A Community Emissions Data System (CEDS): Emissions For CMIP6 and Beyond

Steven J Smith,* Yuyu Zhou, Page Kyle
Joint Global Change Research Institute (JGCRI), PNNL, 5825 University Research Court, Suite 3500, College Park, MD 20740
* Also Department of Atmospheric and Oceanic Science, University of Maryland. ssmith@pnnl.gov

Hailong Wang
Atmospheric Sciences & Global Change Division, PNNL, Richland, WA

Hongbin Yu
Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD & Climate & Radiation Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD

Abstract

Historical estimates for emissions of greenhouse gases, anthropogenic aerosol (BC, OC) and aerosol and ozone precursor compounds (SO_2, NO_x, NH_3, CH_4, CO, NMVOC) are key data needed for Earth System Models, Integrated Assessment Models and research more generally on aerosols, air pollution, and atmospheric chemistry, air quality, related health impacts, and the design of emissions control policies. Current global emissions data sets have a number of shortcomings, including timeliness and transparency. With the involvement of the global stakeholder community and building on current global efforts, we will implement a community emissions data system that will produce annually updated estimates of global emissions using a data-driven, open source framework. The basic methodologies to be used for this system have been documented in recent journal papers and are designed to complement and extend existing emission inventory efforts. Because energy and other driver data will be incorporated as part of the emissions data system, aggregate emissions factors can be estimated in order to examine consistency over time and space. Emissions estimates will be extended to one year before present, albeit with additional uncertainty.

Current global emission estimates are generally not provided with uncertainty values. A central component of the proposed system will be uncertainty estimation at the country and sector levels. The data system will provide transparent, timely, and country-specific estimates that are consistent across species.

The data system is constructed using the R open source platform. The use of an open source platform will enable the system to be released to researchers to perform their own research or develop alternative scenarios. The data system will allow researchers, analysts, and policy-makers to access data by sector, fuel, and country, which will allow an unprecedented community involvement in emissions research and data development.

The emissions data and data system will be developed in close cooperation with national and international emissions inventory teams and stakeholder groups. The project is guided by a steering committee, with international engagement also facilitated in cooperation with the Global Emissions Inventory Activity (GEIA). The community emissions data system project will provide global and regional pollutant emissions that are available up to the most recent full year with consistent uncertainty estimates. The project timeline has been designed to provide an interim historical emissions dataset for use in the CMIP6 model inter-comparison suite of projects that provide foundational data for scientific advances and international assessments, including IPCC assessments. This widespread use of the data produced by this project will result in model projections and analyses that are using up to date annual emission estimates, which will result in modeling results with greater scientific value and policy relevance.

Table Of Contents

1. Introduction

Historical estimates for emissions of greenhouse gases, anthropogenic aerosol (BC, OC) and aerosol and ozone precursor compounds (SO_2, NO_x, NH_3, CH_4, CO, NMVOC) are key data needed for Earth System Models, Integrated Assessment Models, and atmospheric chemistry and transport models, both for general analysis and for model validation through comparisons with observations. Such comparisons, however, are only as good as the quality of the emissions input data used to drive the models. Emissions data are also used for calibrating and improving future projections from integrated assessment models. While detailed historical data exist for GHG concentrations and (to a lesser extent) solar and volcanic forcings, data on aerosol and chemically reactive species are relatively crude. Because of the short atmospheric lifetime of many of these species, spatial location matters, as do the large changes in emissions that have occurred in many regions over decadal or shorter timescales.

Satellite and other Earth-system data are increasingly available in near real-time, but global emission estimates lag by 5-10 years. While satellite data products show promise for a variety of uses, these products cannot currently replace bottom-up emission inventories. Indeed, inversion studies have found that the choice of default emission dataset influences inversion results (Huneeus et al. 2013). Improved inventory estimates will, therefore, enhance the use of satellite data. Inversely, satellite data can potentially be used to improve inventory estimates.

We will implement an open-source data system that will allow the production of global and regional emission datasets with improved temporal and spatial resolution. Seasonality of anthropogenic emissions will be included as well as other characteristics as deemed important by the user community. The data will be policy-relevant (e.g., using country-level inventories where these are complete), consistent over time, as up to date as possible, and will include uncertainty. These data will enable improved validation and assessment of aerosol and cloud model components, improve the emission data needed for both historical attribution and near-term climate predictions, provide uncertainty estimates needed for uncertainty quantification (UQ) research, and allow enhanced use of satellite data products. The goal is to produce annual emission estimates, focusing on aerosol and ozone precursor compounds over the entire industrial era, by country, sector, and fuel (plus spatial emission grids), with uncertainty analysis, as diagrammed in Figure 1 below.

