Geoscientific
Model Development

# The generation of gridded emissions data for CMIP6

**Leyang Feng**[1], **Steven J. Smith**[1], **Caleb Braun**[1], **Monica Crippa**[2], **Matthew J. Gidden**[3,4], **Rachel Hoesly**[1],
**Zbigniew Klimont**[4], **Margreet van Marle**[5], **Maarten van den Berg**[6], and **Guido R. van der Werf**[7]

[1]Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA
[2]European Commission, Joint Research Centre (JRC), Ispra, Italy
[3]Climate Analytics, Berlin, Germany
[4]International Institute for Applied Systems Analysis, Laxenburg, Austria
[5]Deltares, Delft, the Netherlands
[6]PBL Netherlands Environmental Assessment Agency, The Hague, the Netherlands
[7]Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

**Correspondence:** Steven J. Smith (ssmith@pnnl.gov)

**Abstract.** Spatially distributed anthropogenic and open burning emissions are fundamental data needed by Earth system models. We describe the methods used for generating gridded datasets produced for use by the modeling community, particularly for the Coupled Model Intercomparison Project Phase 6. The development of three sets of gridded data for historical open burning, historical anthropogenic, and future scenarios was coordinated to produce consistent data over 1750–2100. Historical data up to 2014 were provided with annual resolution and future scenario data in 10-year intervals. Emissions are provided on a sectoral basis, along with additional files for speciated non-methane volatile organic compounds (NMVOCs). An automated framework was developed to produce these datasets to ensure that they are reproducible and facilitate future improvements. We discuss the methodologies used to produce these data along with limitations and potential for future work.

## 1 Introduction

Anthropogenic activities, from the generation of electricity to the ignition of forest fires, result in the emissions of gases and aerosol species into the atmosphere. These emissions, in turn, alter atmospheric composition, deposition rates, and the Earth's radiative balance. Emissions have, for the most part, increased over the 20th century, leading to higher aerosol concentrations and tropospheric ozone levels as industrial ac- tivities and fossil fuel consumption increased. Over the past couple of decades air pollutant emissions have shifted away from North America and Europe, due in large part to air pollution controls, to East and South Asia, driven by rapid industrialization and population growth in those regions.

One of the important tools used to examine these impacts are models of the Earth system, including chemical transport, climate–chemistry, and Earth system models. These models require gridded emissions data in order to simulate the impact of these emissions within the Earth system. Furthermore, for multi-decadal to century-long model runs, temporally and spatially consistent emissions data are needed. Here we describe the production of several related gridded datasets that were produced, in large part, to facilitate the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016). These datasets contain anthropogenic chemically reactive gases ($CO$, $CH_4$, $NH_3$, $NO_x$, $SO_2$, NMVOC), carbonaceous aerosols (BC and OC) – the principal climatically important primary particle emission species, and fossil $CO_2$ emissions, with non-methane organic hydrocarbons (NMVOCs) also provided as separate species (Table 1). These emission data products are needed for emissions-driven simulations over both historical and future time periods for ScenarioMIP, AerChemMIP, and C4MIP (Eyring et al., 2016; Collins et al., 2017; Jones et al., 2016). Using these emission datasets, models can estimate atmospheric concentrations of primary and secondary fine particulate and tropo-

spheric ozone, which impact human health, ecosystem functioning, and the Earth's radiative balance.

The first dataset considered here are anthropogenic emissions, which are defined as emissions that stem directly from human activities such as energy transformation, buildings, transportation, and agricultural and industrial activities (see Hoesly et al., 2018, for a complete listing). Historical emissions gridding is also discussed in Hoesly et al. (2018), along with an extensive description of the methodologies used to produce country-level historical anthropogenic emissions. We provide a more complete discussion of the gridding methodology, a description of the closely related methods used for gridding future emissions, and related supplementary data such as speciated VOCs. Compared to previous datasets, our anthropogenic data have a greater degree of consistency across species and over time, with seasonality for all species. The historical anthropogenic emissions were produced by the Community Emissions Data System (CEDS). This paper focuses on the data produced for CMIP6 (see Appendix A). Updated emission data releases for general scientific use by the CEDS project are in progress (e.g., Hoesly et al., 2019).

The second dataset is historical open burning emissions, which are defined as forest, grassland, and peatland fires, along with agricultural waste burning (AWB) on fields. While open burning emissions can also have anthropogenic drivers, these emissions are in a separate category here and elsewhere in the literature, as the techniques for estimating are generally distinct from the methods used for "anthropogenic" emissions. The open burning emissions over recent years are from the Global Fire Emissions Database version 4 with small fires (GFED4s; Van der Werf et al., 2017), which have been driven by satellite data since 1997. Estimates for earlier years are based on proxies and fire models. We only briefly discuss this dataset in this paper since those data are inherently in gridded form during their development, as described in detail by van Marle et al. (2017). Note that open burning emissions are often described in the literature as "biomass burning", but we do not use this term to avoid confusion with anthropogenic biofuel combustion such as biofuel use in cookstoves, which is included in the anthropogenic emissions dataset.

The third set of data comprises gridded data over the future (2015–2100) for these same species for both anthropogenic and open burning sectors for selected future scenarios. These future gridded data were produced using a variation of the same gridding methodologies used for the historical anthropogenic data, which is why these are discussed together in this paper. The future emission trajectories are discussed in Gidden et al. (2019), with the gridding methodology described in more detail herein. As discussed below, the future gridded emissions build on these two historical datasets and, in large part, inherit their properties such as within-country spatial distribution and seasonality.

A number of global gridded datasets have been produced over the years. One of the most widely used datasets is the Emissions Database for Global Atmospheric Research (EDGAR), which provides an independent estimate of historical greenhouse gas (GHG) and pollutant emissions by country, sector, and spatial grid ($0.1 \times 0.1°$; Crippa et al., 2016; EC-JRC/PBL, 2016). The most recent release provides annual and monthly data over 1970–2012 (Crippa et al., 2018). The GAINS (Greenhouse gas–Air pollution Interactions and Synergies) model (Amann et al., 2015) has been used to produce regional and global emission estimates for recent years (1990–2010; in 5-year intervals). These estimates formed the basis for ECLIPSE emission data, also available in gridded form (Klimont et al., 2017). Other global datasets include the HTAP v2 (Janssens-Maenhout et al., 2015) emissions data used by the Task Force on Hemispheric Transport of Air Pollutants (HTAP). That dataset merged EDGAR with regional and country-level gridded emissions data for 2008 and 2010.

Lamarque et al. (2010) developed the historical dataset used in the Coupled Model Intercomparison Project Phase 5 (CMIP5), which included global gridded estimates of anthropogenic and open burning emissions for 1850–2000 at 10-year intervals. It was a compilation of the "best available estimates" from many sources including EDGAR-HYDE (van Aardenne et al., 2001), RETRO (Schultz and Sebastian, 2007), and emissions reported largely by Organization for Economic Cooperation and Development (OECD) countries over recent years. One focal point of that work was the compilation of a year 2000 emissions dataset that was used as the starting point for the future projections. See Hoesly et al. (2018) and van Marle et al. (2017) for a comparison of the CMIP6 and CMIP5 anthropogenic and open burning datasets, respectively, and Gidden et al. (2019) for a comparison of the CMIP6 and CMIP5 projections.

We first discuss the overall methodologies for producing the gridded data, then present the gridded data and discuss the underlying properties of these data, focusing on the anthropogenic sources. The paper concludes with a discussion of the issues identified in these data and potential further work to improve their quality.

## 2 Data and methodology

### 2.1 Methodology overview

We first provide an overview of the gridding methodology, with further details provided in the following subsections. The gridding methodology for historical anthropogenic and all future emissions is summarized in Fig. 1. The fundamental underlying data used here are emissions by country and sector. This provides both total emission trends over time as well as changes in the sectoral composition of each emission species. As discussed in Hoesly et al. (2018), emissions in the preindustrial period were generally dominated by biofuel use

**Table 1.** Emission species provided in the emission data files. For the molecular weights assumed for each species, see https://GitHub.com/iiasa/emissions_downscaling/wiki (last access: 1 August 2019).

| Emissions dataset | Emission species provided |
|---|---|
| Anthropogenic bulk emissions | CO, $CH_4$, $NH_3$, $NO_x$, $SO_2$, NMVOC, BC, OC, $CO_2$ |
| Open burning bulk emissions | CO, $CH_4$, $NH_3$, $NO_x$, $SO_2$, NMVOC, BC, OC, $H_2$ |
| Speciated anthropogenic emissions | alcohols, ethane, propane, butanes, pentanes, hexanes_plus_higher_alkanes, ethene, propene, ethyne, other_alkenes_and_alkynes, benzene, toluene, xylene, trimethylbenzenes, other_aromatics, esters |
| Speciated open burning emissions | $C_{10}H_{16}$, $C_2H_2$, $C_2H_4$, $C_2H_4O$, $C_2H_5OH$, $C_2H_6$, $C_2H_6S$, $C_3H_6$, $C_3H_6O$, $C_3H_8$, $C_5H_8$, $C_6H_6$, $C_7H_8$, $C_8H_{10}$, $CH_2O$, $CH_3COCHO$, $CH_3COOH$, $CH_3OH$, HCN, HCOOH, higher-alkanes, higher-alkenes, $HOCH_2CHO$, MEK, toluene-lump |
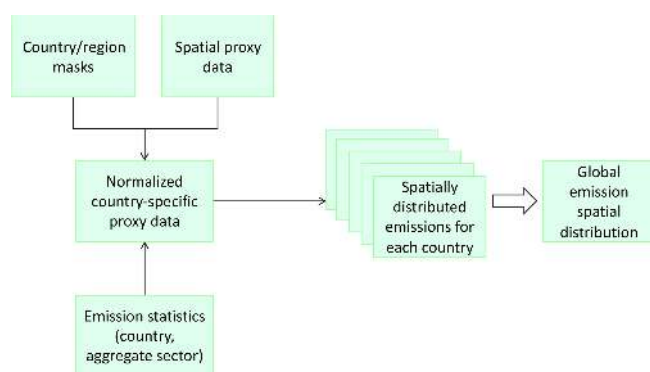


**Figure 1.** Emissions gridding overview. Emissions data at the country and aggregate sector level are mapped to spatial grids separately by country and sector, then combined into global emissions grids as described in the main text.

in the residential sector. As industrialization proceeded emissions from the industrial, energy transformation, and transportation sectors became increasingly important. Emissions are translated to a spatial grid for each country and aggregate gridding sector (see Table 2). The methodology applied to future emissions is similar to that used for historical emissions, although with a lower sectoral and temporal resolution, as detailed below. This comprehensive gridded dataset is therefore produced with a consistent methodology by sector across all emission species spanning 1750 through 2014, with a consistent set of future projections over 2015–2100. The details of the methodology are first described for the historical emissions data (1750–2014), followed by a discussion of areas in which the methodology differs for future emissions (2015–2100). Further details including code are available online for both historical and future gridding (Sect. 5 below).

## 2.2 Historical anthropogenic emissions

Aggregate anthropogenic emissions over the historical period by country and CEDS sector from Hoesly et al. (2018) are aggregated to 16 intermediate sectors (Table 2) and

mapped to a $0.5 \times 0.5°$ grid as described below. The intermediate sectors were selected according to the availability of gridded proxy data in order to provide as much spatial detail as possible. For the final distribution, gridded emissions are further aggregated to nine sectors: agriculture, energy, industrial, transportation, residential–commercial–other, solvents, waste, international shipping, and aircraft. These final sectors were chosen to be consistent with previous practice (e.g., Lamarque et al., 2010), which includes matching the generally lower level of sectoral disaggregation available from future projections, and to reduce overall file size. Gridded emissions data are produced as CF-compliant NetCDF files (http://cfconventions.org/, last access: 4 February 2020) and converted to flux ($kg\,m^{-2}\,s^{-1}$). Data were aggregated into a number of multiyear files, keeping individual file sizes under about 1 Gb. Filenames and NetCDF metadata are generated so that files are ready for distribution using input4MIPS on the Earth System Grid Federation system.

Emissions at the country and gridding sector level are mapped to a spatial grid using a variety of spatial proxy data, as described below. Emissions are distributed into target grid cells $XY$ as

$$\text{emission}_{XY} = \text{emission}_c \times \frac{\text{proxy\_value}_{XY}}{\sum \text{proxy\_value}_{XY}}, \qquad (1)$$

where $\text{emission}_{XY}$ is the emissions value assigned to coordinate $XY$; $\text{emission}_c$ is the total country-level emissions for the gridding sector; $\text{proxy\_value}_{XY}$ is the value of the proxy data for the cell $XY$, and the sum is performed over all coordinates $XY$ that are within the specified country.

