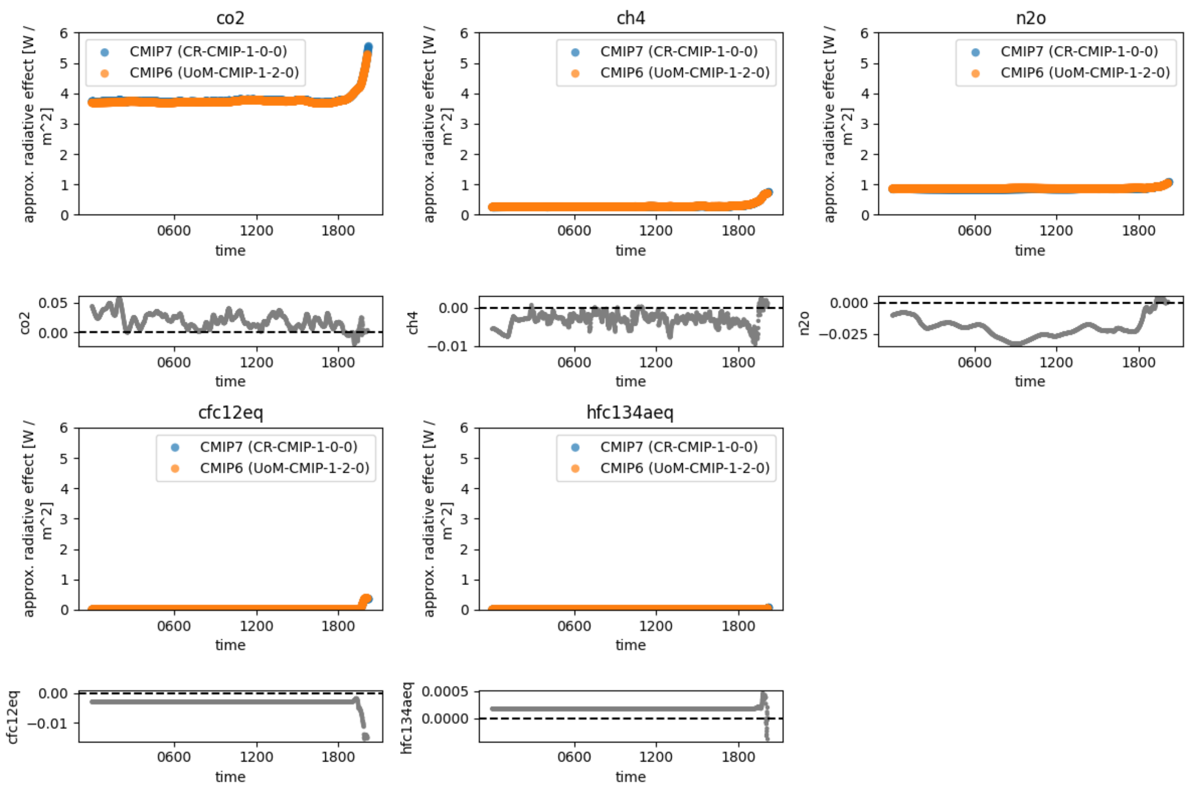
Approximate radiative effect: Year 1 - 2022

As seen above, in atmospheric concentration terms the differences are small. However, this can be put on a common scale by comparing the differences in radiative effect terms. This gives an approximation of the size of the difference that would be seen by an Earth System Model's (ESM's) radiation code. This uses basic linear approximations, assuming that the radiative effect of each gas is simply its atmospheric concentration multiplied by a constant. This isn't the same as effective radiative forcing (ERF). For that comparison, see the later sections focussed on ERF.

Values below come from Table 7.SM.7 of IPCC AR7 WG1 Ch. 7 Supplementary Material[4].