The proxy data values are preprocessed for each country by multiplying the data by the area fraction of each grid cell that is in the specified country at an annual resolution (if annual gridded data are available). In this way grid cells that contain multiple countries will be assigned emissions proportionately. After assigning emissions to spatial grids by country, the spatially distributed emissions from each country are then added into a global spatial matrix. The "country" list includes a global region for emissions related to interna-

**Table 2.** Proxy data used for gridding anthropogenic emissions data, adapted from Hoesly et al. (2018). Gridding proxies marked with an asterisk (*) do not vary in time. Future emissions use the same final gridding sectors, with the addition of four sectors for open burning emissions (e.g., agricultural waste burning on fields (AWB), forest burning, grassland burning, and peatland burning) and an additional sector for net negative $CO_2$ emissions (NEGCO2). A decadal average (2005–2014 avg.) for open burning proxy data was used for future emissions as indicated. Anthropogenic sector definitions can be found in Hoesly et al. (2018). Note that, except for international shipping, all future emissions used proxy data at the country and sector level that did not change over time; see the main text.

| Final gridding sector | Intermediate gridding sector definition | Proxy data source | Years |
|---|---|---|---|
| **Historical anthropogenic emissions** | | | |
| Residential, commercial, other (RCO) | Residential, commercial, other (residential and commercial) | HYDE population (interpolated annually) | 1750–1899 |
| | | EDGAR v4.2 (1970) blended with HYDE | 1900–1969 |
| | | EDGAR v4.2 RCORC | 1970–2008 |
| | Residential, commercial, other (other) | HYDE population (interpolated annually) | 1750–1899 |
| | | EDGAR v4.2 (1970) blended with HYDE | 1900–1969 |
| | | EDGAR v4.2 RCOO | 1970–2008 |
| Agriculture (AGR) | Agriculture | EDGAR v4.2 AGR | 1970–2008 |
| Energy sector (ENE) | Electricity and heat production | EDGAR v4.2 ELEC | 1970– 2008 |
| | Fossil fuel fires | EDGAR v4.2 FFFI | 1970–2008 |
| | Fuel production and transformation | EDGAR v4.2 ETRN | 1970–2008 |
| | Oil and gas fugitive–flaring | ECLIPSE FLR 1990, 2000, 2010 EDGAR v4.2 ETRN (1970–2008) | 1970–2010 |
| Industrial sector (IND) | Industrial combustion | EDGAR v4.2 INDC | 1970–2008 |
| | Industrial process and product use | EDGAR v4.2 INPU | 1970–2008 |
| Transportation section (TRA) | Road transportation | EDGAR v4.3.2 ROAD (2010)* | 1750–2014 |
| | Non-road transportation | EDGAR v4.2 NRTR | 1970–2008 |
| International shipping (SHP) | International shipping | ECLIPSE + additional data (1990–2015) | 1990–2010 |
| | Tanker loading | ECLIPSE + additional data (1990–2015) | 1990–2010 |
| Solvents (SLV) | Solvents production and application | EDGAR v4.2 SLV | 1970–2008 |
| Waste (WST) | Waste | HYDE Population, GPW v3 (modified rural population) | 1750–2014 |
| Aircraft (AIR) | Aircraft | CMIP5 (Lamarque et al., 2010; Lee et al., 2009)* | 1850–2008 |
| **Additional sector for future anthropogenic emissions** | | | |
| Negative $CO_2$ (NEGCO2) | Net negative $CO_2$ emissions | See main text* | 2015–2100 |
| **Future open burning emission sectors** | | | |
| Forests | Forest fires, including deforestation | 2005–2014 average (van Marle et al., 2017) | 2015–2100 |
| Grassland | Grassland fires | 2005–2014 average (van Marle et al., 2017) | 2015–2100 |
| Peat | Peatland fires | 2005–2014 average (van Marle et al., 2017) | 2015–2100 |
| Ag waste (AWB) | Agricultural waste burning on fields | 2005–2014 average (van Marle et al., 2017) | 2015–2100 |

tional shipping that are not associated with a particular country. Aircraft emissions are gridded separately, with one three-dimensional distribution that is scaled uniformly to match the global emissions estimate.

In most cases the proxy data are gridded emissions so as not to duplicate the effort needed to convert raw proxy data into the form needed for emissions inventories. We define two levels of spatial proxy data, the primary gridding proxies and a backup gridding proxy, which is gridded population. The backup gridding proxy is used where the primary gridding proxies are not appropriate due to either (1) the primary proxy not being available (e.g., is equal to zero) for the given country–sector–year combination or (2) the primary proxy being inaccurate for the given country–sector combination. In the latter case, we perform this proxy substitution when the ratio of proxy to sector emissions data for that country–sector combination is an outlier compared to the global distribution of this ratio across countries.

Over recent decades the primary gridding proxy data were from the EDGAR v4.2 (EC-JRC/PBL, 2012) inventory (Table 2) since these data were available over 1970 through 2008. Road transportation uses the EDGAR 4.3.2 road transportation grid, which is significantly improved over previous versions (Crippa et al., 2016) but was only available for 2010 at the time these data were produced, so the country-specific 2010 spatial distribution is used for all years. Flaring emissions use a blend of grids from EDGAR and ECLIPSE (Klimont et al., 2017). The backup gridded population proxy, as well as the proxy for early years for the residential–commercial sector, is a combination of gridded population from the Gridded Population of the World (GPW) (Doxsey-Whitfield et al., 2015) and HYDE (Goldewijk et al., 2011). Aircraft emission distributions are from Lee et al. (2009) and international shipping uses ECLIPSE shipping grids, with additional data from Endresen et al. (2003) for NMVOC emissions from oil tanker venting as used in Lamarque et al. (2010).

The only proxy data that vary before 1970 are those for the RCO (residential, commercial, other) and waste sectors (Table 2). For the RCO sector, for 1900 to 1969, the proxy linearly blends grids from the EDGAR v4.2 RCO 1970 grid and the gridded population, with the gridded population used for years before 1900. The proxy for the waste burning sector is based on rural population. More specific methodological details for certain aspects of the emissions dataset are outlined below.

The EDGAR and GPW proxy datasets that are used for most of the proxy data are initially processed at the highest resolution available (e.g., 0.1° for the EDGAR emissions data) and split into countries, also including a split into land and ocean areas, at this resolution before being aggregated to the final 0.5° resolution used for the data products. Because we allow multiple countries to exist within one 0.5° grid cell, this results in a more accurate distribution of emissions for each country, even when data are being processed

at the lower 0.5° resolution. For the few datasets with a resolution lower than 0.5°, data are subsampled to a resolution of 0.5° using an appropriate template.

## 2.3 Open burning emissions

The development of historical open burning emissions is described in van Marle et al. (2017). In brief, the spatial emissions distribution for these emissions is fundamentally grid-based, with emissions over the satellite era estimated from remotely sensed data (GFED4s; 1997–2015) merged with several existing historical proxies for fires, including charcoal records, for boreal and temperate North America and Europe as well as visibility-based fire emissions for the tropical areas of equatorial Asia and the arc of deforestation. The spatial distribution for the pre-satellite era was based on the 1997–2015 average and uniformly adjusted for large geographic regions based on the proxies (see van Marle et al., 2017, for the basis regions used). For the regions where proxies had limited coverage the output of six different fire models from the Fire Model Intercomparison Project (FireMIP) was used (Rabin et al., 2017) (https://www.imk-ifu.kit.edu/firemip.php, last access: 4 February 2020). Data are reported for 1750–2015. Before 1997 the annual regional fire emissions were distributed over a 0.25 × 0.25° grid based on the GFED4s climatology for six classes (savanna and grassland fires, deforestation fires, boreal forest fires, temperate forest fires, peat fires, and agricultural waste burning).

Future open burning emissions originate from the integrated assessment models (IAMs), which report by model region and broad category (e.g., forest or grassland burning). While many of the models have some additional level of spatial detail, emissions were reported at the model region level to facilitate common data harmonization, downscaling, and gridding routines. Future open burning emissions were therefore mapped to spatial grids using the same methodologies as used for anthropogenic emissions, as further described below. This means that, unlike historical open burning emissions, the spatial distribution of open burning emissions within a category (e.g., forest burning) within any country does not vary in the future scenarios.

## 2.4 Future emissions

Scenario data for future emissions from IAMs are first harmonized to a common 2015 base-year value by native model region and sector. This harmonization process adjusts the native model data to match the 2015 starting year values with a smooth transition forward in time, generally converging to native model results (Gidden et al., 2018a). The production of the harmonized future emissions data is described in Gidden et al. (2019).

The 2015 base-year anthropogenic emissions data by country and sector are extensions of the 2014 historical data largely using emission factor trends for combustion sources

from the GAINS model (ECLIPSE V5a; Stohl et al., 2015; Klimont et al., 2017) and BP fuel consumption statistics (BP, 2016). Noncombustion sources were generally scaled by estimated population. There are potentially large changes in emissions over this period, for example in China (Zheng et al., 2018), which results in uncertainty in these estimates for regions with rapid changes in air pollution control technology deployment. Near-term $SO_2$ emission trajectories in China were adjusted to better match the estimates of Zheng et al. (2018) during the harmonization process.

For open burning, the 2015 base-year emissions data for harmonization and downscaling are an average of the previous 10 years of historical data extracted by country. This is because interannual variability is not captured in the future projections, so a longer-term average is a more appropriate starting point for open burning emissions. While a decadal averaging period was used here, a longer averaging period could also be considered in future work.

The global integrated models (IAMs) that generated the Shared Socioeconomic Pathway (SSP) emission projections each have different numbers of socioeconomic regions (11–32) for which the unharmonized emission data are available. Harmonization to the common 2015 starting dataset therefore occurs at each individual model's native spatial definition in order to preserve as much detail from the model as possible (Gidden et al., 2019). Harmonized emissions for the native model regions are then downscaled to the country level as described in Gidden et al. (2019). The country-level downscaling is performed in order to provide a uniform basis for subsequent mapping to spatial grids but does not necessarily represent specific policies that might be in place for any particular country.

After the downscaling procedure the country-level future emissions projections are at the same level of sectoral resolution as the final gridding sectors in Table 2. These are then mapped to a spatial grid using the same underlying methodology as for the historical anthropogenic data, with a few differences in detail. Spatial proxies for each country are taken to be the 2014 gridded historical emissions discussed above for anthropogenic emissions and the average of the last 10 years (2005–2014) of historical data from van Marle et al. (2017) for open burning emissions. These spatial proxies are therefore constant into the future and do not represent any shifts in the spatial location of emissions within a country. The one exception is international shipping, which uses year-specific spatial distributions from the ECLIPSE project for the years 2015, 2020, 2030, 2040, and 2050. These distributions capture projected changes in shipping fuel sulfur content and the imposition of low sulfur control near coastal areas.

Future open burning emissions are provided from the IAMs in the following categories: agricultural waste burning on fields (AWB), forest burning, grassland burning, and peatland burning. Future trends for these emissions are generated by each model driven by, in large part, changes in land use. Note that these are future projections of climatologically average emission rates over time and do not include interannual variability. Because the future model projections used here do not include data on peatland burning, peatland burning emissions are held constant into the future.

Future emissions were gridded at the aggregate sectoral level, which corresponds to the final gridding sectors in the historical emissions data for anthropogenic emissions (Table 2), and with forest burning combined into one category, aggregating the open burning historical categories of boreal forest fires, temperate forest fires, and tropical deforestation and degradation. This lower level of sectoral detail was used because the models that generated the future scenario data used in this process (Gidden et al., 2019) often lack the finer level of sector detail that was available in the historical emissions datasets. The future data were constructed so that the data were consistent at the grid cell level when moving from the historical (up to 2014) to the future (2015 and forward) dataset.

The second major difference is that future emissions are not produced annually but are provided for 2015, 2020, and at decadal intervals thereafter. This is because long-term models do not provide annual data. The choice was made to only distribute data for years that were originally produced by the models and not interpolated data. This also means that fewer, and smaller, data files need to be downloaded and processed by end users. We note that the format of the open burning emissions in the future data is slightly different than the format used in the historical data. This is because we use the same software for both open burning and anthropogenic future emissions, so a similar data format was used for all future emissions data.

### 2.4.1 Future net negative fossil $CO_2$ emissions

In the default historical and future emissions gridding, emissions from electric power generation and some other large industrial facilitates are generally mapped to a spatial grid as large point sources. For most emission species this is an appropriate representation of emissions currently and also into the future. For future carbon dioxide emissions, however, an inconsistency occurs for net negative $CO_2$ emissions if this method is applied. Net negative $CO_2$ emissions occur in models when biofuel feedstocks such as crop residues, wood, or dedicated biomass crops are used to produce energy (e.g., electricity or liquid biofuels) and the $CO_2$ that is produced as part of either transformation or combustion processes is then sequestered in geologic reservoirs. The models report these net negative emissions at the point of conversion or combustion, that is, at the power plant or biorefinery. Physically, however, carbon dioxide is taken out of the atmosphere at the point at which the biofuel is produced, generally in agricultural lands. The $CO_2$ uptake in scenarios with net negative $CO_2$ emissions would therefore actually occur over a wide area and not at a point source. Representing net neg-

ative $CO_2$ emissions as a point uptake is not only physically incorrect but could cause Earth system model codes to error by causing very low or negative $CO_2$ concentration values.

We therefore distribute net negative $CO_2$ emissions over a global spatial grid. Our procedure is to first downscale harmonized model emissions to the country level (Gidden et al., 2019) and then sum all net negative $CO_2$ emissions. Any remaining positive $CO_2$ emissions are gridded normally by country. The global sum of net negative $CO_2$ emissions is then distributed spatially using a global map of potential biofuel productivity. This map was derived by taking the potential production from the LPJ-ml land model as included in the IMAGE model (Müller et al., 2016). The IMAGE-LPJ model generates potential production maps (in dry matter tons) for several crops taking into consideration the climatic and soil conditions for that specific grid cell. We have used the maximum yield of woody and grassy bioenergy crops for each $0.5 \times 0.5$ grid cell, resulting a global map of potential bioenergy production. Only those cells allocated as 2010 agricultural area (including bioenergy production) have been used. Global total net negative $CO_2$ emissions are then spatially allocated at each time step following this bioenergy production map. The spatial allocation distribution is therefore constant in time, although the absolute magnitude of total net negative $CO_2$ emissions varies with time and scenario (Gidden et al., 2019).

A global allocation was chosen because biomass trade information was not available from the models. While a more detailed time-changing map could be produced, carbon cycle modelers indicated that a global distribution would be sufficient for use in future scenarios. Note that net negative $CO_2$ emissions are allocated to a new gridded emissions sector, which means that the future $CO_2$ emission files have one more element in the sector dimension than the historical $CO_2$ emission files.

Note that this procedure is only applied for net negative $CO_2$ emissions. The combustion of biomass-based fuels without $CO_2$ capture and sequestration is not included in the $CO_2$ emissions data. Any net change in $CO_2$ emissions due to the use of biofuels is accounted for in either net land-use change emissions (Gidden et al., 2019) or as part of changing land use and land cover (Hurtt et al., 2019).

## 2.5 Seasonality

Monthly seasonality is applied to the gridded emissions by applying a set of spatially explicit seasonality fractions. The primary source for the seasonality fractions was the ECLIPSE dataset, with the addition of EDGAR v4.3 for international shipping and Lamarque et al. (2010) for aircraft. The dataset was processed to a consistent seasonality that matched the monthly calendar used in the final emissions data (e.g., a 365 d year). As part of this process some minor inconsistencies in the anthropogenic seasonality data between sectors in terms of the number of days assumed in

the year were corrected so that the month distribution was consistent with the 365 d year used in the anthropogenic dataset. Seasonality for future open burning emissions was taken from a 10-year average from the historical dataset.

We note that, after the distribution of this dataset, it was found that the industrial sector in some regions has a high level of seasonality in the ECLIPSE dataset, and this distribution was carried through to the CMIP6 data (as noted on the online GitHub issues list for this dataset; see below). In general, we would expect industrial sector emissions to be fairly constant as most industrial activities operate year-round. The magnitude of seasonality overall in some regions may therefore be overestimated. This will be reexamined in future data releases.

## 2.6 VOC speciation

Modeling atmospheric chemistry requires emissions for specific species, or species groups, of reactive compounds instead of the total mass of volatile hydrocarbon emissions (NMVOCs) generally tabulated in inventories. Anthropogenic VOC emissions were speciated by applying speciation profiles by gridding sector and country, largely derived from the RETRO project. Emissions for 23 anthropogenic VOC groups were extracted by country and broad sector from gridded HTAP v2 emissions data (http://iek8wikis.iek.fz-juelich.de/HTAPWiki/WP1.1, last access: 13 January 2016). These were normalized, converted to percentages, and applied to the emissions data at the country and final gridding sector level. In the future, it would be useful to use country- and sector-specific profiles, such as those in Huang et al. (2017). The speciated emissions were gridded in the same manner as bulk VOC emissions, thereby providing emission grids by species.

While speciation profiles by gridding sector are held constant in time, the aggregate VOC speciation for a country will change over time. This is because of different speciation profiles for each sector combined with a changing sectoral contribution to total VOC emissions over time.

Figure 2 shows the ratio of butanes to hexanes+ in the RETRO dataset used in CEDS and the newer estimate by Huang et al. (2017), which we present as an illustration of differences in the speciation profiles used in these two datasets. While there is considerable overlap between the two datasets, the RETRO data have a lower average ratio for these two species in the road transportation sector compared to EDGAR, while the opposite is the case for industrial combustion. We note that variation across, and even within, countries is expected. For example, Chin and Batterman (2012) report that the fraction of $n$-heptane (a higher alkane) in gasoline vapors varies between 6 % and 19 %, depending on location and season within the United States.

For both historical and future emissions, however, VOC speciation profiles can potentially change with technology deployment separately from changes in NMVOC emission
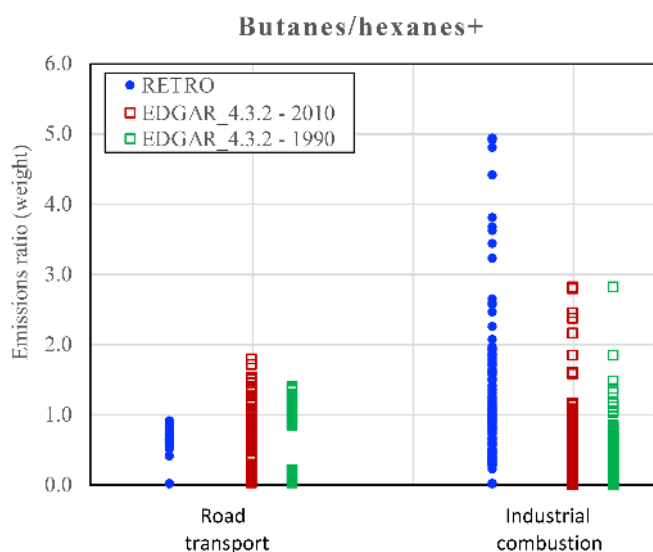
**Figure 2.** Range across countries of the butane / hexanes (and higher alkanes) in the RETRO speciation used in CEDS and the results from Huang et al. (2017) for 2010 and 1990. Each symbol represents one country. A lower ratio indicates more of the higher-weight alkanes in the assumed speciation profile. The difference between the 1990 and 2010 EDGAR data is due to a changing mix of fuels used for those sectors between those two time periods, for example shifts between different amounts of gasoline vs. diesel fuel consumption.

rates. For example, regulations to limit ozone formation not only result in altered bulk NMVOC emission amounts, but also change the specific VOC species emitted (Kirchstetter et al., 1999). Other examples include changes over time in the composition of consumer products (McDonald et al., 2018) and changing formulation of paints and other coating materials.

It is not known how much the differences illustrated in Fig. 2 reflect updated or different data sources or actual changes in VOC speciation over time. The RETRO dataset is targeted at 1990–1995 and uses speciation profiles available in time period, while the EDGAR data take in newer measurements. While the emission time series presented here, and also the EDGAR data, capture the broad changes in VOC speciation due to changing sectoral contributions to VOC emissions over time, these datasets do not capture underlying changes in speciation profiles due to changing regulations or other fundamental changes to speciation profiles over time. It would therefore be useful to determine the importance of such speciation changes for the historical modeling of atmospheric chemistry compared to the broader changes in VOC emission magnitudes and speciation changes due to sectoral shifts over time. Note that the Huang et al. (2017) data do capture changes in speciation over time due to changing mixes of fuels in each end-use sector, which is illustrated in Fig. 2.

For future open burning emissions, speciation profiles by country and open burning sector were extracted from the historical open burning emissions data files and applied to the future scenario bulk NMVOC emissions. For details on the historical open burning VOC speciation see van Marle et al. (2017). Note that the species provided are different for the anthropogenic emissions and open burning emissions, following the past conventions used by each of those communities.

## 3 Results

In this section we provide a brief overview of the gridded data products, focusing largely on diagnostic graphics that illustrate the underlying structure and potential issues that impact the use and interpretation of the gridded data.

### 3.1 Gridded products

Table 4 provides a summary of the gridded data files described in this paper with the emission species listed in Table 1. The files listed there, plus historical gridded open burning emissions (van Marle et al., 2017), provide the complete set of emissions data required for CMIP6 experiments (Eyring et al., 2016). As discussed in Hoesly et al. (2018) the anthropogenic emissions are consistently generated from the same driver data across sectors and emission species. The gridded data files are generated using consistent spatial proxy data, seasonality, and sector definitions. This means that some emission species, such as $CO_2$, which have been provided as only bulk emissions in CMIP5, are provided here with greater sectoral detail. For applications that do not require this level of sectoral detail, emissions can be summed across sectors. Note that, for all species, total anthropogenic emissions consist of the sum of all eight (historical) or nine (future)[1] surface emission sectors plus aircraft emissions.

Supplementary data files include speciated NMVOC emissions for all time periods. Also provided are historical emissions from solid biomass fuels (wood, agricultural residues, etc.) as these are used as supplementary data in some models.[2] The solid biofuel emissions are a subset of the total emissions in the main data files and are provided for models that need this information. Solid biofuel emissions were not reported separately for the future scenarios so these are not provided in the gridded data. If these emissions are needed for the future scenarios, we recommend that the ratio of solid biofuel to total emissions by sector or grid cell from the last few historical years times the future emissions data for each species be used to derive future grids for solid biofuel emis-

---

[1]As discussed above, future emissions contain an additional sector to represent net negative $CO_2$ emissions.

[2]Some models use information on solid biofuel combustion to derive information on associated VOC emissions and/or primary aerosol (e.g., BC, OC) characteristics.

**Table 3.** Total anthropogenic and open burning global emissions for selected time periods. Values for 2050 and 2100 are from the SSP2-45 scenario. Only anthropogenic $CO_2$ emissions are shown since net open burning $CO_2$ emissions depend on land-use change history and are not provided in this dataset.

| Species | Source | 1850 | 1980 | 2015 | 2050 | 2100 | Units |
|---|---|---|---|---|---|---|---|
| BC | Anthro | 934 | 4975 | 7968 | 4769 | 1415 | |
| | Open burn | 1644 | 1646 | 1748 | 1548 | 1211 | ktC |
| $CH_4$ | Anthro | 31 306 | 271 364 | 373 737 | 342 274 | 280 322 | |
| | Open burn | 11 919 | 12 086 | 14 336 | 14 893 | 14 830 | $ktCH_4$ |
| CO | Anthro | 118 527 | 593 946 | 606 222 | 445 376 | 162 521 | |
| | Open burn | 294 472 | 297 994 | 327 549 | 284 852 | 220 688 | ktCO |
| $NH_3$ | Anthro | 6028 | 42 442 | 61 383 | 71 465 | 60 414 | |
| | Open burn | 3599 | 3560 | 3894 | 4564 | 4953 | $ktNH_3$ |
| NMVOC | Anthro | 10 782 | 137 690 | 163 887 | 143 351 | 80 164 | |
| | Open burn | 56 315 | 57 025 | 63 270 | 53 646 | 40 272 | ktNMVOC |
| $NO_x$ | Anthro | 558 | 99 289 | 138 653 | 98 457 | 56 610 | |
| | Open burn | 19 838 | 20 353 | 21 060 | 18 146 | 13 493 | $ktNO_2$ |
| OC | Anthro | 4262 | 11 313 | 19 538 | 13 441 | 3823 | |
| | Open burn | 14 045 | 14 008 | 15 202 | 13 417 | 10 702 | ktC |
| $SO_2$ | Anthro | 2481 | 135 141 | 98 326 | 50 649 | 28 077 | |
| | Open burn | 2073 | 2090 | 2162 | 2170 | 2031 | $ktSO_2$ |
| $CO_2$ | Anthro | 188 | 20 331 | 11 464 | 22 585 | 31 396 | $MtCO_2$ |

sions. More detailed emissions data by sector and fuel from individual models could be used to assess the adequacy of this assumption.

Spatial distributions for anthropogenic reactive gases and carbonaceous aerosol emissions (e.g., BC and OC) for selected historical and future years are shown in Figs. 3–6, with numerical values in Table 3. An overview of large-scale trends is provided here, with the emission trajectories and their derivation described in more detail in the underlying literature for the historical emissions (van Marle et al., 2017; Hoesly et al., 2018), future scenarios (Calvin et al., 2017; Fricko et al., 2017; Fujimori et al., 2017; Kriegler et al., 2017; Riahi et al., 2017; van Vuuren et al., 2017), and the harmonized CMIP6 data (Gidden et al., 2019).

Global emissions of species that are the product of incomplete combustion (BC, CO, $NO_x$, OC, NMVOC) are dominated by open burning in 1850 (Table 3, Figs. 3–4). Anthropogenic combustion emissions in 1850 are dominated by solid biofuel use for cooking and heating. However, by the mid-20th century, anthropogenic processes dominate for all species except for OC (Table 3). As industrialization proceeds, emissions become large in the regions that experienced early industrialization such as Europe, North America, Japan, and Australia by the middle to late 20th century (e.g., Fig. 5). By the early 21st century (e.g., 2015; Fig. 6) emissions controls have decreased pollutant emissions somewhat in Europe and North America but increased dramatically compared to mid-20th century levels in emerging economy regions such as China, India, and the Middle East. Anthropogenic $CO_2$ emissions, in general, do not follow the same trajectory as air pollutant emissions.