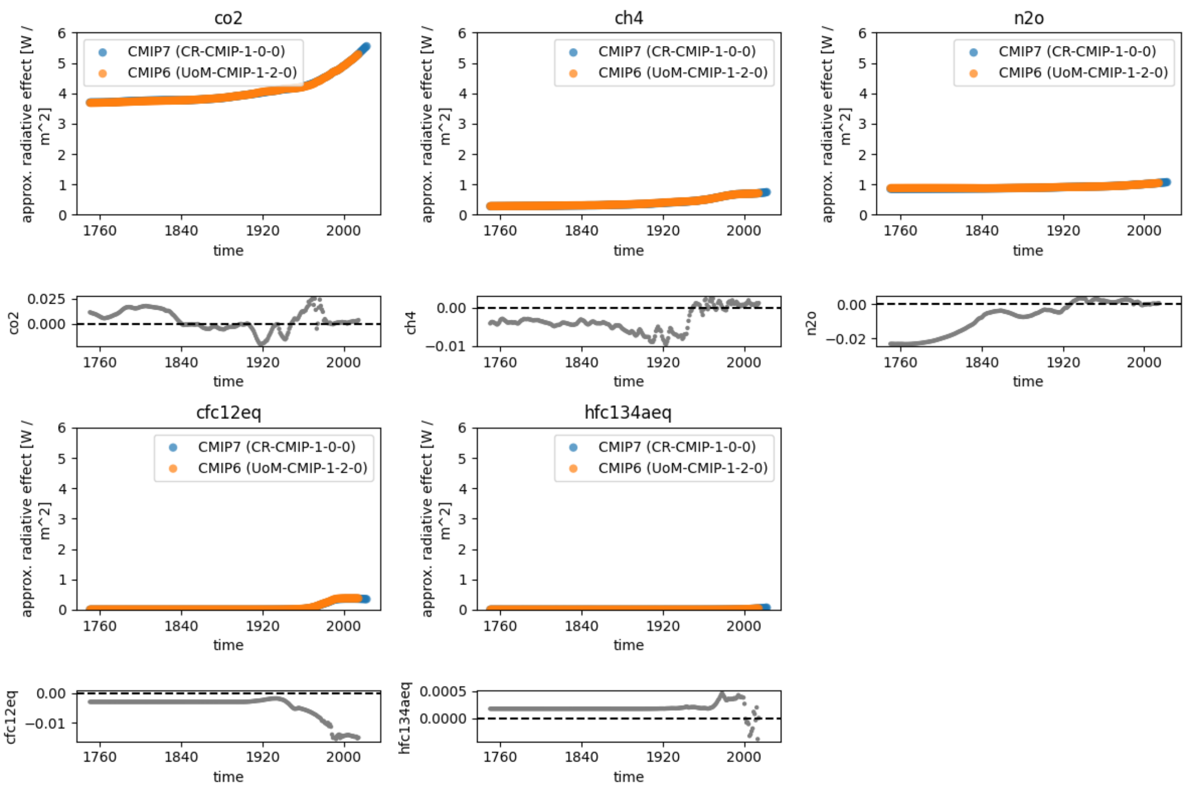
```
from openscm_units import unit_registry

Q = unit_registry.Quantity

RADIATIVE_EFFICIENCIES = {
    "co2": Q(1.33e-5, "W / m^2 / ppb"),
    "ch4": Q(3.88e-4, "W / m^2 / ppb"),
    "n2o": Q(3.2e-3, "W / m^2 / ppb"),
    "cfc12eq": Q(0.358, "W / m^2 / ppb"),
    "hfc134aeq": Q(0.167, "W / m^2 / ppb"),
}
```

[4] https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter07_SM.pdf



Approximate radiative effect: Year 1750 - 2022

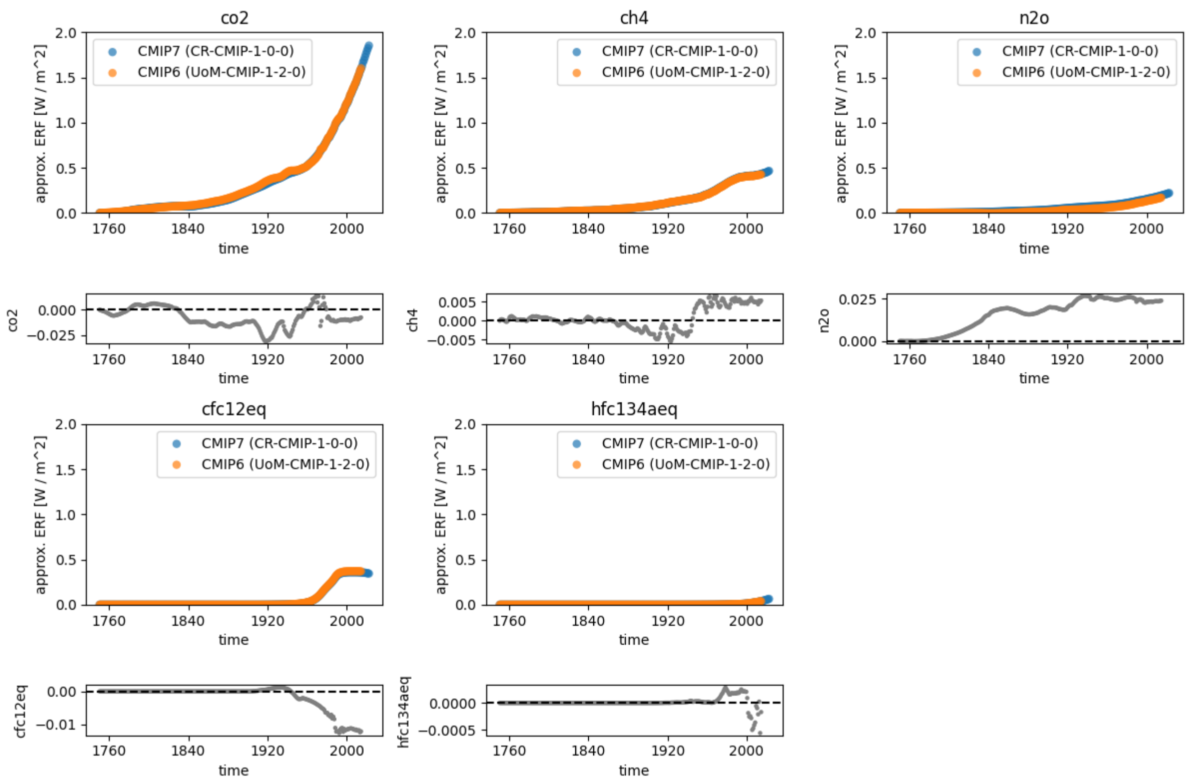
This is the period relevant for historical simulations in CMIP.