Future emissions diverge depending on the emission scenario and follow trends as described in Gidden et al. (2019). A few illustrative graphics are provided here. By 2050 in SSP2, the "middle of the road" scenario, emissions have decreased from 2015 levels in most world regions (Fig. 7), while they have decreased to quite low levels in the SSP1-26 scenario (Fig. S7 ). In contrast, emissions do not decrease nearly as much in the SSP3-70 scenario (Fig. S9). By 2100, under the SSP2-45 scenario, anthropogenic emissions are generally larger than open burning emissions, but anthropogenic + open burning emissions are generally smaller than 2015 levels. Future open burning emissions also vary by scenario, but not as dramatically as anthropogenic emissions (Figs. S3–S6 in the Supplement).

## 3.2 Gridded data diagnostics and interpretation

While 0.5° gridded data are the primary product of gridding, maps at that resolution can be of limited use for understanding spatial and temporal patterns in the data or diagnosing potential issues. In this section we discuss a number of diagnostic graphs that illustrate various dimensions of the dataset. We first discuss the temporal structure of the anthropogenic emissions data over different sectoral and spatial scales.
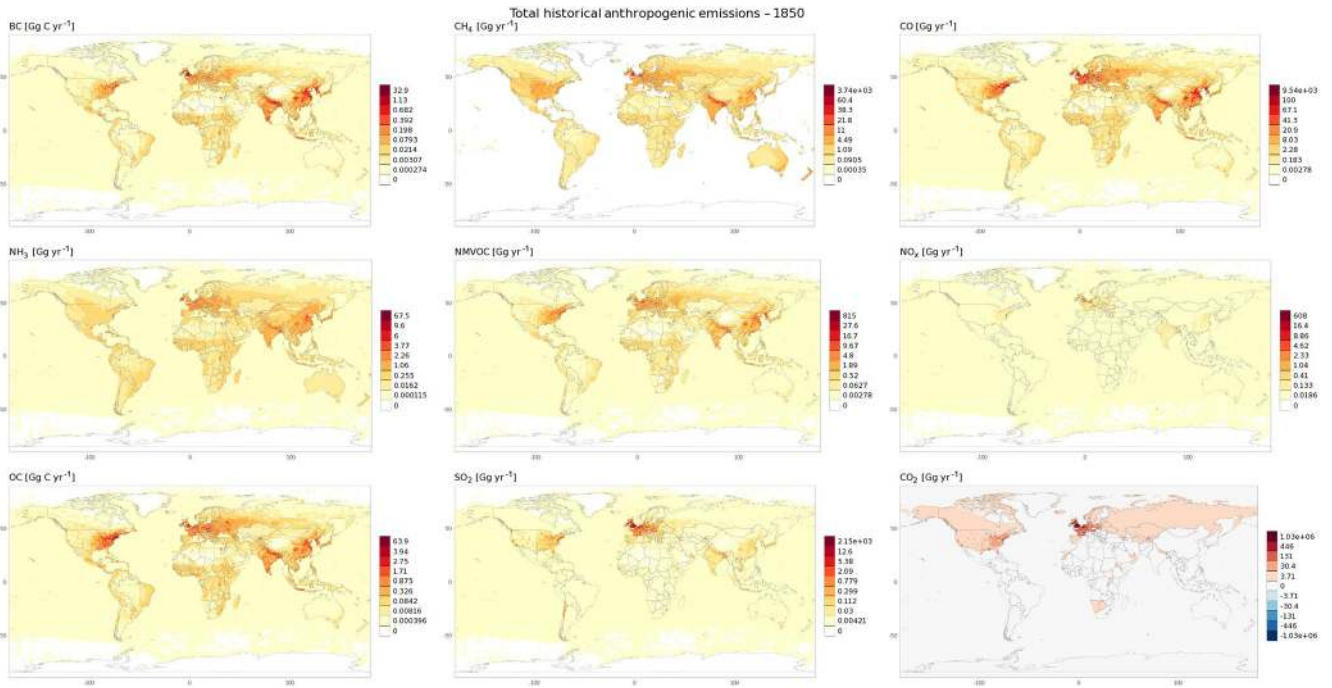
**Figure 3.** Total anthropogenic emissions in 1850 by species. The color scale is the same for each emission species across all gridded figures for comparison. Note that the $CO_2$ color scale accommodates the net negative emissions found in some future scenarios.
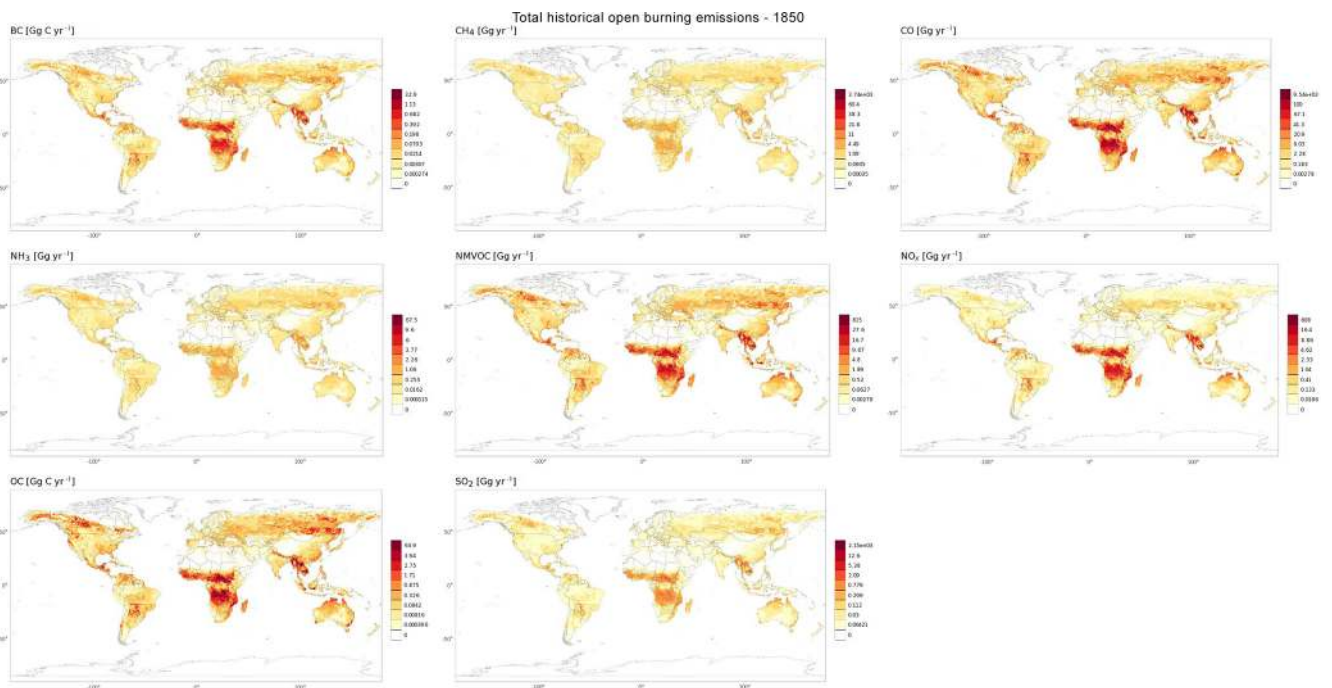


**Figure 4.** Total open burning emissions in 1850 by species. The color scale is the same for each emission species across all gridded figures for comparison.
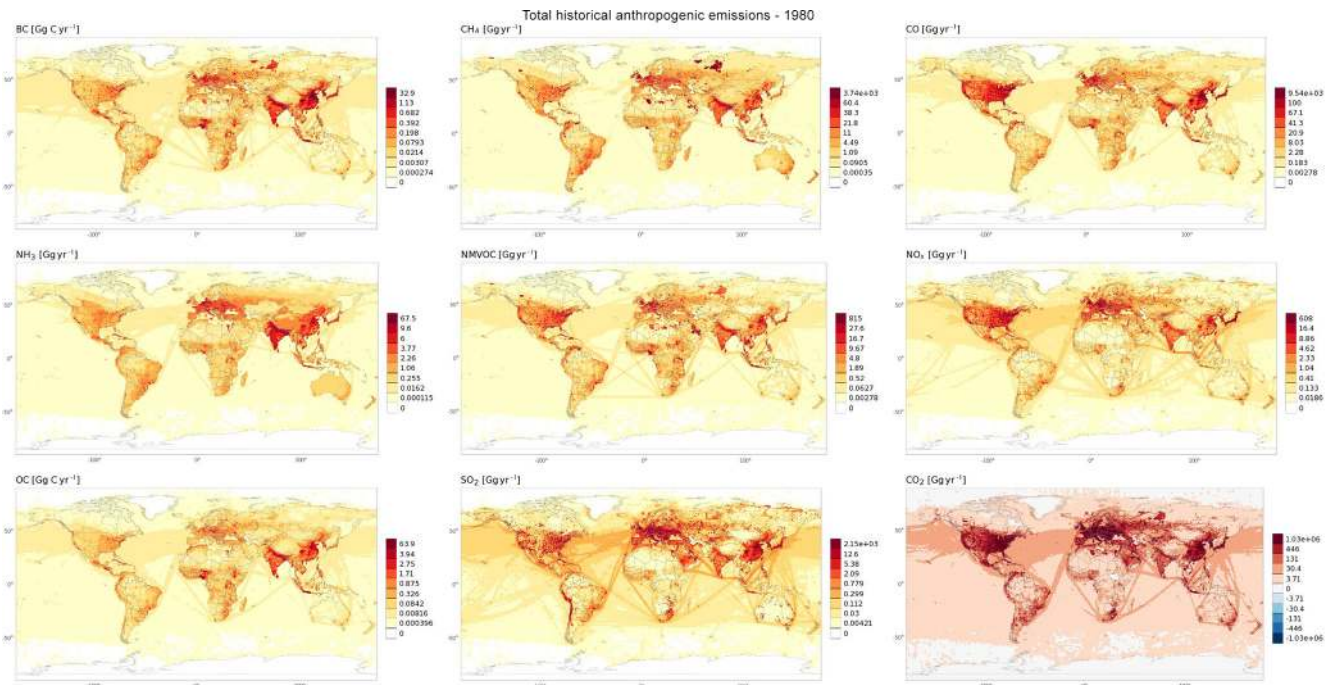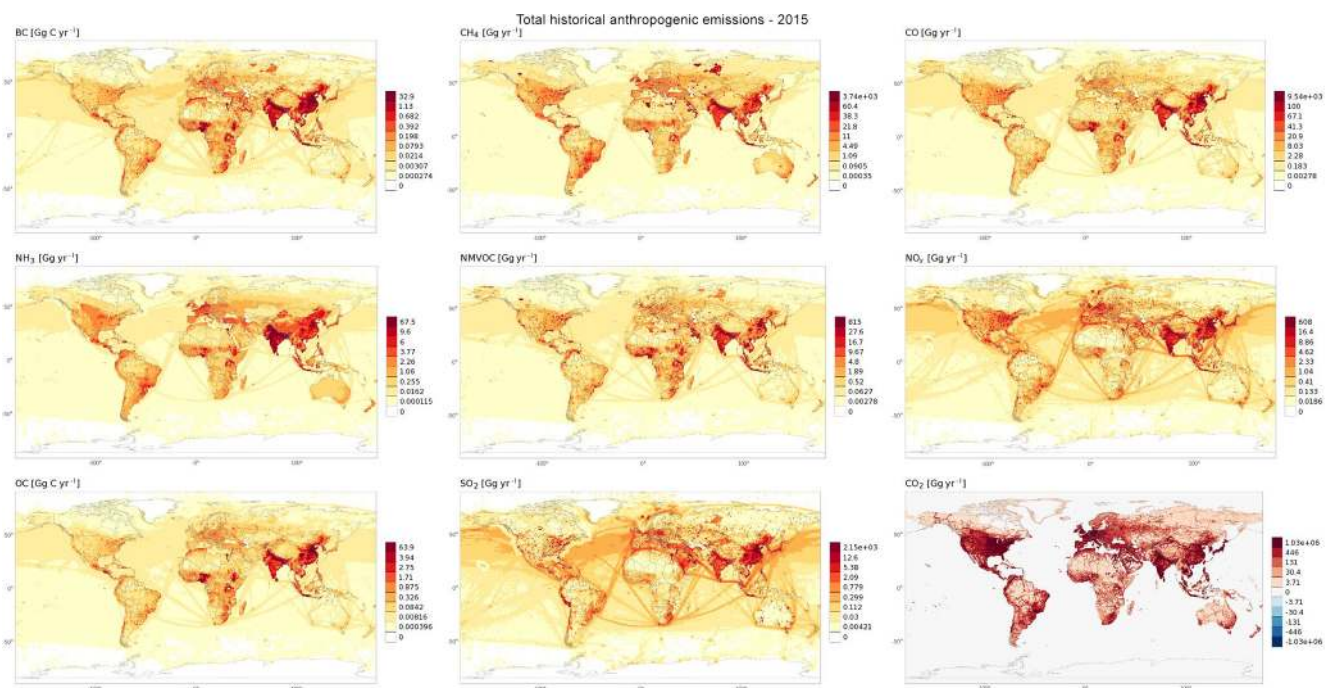
**Figure 5.** Total anthropogenic emissions in 1980 by species. The color scale is the same for each emission species across all gridded figures for comparison. Note that the $CO_2$ color scale accommodates the net negative emissions found in some future scenarios.



**Figure 6.** Total anthropogenic emissions in 2015 by species. The color scale is the same for each emission species across all gridded figures for comparison. The $CO_2$ color scale accommodates the net negative emissions found in some future scenarios.

**Table 4.** Gridded historical anthropogenic and future scenario emission data files provided for CMIP6. The complete historical anthropogenic dataset consists of 327 files. The complete future scenarios dataset consists of 1017 files. Historical emissions are provided annually, while future emissions are provided for 2015, 2020, and in decadal intervals after that. All emissions are provided with monthly seasonality. Full data citations, including digital object identifiers (DOIs), are provided in the "Data availability" section at the end of this paper.

| Data description | Year range | Example file name on ESGF |
|---|---|---|
| Anthropogenic emissions – surface | Historical (1750–2014) | CO-em-anthro_input4MIPs_emissions_CMIP_CEDS-2017-05-18_gn_185101-189912.nc |
| Anthropogenic emissions – aviation | Historical (1750–2014) | CO-em-AIR-anthro_input4MIPs_emissions_CMIP_CEDS-2017-10-05_gn_185001-185012.nc |
| Speciated anthropogenic emissions | Historical (1750–2014) | VOC01-alcohols-em-speciated-VOC-anthro_input4MIPs_emissions_CMIP_CEDS-2017-05-18-supplemental-data_gn_180001-184912.nc |
| Solid biofuel emissions | Historical (1750–2014) | OC-em-SOLID-BIOFUEL-anthro_input4MIPs_emissions_CMIP_CEDS-2017-05-18-supplemental-data_gn_195001-199912.nc |
| Extended anthropogenic $CH_4$ emissions | Historical (1750–2014) | CH4-em-anthro_input4MIPs_emissions_CMIP_CEDS-2017-05-18-supplemental-data_gn_185001-196012.nc |
| Anthropogenic emissions – surface | Future (2015–2100) | CO-em-anthro_input4MIPs_emissions_ScenarioMIP_IAMC-MESSAGE-GLOBIOM-ssp245-1-1_gn_201501-210012.nc |
| Anthropogenic emissions – aviation | Future (2015–2100) | CO2-em-AIR-anthro_input4MIPs_emissions_AerChemMIP_IAMC-AIM-ssp370-lowNTCF-1-1_gn_201501-210012.nc |
| Speciated anthropogenic emissions | Future (2015–2100) | VOC07-ethene-em-speciated-VOC-anthro_input4MIPs_emissions_ScenarioMIP_IAMC-IMAGE-ssp126-1-1-supplemental-data_gn_201501-210012.nc |
| Open burning emissions | Future (2015–2100) | NO$_x$-em-anthro_input4MIPs_emissions_ScenarioMIP_IAMC-IMAGE-ssp119-1-1_gn_201501-210012.nc |
| Open burning sector shares | Future (2015–2100) | CO-openburning-share_input4MIPs_emissions_ScenarioMIP_IAMC-IMAGE- ssp119-1-1_gn_201501-210012.nc |
| Speciated open burning emissions | Future (2015–2100) | NMVOC-C6H6-em-speciated-VOC-openburning_input4MIPs_missions_ScenarioMIP_IAMC-REMIND-MAGPIE-ssp534-over-1-1-supplemental-data_gn_201501-210012.nc |
| Speciated open burning sector shares | Future (2015–2100) | NMVOC-CH2O-speciated-VOC-openburning-share_input4MIPs_emissions_ScenarioMIP_IAMC-IMAGE-ssp126-1-1-supplemental-data_gn_201501-210012.nc |

Historical emission files are generally provided in 50-year segments in order to keep the size of individual files reasonable (historical $CH_4$ emissions are provided over 1970–2014). Supplementary emissions files provide an approximate extension back to 1850. Data for the CMIP6 starting year 1850 are provided in one file. Future emissions were provided for the entire projection period (2015–2100) in one file per species and/or emissions type enumerated in this table. All data except aviation are provided as two-dimensional grids, generally with an additional sector dimension. Aviation emissions are provided as three-dimensional data. All data available at https://esgf-node.llnl.gov/projects/input4mips/ (last access: 4 February 2020).

Figure 8a shows a time series plot of $SO_2$ emissions at one grid cell for the industrial (IND) gridding sector (e.g., combustion + process emissions; see Table 2). These emissions are the result of a combination of changes in the proxy data over time, in this case from EDGAR emissions, multiplied by the estimated time series for the United States for the underlying detailed sectors. Emissions from this same cell for the energy sector (Fig. 8b) show a sharp increase due to (presumably) a power plant coming into service, as represented in the proxy dataset.

Total anthropogenic $SO_2$ emissions for the same grid cell (Fig. 8c) show a different temporal structure as a result of combing the differing temporal trends from different sectors. Emissions become smoother if we examine total $SO_2$ emissions summed over 16 nearby grid cells (Fig. 8d). However, emissions at either the grid cell or for a sum of several grid cells will generally show more structure over time compared to the corresponding emissions for the entire USA region (Fig. 9) due to the additional temporal structure in the proxy dataset at smaller spatial resolutions.

For any particular spatial location, the accuracy of the gridded emissions depends on the accuracy of the underlying proxy dataset as well as underlying the country-level emissions data. Global proxy datasets that are available for
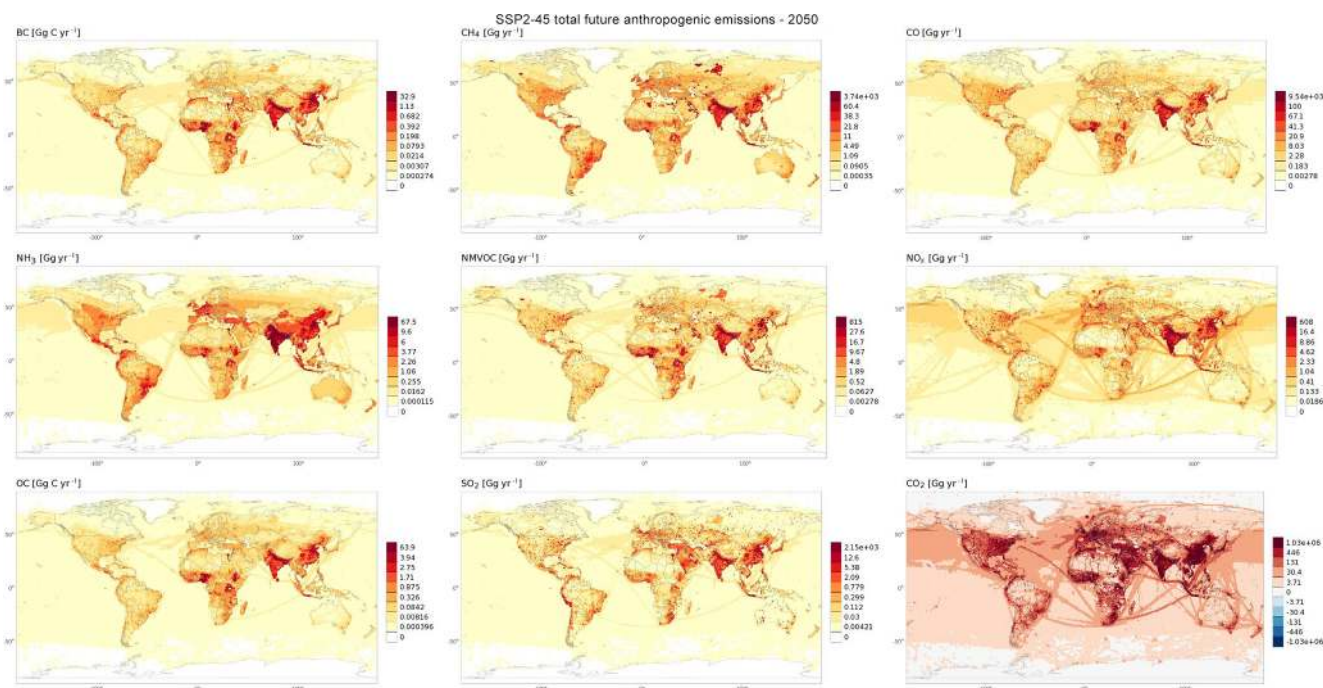
**Figure 7.** Total anthropogenic emissions in 2050 by species in the SSP2-45 scenario. The color scale is the same for each emission species across all gridded figures for comparison. The $CO_2$ color scale accommodates the net negative emissions found in some future scenarios.

use in projects such as this, including projects of similar scope such as EDGAR (from which many of the spatial distributions used here are drawn), will not necessarily correspond exactly to the actual spatial location and magnitude of emission sources. For example, databases of global road networks, while they are improving (e.g., Crippa et al., 2018), will not exactly correspond to either actual road traffic at every point, nor are they likely to be complete for all regions. While large emission sources, such as power plants, are more likely to be in global databases, such data can still be incomplete, particularly for smaller plants (Liu et al., 2015). For other sectors, it is difficult to capture in proxy datasets historical changes over time in emission strengths in specific spatial locations.

At the grid cell level, discontinuities over time can therefore potentially be due to (a) discontinuities in the country-level emission time series, (b) actual changes in the underlying emitting processes as represented in the proxy dataset, or (c) data breaks or inconsistencies in the proxy dataset that do not represent an actual historical change in emissions. As either the sectoral aggregation or the spatial scale increases, however, the data will tend to become more robust, at least to the extent that the country-level emission time series used for calibration in this project are accurate (see Hoesly et al., 2018, for a further discussion).

Monthly time series plots visualizing the seasonal cycle of the emissions data, as shown in Fig. 10, are also useful

diagnostics. Note, again, that discontinuities at the single cell level are due to temporal structure in the proxy datasets.

The temporal signature of the open burning data is quite different, as interannual variations in those data are dominated in large part by year-to-year changes in local meteorological conditions. At least for the modern era, the interannual variations are inferred from satellite observations. As discussed below, the increased use of remote sensing and regional bottom-up data has the potential to improve the spatial and temporal accuracy of anthropogenic emissions data as well.

Spatial data presented at a lower resolution can also be useful in understanding overall patterns and spatial differences between datasets. As discussed in Hoesly et al. (2018), we have found that spatial maps at a 10° resolution are useful in showing large-scale differences in emissions magnitude and spatial distribution between two datasets. For anthropogenic $SO_2$ emissions in 1900, for example (e.g., Fig. 11), the CMIP6 dataset shows some shifts in emissions from northwest to eastern Europe and a different distribution of emissions over the oceans compared to the CMIP5 dataset.