Approximate effective radiative forcing: Year 1750 - 2022

The above isn't effective radiative forcing. For that, you have to normalise the data to some reference year. There are a few different choices for this reference year. In IPCC reports, it is 1750 so that is what we show here. It should be noted that some ESMs may make other choices, but these would not have a great effect on the interpretation of the difference between the CMIP6 and CMIP7 datasets.

Note that this approximation is linear, which is a particularly strong approximation for CO2 because of its logarithmic forcing nature. We show this approximation here nonetheless because it provides an order of magnitude estimate for the change from CMIP6 in ERF terms. The forthcoming manuscripts will explore the subtleties of this quantification in more detail.

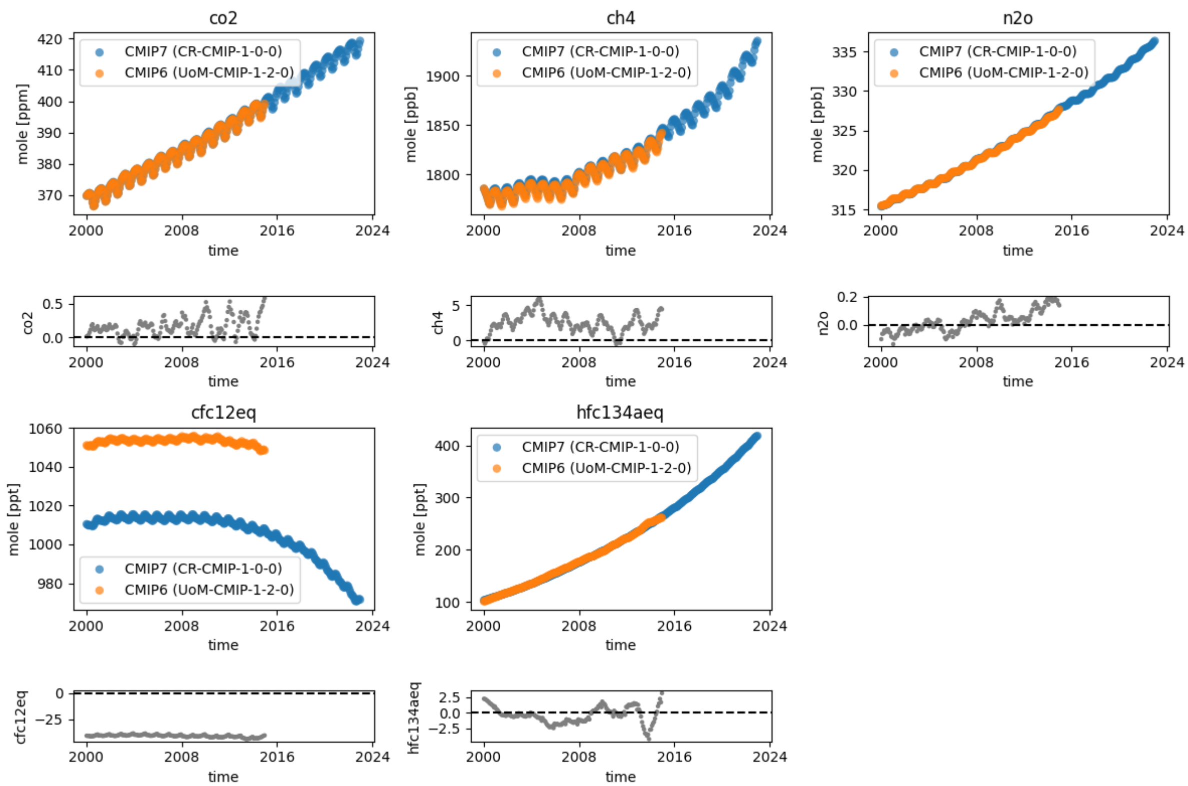


In summary, in ERF terms, the differences from CMIP6 are very small. For all gases, they are less than around 0.025 W / m2. Compared to the estimated total greenhouse gas forcing and uncertainty in IPCC AR6 (see Section 7.3.5.2 of AR6 WG1 Chapter 7[5]), estimated to be 3.84 W / m2 (very likely range of 3.46 to 4.22 W / m2), such differences are particularly small.