Also note that the CMIP6 dataset has a different distribution of $SO_2$ emissions over North America than in the CMIP5 data, with more $SO_2$ emissions over the western areas and lower emissions over the east. This has been recognized (Yang et al., 2018) as a bias in the proxy dataset over the United States used here (e.g., EDGAR $SO_2$ emissions). As discussed above, national-level anthropogenic emissions
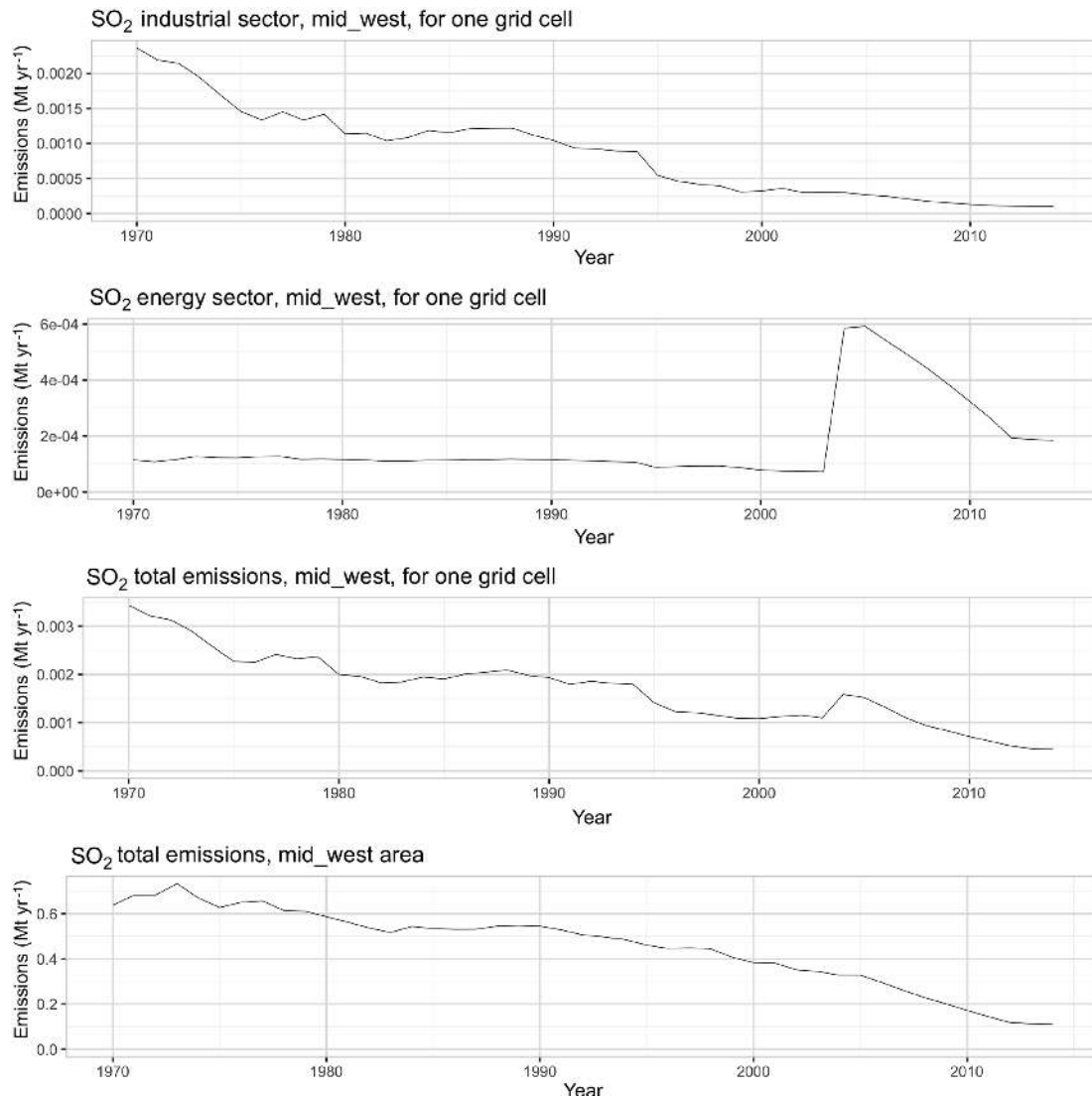
**Figure 8. (a)** Time series plot for the industrial gridding sector for a single 0.5° cell in the US midwest (lat, long: 41.25, −87.75). **(b)** Time series plot for the energy sector for the same grid cell. **(c)** Time series plot for total anthropogenic emissions for the same grid cell. **(d)** Time series plot for total anthropogenic emissions summed over 16 grid cells including the cell depicted in **(a)**–**(c)**.

were distributed within each country using proxy data. In this case, the proxy data likely do not account for the lower sulfur content of coal in the western United States. As discussed below, one method of accounting for such issues is to estimate emission at smaller geographic units, such as US states, and then map to spatial grids. For further comparisons between the CMIP5 and CMIP6 anthropogenic emissions data, see Hoesly et al. (2018).

## 4 Discussion and potential future work

The spatially distributed emissions data discussed here represent a number of improvements over previous century-scale gridded datasets. These improvements include anthropogenic

emissions that were consistently generated from the same driver data and mapped to spatial grids using the same proxy data across emission species with consistent seasonality. Future emissions were downscaled to the country level and then consistently mapped to spatial grids using largely the same gridding methodologies as used for the historical anthropogenic emissions. Further, the software used for historical anthropogenic emissions and future scenario emissions has been made publicly available as open-source software.

Emission estimates remain uncertain owing to data gaps, biases and errors in activity data, biases in emission factors, and spatial and temporal variations in real-world conditions. Relative uncertainties will generally increase going back in time as less detailed sectoral, spatial, and temporal informa-
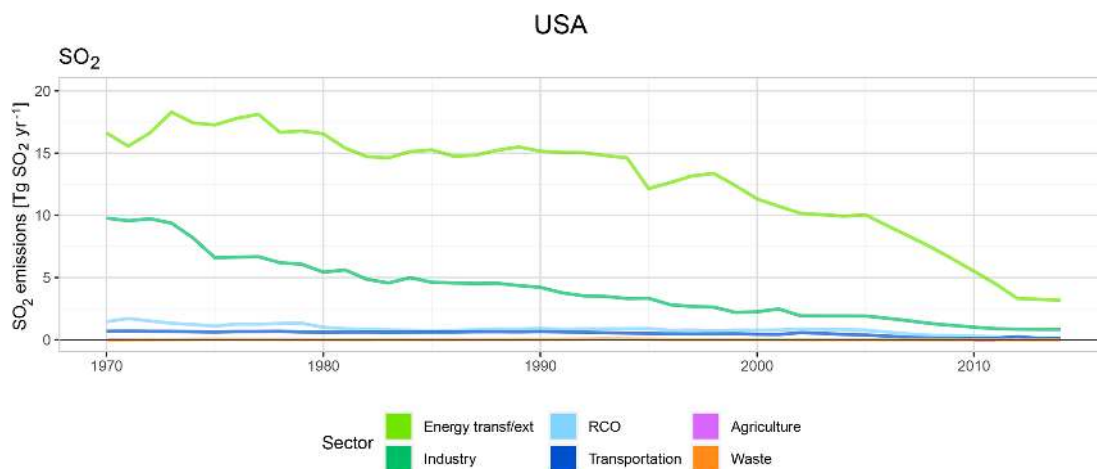
**Figure 9.** Time series plot for $SO_2$ by sector for the USA as a whole.
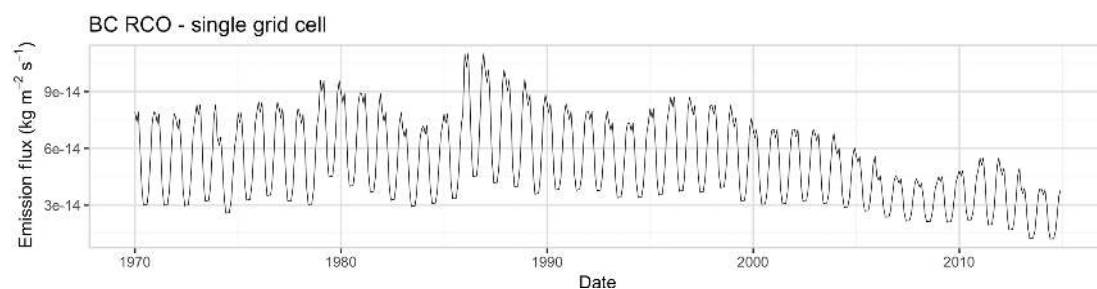


**Figure 10.** Time series plot for total BC emissions over time at the monthly level for a grid cell in Finland.
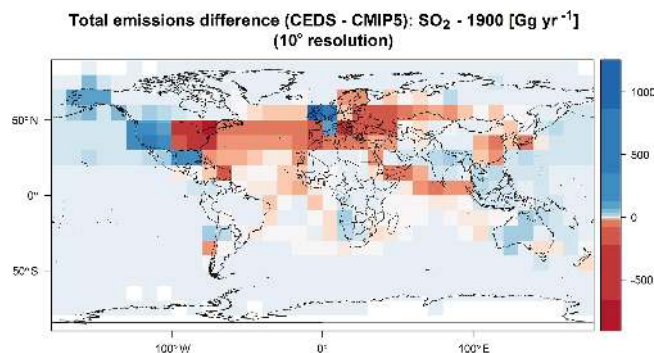


**Figure 11.** Difference in CMIP6 (e.g., CEDS) and CMIP5 gridded total $SO_2$ emissions in 1900. For a sector-to-sector comparison, these figures do not include aviation or agricultural waste burning on fields.

tion is available. Future emissions projections are inherently uncertain, which is why multiple scenarios are presented. Within any one scenario, no attempt was made in these data products to estimate how the spatial distributions of emissions within countries might change into the future. What is not clear, however, is the practical value of attempting such detail given the overall uncertainty in future trajectories.

The historical open burning emissions have specific sources of uncertainties, as discussed further in van Marle et al. (2017). These start with uncertainties of the different products used in the construction of the overall dataset: GFED4s, emissions stemming from visibility observations, charcoal-based fire records, and modeled emissions. The choice of which fire models are used before the satellite-observed fires (1997–2015) can influence the results, as will the land-use change data and other assumptions used to drive those models. The use of GFED4s climatology to distribute emissions spatially and temporally (e.g., seasonally) from regional estimates before 1997 assumes that subregional distributions stayed constant before 1997.

While any errors in aggregate emission estimates will flow down to errors in spatial distribution, here we are particularly interested in potential errors in allocating emissions spatially. Systematic analyses of uncertainties in spatial emission patterns, however, have not been assessed to date. The methodologies used here for historical anthropogenic and future scenario emissions rely on underlying spatial proxy data. One source of error is when the spatial proxy data do not match the actual spatial distribution of emissions. One example was discussed above in the case of $SO_2$ emissions in the United States. In this case, the global proxy dataset used to distribute these emissions spatially did not contain country-specific in-

formation on the subregional distribution of emissions. Overall, as the emissions data are considered over larger spatial areas, we expect that the accuracy is likely to improve.

Capturing the distinctions between urban and rural emissions, and the finer distinctions in between, is an ongoing challenge for emission inventories, for example road transport emissions, open refuse burning, and fuel combustion for residential heating and/or cooking. Driving conditions and emission profiles tend to differ between urban and rural conditions, resulting in different fuel consumption and emission factors. To the extent that regional inventories become better able to capture these distinctions, those methodologies could ultimately be incorporated into the methods for generating proxy data for global datasets.

While the focus here was the use of consistent spatial proxy data across species, in some cases it might be desirable to use varying proxies between pollutants; e.g., coal use in the residential sector will produce significant $SO_2$, while heating or cooking with wood will not. Some of these differences, however, may be captured in the EDGAR emissions distributions used as proxy data. Because country-level anthropogenic emissions estimates were used for historical emissions gridding, potential misallocation of emissions is most important for countries with large spatial extents.

More general spatial errors can also occur because of mismatches between proxy data and actual emission sources. Traffic volume, for example, is not likely to scale purely with the class of roadway in global road datasets. Country-level emissions from power plants, for example, are allocated spatially using proxy data such as power plant capacity and the presence of emission control devices, which is the primary information available in global proxy datasets. Actual emissions, however, will not necessarily be proportional to facility capacity, resulting in spatial errors. Further, point sources can be missing in global proxy datasets, particularly in developing countries (Liu et al., 2015). In addition, the spatial location of individual sources might not always accurate in proxy datasets.

Additional kinds of errors can occur in the temporal dimension. Even if the time series at the aggregate country level is correct, emissions at any spatial location will likely not follow the same temporal pathway. The emissions rate of localized sources is likely to change over time, and individual point sources could close down, or new sources could be put into operation. Temporal consistency over time at the grid cell level is perhaps one of the most difficult challenges for long-term gridded datasets. Improved information on emissions over time for large point sources (see below) is the most important issue here, as year-to-year changes in large sources could be a large source of temporal changes in emissions.

The distribution of emissions over the year, represented in these datasets as monthly average seasonality, is another area in which improvement is likely needed. The seasonal distribution for anthropogenic emissions is based on global sectoral analysis, for example mapped by heating degree days (HDDs), which is not likely to incorporate country-specific details on how HDDs map to actual fuel use. The rate of heating fuel consumption relative to temperature will likely be different in a cold region of a country (where heating demand is high) compared to a region with a milder climate using different heating technologies. The seasonal distribution of ammonia emissions depends on the sector and region (Paulot et al., 2014), and it is not clear if there is a general consensus on these seasonal distributions.

There are several potential strategies for reducing such biases and errors in gridded datasets. One general approach is to improve the completeness and representativeness (relative to actual emissions) of global proxy datasets to reduce the issues discussed above. One example is community efforts to improve global databases, such as those for power plants and road networks, used in emissions work.

Another method is to estimate emissions at smaller geographical units, such as states or provinces, and allocate emissions spatially at that level. Emissions information is often available at subnational levels, for example for US states (US EPA), Chinese provinces (Liu et al., 2015), and Indian states (Klimont et al., 2017). Allocating emissions information at this level would result in an improved overall distribution of emissions over large countries, even if the same global proxy datasets were used for each individual state or province. For global-scale models this would likely be sufficient accuracy, assuming the state-level emission time series were accurate. An effort is underway, for example, to implement a state-level disaggregation for the USA in the CEDS, which would fix the east–west $SO_2$ emissions distribution issue mentioned above.

Detailed spatial data for emissions have been developed over many regions for air quality modeling purposes, and it is likely that incorporating these data into global inventories could result in improved spatial data overall. One method is to simply "stitch" together detailed spatial maps where they are available (Janssens-Maenhout et al., 2015). One disadvantage of this approach, however, is that the resulting datasets are often inconsistent because detailed spatial data are not available for every year; different years in different regions need to be used. Inconsistencies also arise at country borders where grid cells overlap. Another disadvantage is that these detailed data are often time-consuming to generate, which means that these regional datasets are not available for the most recent years. An alternative to binding together emission data grids is to incorporate the underlying spatial proxy distributions into workflows such as those described here. This would presumably result in improved spatial distributions for those regions with data but also be flexible enough to allow for greater temporal consistency between regions. One challenge, however, is that these detailed datasets are often only available for recent decades. For longer-term modeling, some method of extending these proxy data back into time would be needed, perhaps by blending with simpler proxy data such as population distributions.

For large point sources in some regions direct measurements of emissions through continuous emission monitoring (CEM) systems are in place, which potentially provide highly accurate emission estimates for those specific sources. Another source of emission information for large sources is satellite remote sensing data. Fioletov et al. (2016), for example, have used satellite data to develop a catalog of emission sources "missing" in existing gridded inventories. Either of these data sources could be combined with bottom-up estimates to produce a more robust emission dataset, with the spatial location and temporal (e.g., year-to-year) variations of emissions from large sources provided by CEMs or remote sensing analysis. This was done, in a limited manner, for $SO_2$ emissions from petroleum refining in the CMIP6 dataset (https://GitHub.com/JGCRI/CEDS/wiki/Data_and_Assumptions#so-sub-2-sub-1, last access: 1 August 2019; Hoesly et al., 2018). Note, however, that there are limits to such techniques in that satellite data are only reliably estimated for the largest sources. Spatial and temporal biases in remote sensing data need to be carefully considered, for example the greater difficulty of detecting sources in winter compared to summer.

Additional data sources on specific large sources, such as literature sources and corporate reports, can also be valuable for providing temporal emission detail. While such information was used in the CEDS CMIP6 release, for example for $SO_2$ emissions from several large ore smelting complexes,[3] this was only applied at the country level, not at the level of point sources. Information on industrial sources is particularly difficult to obtain given the wide diversity of source types and often confidentiality issues preventing data availability for specific facilities. Given that large point sources often account for a large portion of emissions for some species, adding a general capability to represent specific point sources and their emission levels over time, as distinct from more numerous smaller sources, would improve the accuracy of emissions data.

While there are many aspects of emissions data that can potentially be improved, is will be important to better assess which improvements are most critical for focused efforts. For example, is a 10 % error in overall emissions level in a region more or less important than a 20 % east–west error in spatial allocation? These trade-offs will depend on the model and specific application. For example, smaller-scale details will be less important for coarser-scale models. It would be useful to more systematically test the sensitivity of models to spatial allocation. While localized results will depend on spatial distribution, it is not clear how sensitive larger-scale results, such as circulation, temperature, and precipitation changes,

are to differences in spatial allocation. One work directly examining this sensitivity (Geng et al., 2017) found that, once power plant point source and road network proxy data were used, there was a significant benefit from using industrial economic activity data to spatially allocate industrial sector emissions.

The current datasets produced for CMIP6 were produced at either 0.5° (historical anthropogenic and future) or 0.25° (open burning over 1750–2015). Given that the primary purpose of these datasets is for global models, this resolution is appropriate given that this is finer than the spatial resolution of most current global models, particularly those used to conduct century-scale simulations. Higher-resolution datasets would be useful for regional modeling or next-generation models with higher resolution. In principle, given the availability of 0.1° EDGAR emissions data used as proxy data within CEDS, emissions data up to this level of resolution are feasible to produce. It would be useful, before committing to the production of long-term higher-resolution emissions data, to examine the accuracy of the global proxy datasets at this level of spatial detail (and also the accuracy required for a specific purpose). This could be done, for example, by comparing to higher-resolution regional data.

Emission datasets would also benefit from more systematic comparisons with observations. Some observations, such as concentration ratios, can be directly compared to inventories (e.g., Hoesly et al., 2018), while, more generally, modeled concentrations need to be compared to observations. Where model biases can be ruled out, differences with observations can be used to inform improvements to inventory data. At present, however, this is performed in an ad hoc manner, with no regular systematic comparison across species and regions.

Finally, it would be useful if the types of long-term emission datasets discussed here could be more regularly updated. In the past, most of these datasets were compiled once every 5 or so years in order to produce data for the next CMIP exercise. A paradigm of continuous improvement would allow for more incremental releases, allow new emissions data to be more thoroughly tested, and result in fewer differences between each subsequent dataset release. This is already the case for the GFED open burning data for the satellite era and is in progress for the CEDS anthropogenic data.

---

[3]For example, the Kazakhmys smelter in Kazakhstan, Raahe sintering plant in Finland, and the Algoma sintering plant in Canada. See the CEDS online documentation (https://GitHub.com/JGCRI/CEDS/wiki/Data_and_Assumptions#so-sub-2-sub-1) for details.

## Appendix A:  Gridded data release history

Below we provide a summary of the release history for gridded emissions data.

### Historical anthropogenic emissions

- June and July 2016: preindustrial (1750–1850) data (ver. 2016-06-18) and historical 1851–2014 (ver. v2016-07-26) for aerosol and reactive gas species were published on ESGF.

- September 2016: files were republished in a new format (with sectors contained within a dimension) due to a limitation in ESGF software on some nodes (version numbers were amended with -sectordim). Data values did not change.

- May 2017: new data version (ver. 2017-05-18) that corrected some errors in the previous gridded data (ver. 2016-06-18 and 2016-07-26; see README and analysis of difference between the new and old datasets at https://goo.gl/tGrc27, last access: 13 June 2017). The new release also included $CO_2$ emissions (all years) and $CH_4$ emissions (from 1970 forward). Note that aggregate country emissions by sector did not change, only the spatial distribution within each country.

- August 2017: "rough cut" extension of $CH_4$ emissions between 1850 and 1970 (in decadal increments) for models that require $CH_4$ emissions over all years. Labeled as supplemental data (ver. 2017-05-18) on ESGF.

- September 2017: new versions of aircraft gridded emissions (ver. 2017-08-30 and 2017-10-05 for $SO_2$) were published on ESGF. This is a cosmetic change only, whereby emissions were set to zero before 1920.

### Historical open burning emissions

- V1.2 (December 2016). Sectoral contributions for all species are now included (in v1.1 only for CO). Minor modifications to the interannual variability between 1960 and 1997 for tropical regions.

- V1.1 (November 2016). Updated emissions, mainly for the boreal regions. Global total fire carbon emissions decreased by about 5 % ($\sim 100\,\mathrm{Tg\,C\,yr^{-1}}$) over 1750–2015.

- V1.0 ($\sim$ September 2016). First release of all emission species including NMVOC speciation.

### Future emissions

As of this writing only one version of the future emissions has been published. Most species and data files were published on ESGF on 28 June 2018. Fossil $CO_2$ emissions and $H_2$ emissions from open burning were published on 19 September 2018.

## References

Amann, M., Bertok, I., Borken-Kleefeld, J., Cofala, J., Heyes, C., Hoglund-Isaksson, L., Kiesewetter, G., Klimont, Z., Schopp, W., Vellinga, N. and Winiwarter, W.: Adjusted historic emission data, projections, and optimized emission reduction targets for 2030 – A comparison with COM data 2013, IIASA, Laxenburg, Austria, available at: http://ec.europa.eu/environment/air/pdf/review/TSAP_16a.pdf (last access: 3 February 2020), 2015.

BP: BP Statistical Review of World Energy June 2016, http://bp.com/statisticalreview, last access: 11 June 2016.

Calvin, K., Bond-Lamberty, B., Clarke, L., Edmonds, J., Eom, J., Hartin, C., Kim, S., Kyle, P., Link, R., Moss, R., McJeon, H., Patel, P., Smith, S., Waldhoff, S., and Wise, M.: The SSP4: A world of deepening inequality, Global Environ. Chang., 42, 284–296, https://doi.org/10.1016/j.gloenvcha.2016.06.010, 2017.

Chin, J. Y. and Batterman, S. A.: VOC composition of current motor vehicle fuels and vapors, and collinearity analyses for receptor modeling, Chemosphere, 86, 951–958, https://doi.org/10.1016/j.chemosphere.2011.11.017, 2012.

Collins, W. J., Lamarque, J.-F., Schulz, M., Boucher, O., Eyring, V., Hegglin, M. I., Maycock, A., Myhre, G., Prather, M., Shindell, D., and Smith, S. J.: AerChemMIP: quantifying the effects of chemistry and aerosols in CMIP6, Geosci. Model Dev., 10, 585–607, https://doi.org/10.5194/gmd-10-585-2017, 2017.

Crippa, M., Janssens-Maenhout, G., Dentener, F., Guizzardi, D., Sindelarova, K., Muntean, M., Van Dingenen, R., and Granier, C.: Forty years of improvements in European air quality: regional policy-industry interactions with global impacts, Atmos.

Chem. Phys., 16, 3825–3841, https://doi.org/10.5194/acp-16-3825-2016, 2016.

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., van Aardenne, J. A., Monni, S., Doering, U., Olivier, J. G. J., Pagliari, V., and Janssens-Maenhout, G.: Gridded emissions of air pollutants for the period 1970–2012 within EDGAR v4.3.2, Earth Syst. Sci. Data, 10, 1987–2013, https://doi.org/10.5194/essd-10-1987-2018, 2018.

Doxsey-Whitfield, E., MacManus, K., Adamo, S. B., Pistolesi, L., Squires, J., Borkovska, O., and Baptista, S. R.: Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4, Papers in Applied Geography, 1, 226–234, https://doi.org/10.1080/23754931.2015.1014272, 2015.

EC-JRC/PBL: Emission Database for Global Atmospheric Research (EDGAR), release EDGARv4.2 FT2012, available at: http://edgar.jrc.ec.europa.eu (last access: 1 August 2019), 2012.

EC-JRC/PBL: Emission Database for Global Atmospheric Research (EDGAR), release version 4.3.1, available at: http://edgar.jrc.ec.europa.eu/overview.php?v=431 (last access: 1 August 2019), 2016.

Endresen, Ø., Sørga°rd, E., Sundet, J. K., Dalsøren, S. B., Isaksen, I. S. A., Berglen, T. F., and Gravir, G.: Emission from international sea transportation and environmental impact, J. Geophys. Res., 108, 4560, https://doi.org/10.1029/2002JD002898, 2003.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Feng, L., Gutierrez, X., Braun, C., Smith, S., and Gidden, M.: iiasa/emissions_downscaling: First Public Release (Version v0.1.0), Zenodo, https://doi.org/10.5281/zenodo.1476850, 2018.

Fioletov, V. E., McLinden, C. A., Krotkov, N., Li, C., Joiner, J., Theys, N., Carn, S., and Moran, M. D.: A global catalogue of large $SO_2$ sources and emissions derived from the Ozone Monitoring Instrument, Atmos. Chem. Phys., 16, 11497–11519, https://doi.org/10.5194/acp-16-11497-2016, 2016.

Fricko, O., Havlik, P., Rogelj, J., Klimont, Z., Gusti, M., Johnson, N., Kolp, P., Strubegger, M., Valin, H., Amann, M., Ermolieva, T., Forsell, N., Herrero, M., Heyes, C., Kindermann, G., Krey, V., McCollum, D. L., Obersteiner, M., Pachauri, S., Rao, S., Schmid, E., Schoepp, W., and Riahi, K.: The marker quantification of the Shared Socioeconomic Pathway 2: A middle-of-the-road scenario for the 21st century, Global Environ. Change, 42, 251–267, https://doi.org/10.1016/j.gloenvcha.2016.06.004, 2017.

Fujimori, S., Hasegawa, T., Masui, T., Takahashi, K., Herran, D. S., Dai, H., Hijioka, Y., and Kainuma, M.: SSP3: AIM implementation of Shared Socioeconomic Pathways, Global Environ. Chang., 42, 268–283, https://doi.org/10.1016/j.gloenvcha.2016.06.009, 2017.

Geng, G., Zhang, Q., Martin, R. V., Lin, J., Huo, H., Zheng, B., Wang, S., and He, K.: Impact of spatial proxies on the representation of bottom-up emission inventories: A satellite-based analysis, Atmos. Chem. Phys., 17, 4131–4145, https://doi.org/10.5194/acp-17-4131-2017, 2017.

Gidden, M. J., Fujimori, S., van den Berg, M., Klein, D., Smith, S. J., van Vuuren, D. P., and Riahi, K.: A methodology and implementation of automated emissions harmonization for use in Integrated Assessment Models, Environmental Modell. Softw., 105, 187–200, 2018a.