Atmospheric concentrations including seasonality: Year 2000 - 2022

The final comparisons we show are atmospheric concentrations including seasonality. Given that most greenhouse gases are well-mixed with lifetimes much greater than a year, these differences are unlikely to be of huge interest to ESMs. However, for other applications, such seasonality differences may matter more.

[5] https://www.ipcc.ch/report/ar6/wg1/chapter/chapter-7/



Like the annual-means, the atmospheric concentrations including seasonality are reasonably consistent between CMIP6 and CMIP7. There are some areas of change. Full details of these changes will be provided in the forthcoming manuscripts.

BIBLIOGRAPHY

[1] Malte Meinshausen, Elisabeth Vogel, Alexander Nauels, Katja Lorbacher, Nicolai Meinshausen, David M. Etheridge, Paul J. Fraser, Stephen A. Montzka, Peter J. Rayner, Cathy M. Trudinger, Paul B. Krummel, Urs Beyerle, Josep G. Canadell, John S. Daniel, Ian G. Enting, Rachel M. Law, Chris R. Lunder, Simon O'Doherty, Ron G. Prinn, Stefan Reimann, Mauro Rubino, Guus J. M. Velders, Martin K. Vollmer, Ray H. J. Wang, and Ray Weiss. Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, 10(5):2057–2116, 2017. doi:10.5194/gmd-10-2057-2017.

[2] X. Lan, J.W. Mund, A.M. Crotwell, K.W. Thoning, E. Moglia, M. Madronich, K. Baugh, G. Petron, M.J. Crotwell, D. Neff, T. Wolter, T. Mefford, and S. DeVogel. Atmospheric Carbon Dioxide Dry Air Mole Fractions from the NOAA GML Carbon Cycle Cooperative Global Air Sampling Network, 1968-2024. 2025. URL: https://www.esrl.noaa.gov/gmd/ccgg/arc/?id=132 (visited on 2025-07-14), doi:10.15138/WKGJ-F215.

[3] X. Lan, J.W. Mund, A.M. Crotwell, K.W. Thoning, E. Moglia, M. Madronich, K. Baugh, G. Petron, M.J. Crotwell, D. Neff, T. Wolter, T. Mefford, and S. DeVogel. Atmospheric Methane Dry Air Mole Fractions from the NOAA GML Carbon Cycle Cooperative Global Air Sampling Network, 1983-2024. 2025. URL: https://www.esrl.noaa.gov/gmd/ccgg/arc/?id=132 (visited on 2025-07-14), doi:10.15138/VNCZ-M766.

[4] R. G. Prinn, R. F. Weiss, P. J. Fraser, P. G. Simmonds, D. M. Cunnold, F. N. Alyea, S. O'Doherty, P. Salameh, B. R. Miller, J. Huang, R. H. J. Wang, D. E. Hartley, C. Harth, L. P. Steele, G. Sturrock, P. M. Midgley, and A. McCulloch. A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE. *Journal of Geophysical Research: Atmospheres*, 105(D14):17751–17792, 2000. Publisher: American Geophysical Union (AGU). doi:10.1029/2000jd900141.

[5] Ronald G Prinn, Ray F Weiss, Jgor Arduini, Tim Arnold, H Langley DeWitt, Paul J Fraser, Anita L Ganesan, Jimmy Gasore, Christina M Harth, Ove Hermansen, and others. History of chemically and radiatively important atmospheric gases from the advanced global atmospheric gases experiment (agage). *Earth System Science Data Discussions*, 2018:1–39, 2018. doi:10.5194/essd-10-985-2018.

[6] Matt Rigby, Ronald G Prinn, Paul Joseph Fraser, Peter Guy Simmonds, RL Langenfelds, Jin Huang, Derek Martin Cunnold, Lloyd Paul Steele, Paul Brian Krummel, Ray F Weiss, and others. Renewed growth of atmospheric methane. *Geophysical research letters*, 2008. doi:10.1029/2008GL036037.

[7] Matthew Rigby, Stephen A Montzka, Ronald G Prinn, James WC White, Dickon Young, Simon O'doherty, Mark F Lunt, Anita L Ganesan, Alistair J Manning, Peter G Simmonds, and others. Role of atmospheric oxidation in recent methane growth. *Proceedings of the National Academy of Sciences*, 114(21):5373–5377, 2017. doi:10.1073/pnas.1616426114.