Gidden, M., Riahi, K., Smith, S., Fujimori, S., Luderer, G., Kriegler, E., van Vuuren, D., van den Berg, M., Feng, L., Klein, D., Calvin, K., Doelman, J., Frank, S., Fricko, O., Harmsen, M., Hasegawa, T., Havlik, P., Hilaire, J., Hoesly, R., Horing, J., Popp, A., Stehfest, E., and Takahashi, K.: input4MIPs.CMIP6.ScenarioMIP.IAMC, Earth System Grid Federation, https://doi.org/10.22033/ESGF/input4MIPs.10464, 2018b.

Gidden, M., Riahi, K., Smith, S., Fujimori, S., Luderer, G., Kriegler, E., van Vuuren, D., van den Berg, M., Feng, L., Klein, D., Calvin, K., Doelman, J., Frank, S. Fricko, O., Harmsen, M., Hasegawa, T., Havlik, P., Hilaire, J., Hoesly, R., Horing, J., Popp, A., Stehfest, E., and Takahashi, K.: input4MIPs.CMIP6.AerChemMIP.IAMC, Earth System Grid Federation, https://doi.org/10.22033/ESGF/input4MIPs.10443, 2018c.

Gidden, M. J., Riahi, K., Smith, S. J., Fujimori, S., Luderer, G., Kriegler, E., van Vuuren, D. P., van den Berg, M., Feng, L., Klein, D., Calvin, K., Doelman, J. C., Frank, S., Fricko, O., Harmsen, M., Hasegawa, T., Havlik, P., Hilaire, J., Hoesly, R., Horing, J., Popp, A., Stehfest, E., and Takahashi, K.: Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century, Geosci. Model Dev., 12, 1443–1475, https://doi.org/10.5194/gmd-12-1443-2019, 2019.

Goldewijk, K. K., Beusen, A., van Drecht, G., and de Vos, M.: The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years, Global Ecol. Biogeogr., 20, 73–86, https://doi.org/10.1111/J.1466-8238.2010.00587.X, 2011.

Hoesly, R., Smith, S., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J., Li, M., Liu, L., Lu, Z., Moura, M., Cecilia, P., O'Rourke, P. R., and Zhang, Q.: Historical Emissions (1750–2014) – CEDS – v2017-05-18. Version YYYYMMDD[1], Earth System Grid Federation, https://doi.org/10.22033/ESGF/input4MIPs.1241, 2017a.

Hoesly, R., Smith, S., Feng, L.,; Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Supplemental Historical Emissions (1750–2014) – CEDS – v2017-05-18, Version YYYYMMDD[1].Earth System Grid Federation, https://doi.org/10.22033/ESGF/input4MIPs.1242, 2017b.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geosci. Model Dev., 11, 369–408, https://doi.org/10.5194/gmd-11-369-2018, 2018.

Hoesly, R., O'Rourke, P., Braun, C., Feng, L., Smith, S. J., Pitkanen, T., Seibert, J. J., Vu, L., Presley, M., Bolt, R., Goldstein, B., and Kholod, N.: Community Emissions Data System (Version Dec-

23-2019), Zenodo, https://doi.org/10.5281/zenodo.3592073, 2019.

Huang, G., Brook, R., Crippa, M., Janssens-Maenhout, G., Schieberle, C., Dore, C., Guizzardi, D., Muntean, M., Schaaf, E., and Friedrich, R.: Speciation of anthropogenic emissions of non-methane volatile organic compounds: a global gridded data set for 1970–2012, Atmos. Chem. Phys., 17, 7683–7701, https://doi.org/10.5194/acp-17-7683-2017, 2017.

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J., Fisk, J., Fujimori, S., Goldewijk, K. K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J., Lawrence, D., Lawrence, P., Mertz, O., Popp, A., Stehfest, E., Thornton, P., van Vuuren, D., and Zhang, X.: Harmonization of global land use scenarios (LUH2): Historical v2.1h, Earth System Grid Federation, https://doi.org/10.22033/ESGF/input4MIPs.11261, 2019.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., Keating, T., Zhang, Q., Kurokawa, J., Wankmüller, R., Denier van der Gon, H., Kuenen, J. J. P., Klimont, Z., Frost, G., Darras, S., Koffi, B., and Li, M.: HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution, Atmos. Chem. Phys., 15, 11411–11432, https://doi.org/10.5194/acp-15-11411-2015, 2015.

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., Graven, H., Hoffman, F., Ilyina, T., John, J. G., Jung, M., Kawamiya, M., Koven, C., Pongratz, J., Raddatz, T., Randerson, J. T., and Zaehle, S.: C4MIP – The Coupled Climate–Carbon Cycle Model Intercomparison Project: experimental protocol for CMIP6, Geosci. Model Dev., 9, 2853–2880, https://doi.org/10.5194/gmd-9-2853-2016, 2016.

Kirchstetter, T. W., Singer, B. C., Harley, R. A., Kendall, G. R., and Hesson, J. M.: Impact of California reformulated gasoline on motor vehicle emissions. 2. Volatile organic compound speciation and reactivity, Environ. Sci. Technol., 33, 329–336, 1999.

Klimont, Z., Kupiainen, K., Heyes, C., Purohit, P., Cofala, J., Rafaj, P., Borken-Kleefeld, J., and Schöpp, W.: Global anthropogenic emissions of particulate matter including black carbon, Atmos. Chem. Phys., 17, 8681–8723, https://doi.org/10.5194/acp-17-8681-2017, 2017.

Kriegler, E., Bauer, N., Popp, A., Humpenöder, F., Leimbach, M., Strefler, J., Baumstark, L., Bodirsky, B. L., Hilaire, J., Klein, D., Mouratiadou, I., Weindl, I., Bertram, C., Dietrich, J.-P., Luderer, G., Pehl, M., Pietzcker, R., Piontek, F., Lotze-Campen, H., Biewald, A., Bonsch, M., Giannousakis, A., Kreidenweis, U., Müller, C., Rolinski, S., Schultes, A., Schwanitz, J., Stevanovic, M., Calvin, K., Emmerling, J., Fujimori, S., and Edenhofer, O.: Fossil-fueled development (SSP5): An energy and resource intensive scenario for the 21st century, Global Environ. Chang., 42, 297–315, https://doi.org/10.1016/j.gloenvcha.2016.05.015, 2017.

Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Liousse, C., Mieville, A., Owen, B., Schultz, M. G., Shindell, D., Smith, S. J., Stehfest, E., Van Aardenne, J., Cooper, O. R., Kainuma, M., Mahowald, N., McConnell, J. R., Naik, V., Riahi, K., and van Vuuren, D. P.: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodol-

ogy and application, Atmos. Chem. Phys., 10, 7017–7039, https://doi.org/10.5194/acp-10-7017-2010, 2010.

Lee, D. S., Pitari, G., Grewe, V., Gierens, K., Penner, J. E., Petzold, A., Prather, M. J., Schumann, U., Bais, A., Berntsen, T., Iachetti, D., Lim, L. L., and Sausen, R.: Transport impacts on atmosphere and climate: Aviation, Atmos. Environ., 44, 4678–4734, https://doi.org/10.1016/j.atmosenv.2009.06.005, 2009.

Liu, F., Zhang, Q., Tong, D., Zheng, B., Li, M., Huo, H., and He, K. B.: High-resolution inventory of technologies, activities, and emissions of coal-fired power plants in China from 1990 to 2010, Atmos. Chem. Phys., 15, 13299–13317, https://doi.org/10.5194/acp-15-13299-2015, 2015.

McDonald, B. C., De Gouw, J. A., Gilman, J. B., Jathar, S. H., Akherati, A., Cappa, C. D., Jimenez, J. L., Lee-Taylor, J., Hayes, P. L., McKeen, S. A., and Cui, Y. Y.: Volatile chemical products emerging as largest petrochemical source of urban organic emissions., Science, 359, 760–764, 2018.

Müller, C., Stehfest, E., van Minnen, J. G., Strengers, B., von Bloh, W., Beusen, A. H., Schaphoff, S., Kram, T., and Lucht, W.: Drivers and patterns of land biosphere carbon balance reversal, Environ. Res. Lett., 23, 044002, https://doi.org/10.1088/1748-9326/11/4/044002, 2016.

Paulot, F., Jacob, D. J., Pinder, R. W., Bash, J. O., Travis, K., and Henze, D. K.: Ammonia emissions in the United States, European Union, and China derived by high-resolution inversion of ammonium wet deposition data: Interpretation with a new agricultural emissions inventory (MASAGE_NH3), J. Geophys. Res.-Atmos., 119, 4343–4364, https://doi.org/10.1002/2013JD021130, 2014.

Rabin, S. S., Melton, J. R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., Kaplan, J. O., Li, F., Mangeon, S., Ward, D. S., Yue, C., Arora, V. K., Hickler, T., Kloster, S., Knorr, W., Nieradzik, L., Spessa, A., Folberth, G. A., Sheehan, T., Voulgarakis, A., Kelley, D. I., Prentice, I. C., Sitch, S., Harrison, S., and Arneth, A.: The Fire Modeling Intercomparison Project (FireMIP), phase 1: experimental and analytical protocols with detailed model descriptions, Geosci. Model Dev., 10, 1175–1197, https://doi.org/10.5194/gmd-10-1175-2017, 2017.

Riahi, K., Van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., and Lutz, W.: The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview, Global Environ. Chang., 42, 153–168, https://doi.org/10.1016/j.gloenvcha.2016.05.009, 2017.

Schultz, M. and Sebastian, R.: Emission data sets and methodologies for estimating emissions, available at: http://retro-archive.iek.fz-juelich.de/data/documents/reports/D1-6_final.pdf (last access: 19 April 2014), 2007.

Stohl, A., Aamaas, B., Amann, M., Baker, L. H., Bellouin, N., Berntsen, T. K., Boucher, O., Cherian, R., Collins, W., Daskalakis, N., Dusinska, M., Eckhardt, S., Fuglestvedt, J. S., Harju, M., Heyes, C., Hodnebrog, Ø., Hao, J., Im, U., Kanakidou, M., Klimont, Z., Kupiainen, K., Law, K. S., Lund, M. T., Maas, R., MacIntosh, C. R., Myhre, G., Myriokefalitakis, S., Olivié, D., Quaas, J., Quennehen, B., Raut, J.-C., Rumbold, S. T., Samset, B. H., Schulz, M., Seland, Ø., Shine, K. P., Skeie, R. B., Wang, S., Yttri, K. E., and Zhu, T.: Evaluating the climate and air quality impacts of short-lived pollutants, Atmos. Chem.

Phys., 15, 10529–10566, https://doi.org/10.5194/acp-15-10529-2015, 2015.

van Aardenne, J. A., Dentener, F. J., Olivier, J. G. J., Goldewijk, C. G. M. K., and Lelieveld, J.: A 1° × 1° resolution data set of historical anthropogenic trace gas emissions for the period 1890–1990, Global Biogeochem. Cy., 15, 909–928, https://doi.org/10.1029/2000GB001265, 2001.

van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S.: Global fire emissions estimates during 1997–2016, Earth Syst. Sci. Data, 9, 697–720, https://doi.org/10.5194/essd-9-697-2017, 2017.

van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., Arneth, A., Forrest, M., Hantson, S., Kehrwald, N. M., Knorr, W., Lasslop, G., Li, F., Mangeon, S., Yue, C., Kaiser, J. W., and van der Werf, G. R.: input4MIPs.CMIP6.CMIP.VUA, Version YYYYMMDD[1].Earth System Grid Federation, https://doi.org/10.22033/ESGF/input4MIPs.10455, 2016.

van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., Arneth, A., Forrest, M., Hantson, S., Kehrwald, N. M., Knorr, W., Lasslop, G., Li, F., Mangeon, S., Yue, C., Kaiser, J. W., and van der Werf, G. R.: Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015), Geosci. Model Dev., 10, 3329–3357, https://doi.org/10.5194/gmd-10-3329-2017, 2017.

van Vuuren, D. P., Stehfest, E., Gernaat, D. E. H. J., Doelman, J. C., van den Berg, M., Harmsen, M., de Boer, H. S., Bouwman, L. F., Daioglou, V., Edelenbosch, O. Y., Girod, B., Kram, T., Lassaletta, L., Lucas, P. L., van Meijl, H., Müller, C., van Ruijven, B. J., van der Sluis, S., and Tabeau, A.: Energy, land-use and greenhouse gas emissions trajectories under a green growth paradigm, Global Environ. Chang., 42, 237–250, https://doi.org/10.1016/j.gloenvcha.2016.05.008, 2017.

Yang, Y., Wang, H., Smith, S. J., Zhang, R., Lou, S., Yu, H., Li, C., and Rasch, P. J.: Source apportionments of aerosols and their direct radiative forcing and long-term trends over continental United States, Earth's Future, 6, https://doi.org/10.1029/2018EF000859, 2018.

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., Li, H., Li, X., Peng, L., Qi, J., Yan, L., Zhang, Y., Zhao, H., Zheng, Y., He, K., and Zhang, Q.: Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions, Atmos. Chem. Phys., 18, 14095–14111, https://doi.org/10.5194/acp-18-14095-2018, 2018.