The following sections provide background information, a description of the proposed data system, and overall project timeline.

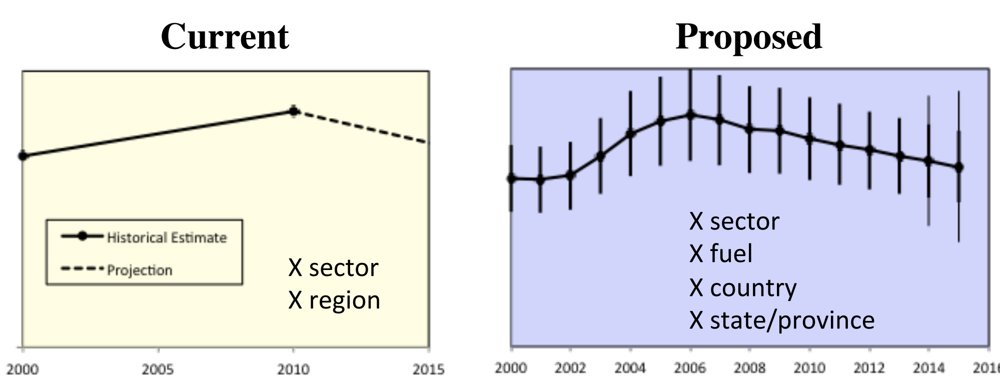

Figure 1. This project will provide emissions data more frequently, extending to more recent years (starting from 1750), and with uncertainty estimates. Emissions for the most recent years will be subject to additional uncertainties.

2. Background

2.1 Need for and use of emissions data

Emissions estimates for anthropogenic aerosol and chemically active compounds are used for a variety of purposes. Simulations of climate and atmospheric chemistry require gridded inputs of emissions of these compounds. Model performance in areas such as aerosol formation, transport, cloud interactions and deposition, and ozone concentrations can be evaluated by comparing to surface, air, and satellite observations. Such comparisons, however, are only as good as the quality of the emissions input data used to drive the models. Many other modeling efforts use the outputs of such analysis, thus indirectly making use of emission data. Examples include ozone and oxidant fields used in climate models lacking an atmospheric chemistry component and nitrogen and sulfur deposition fields used by ecological and agricultural models.

Uncertainty quantification, as related to Earth System Models, requires estimates of uncertainties in both input data and observational quantities. At this point, however, uncertainty information for emissions data is lacking and supplying such information is one goal of the current project.

Emissions data are also used as a starting point for future projections from integrated assessment models. Longer time series of emissions data can be used to examine the relationship between emissions trends, driving forces such as fuel use, and socio-economic variables such as income. These relationships then inform the development of future projections.

Emissions data are also used directly for policy purposes. Because of their adverse impact on human health, and natural and managed ecosystems, pollutant emission amounts are estimated by environmental agencies in nearly every high-income country, and an increasing number of developing countries. Information on emissions at various levels (sector, fuel, technology, and region) is used to evaluate pollution and greenhouse gas mitigation strategies. Greenhouse gas and pollution policies overlap in analysis of Short-Lived Climate Forcing (SLCF) agents. The output of Earth system and atmospheric chemistry/transport models are also regularly used in the policy process to guide decision-making, making indirect use of emission datasets.

2.2 Current Global Historical Emissions Data

There are various sources of pollutant and greenhouse gas emission estimates. Environmental agencies in many countries conduct detailed assessments of air pollutant emissions for regulatory purposes, and many of these are available in a common format through UNFCCC (United Nations Framework Convention on Climate Change) reporting. Estimates are generally made available 2-3 years before the present, although with varying levels of detail. Comprehensive estimates for pollutant emissions in the United States, for example, are conducted every few years as the National Emissions Inventory (NEI). Trends for intermediate years are interpolated using a combination of methods, including continuous emission monitoring data available from most electric power plants, detailed modeling for mobile sources, and simple trend estimates.

One of the most important global emission data efforts is the EDGAR (Emissions Database for Global Atmospheric Research) project, which produces comprehensive estimates of air pollutant and greenhouse gas emissions, currently available from 1970 to 2008 (version 4.2) for pollutant emissions. EDGAR estimates annual emissions by substance, detailed sector and country. EDGAR produces emissions estimates that are independent of official national statistics. While this provides a useful check on emissions estimates, trends are not necessarily consistent with country-level data, resulting in significant differences in magnitude and trend in some cases (Figure 2), differences that provide insights into uncertainty in emission estimates.

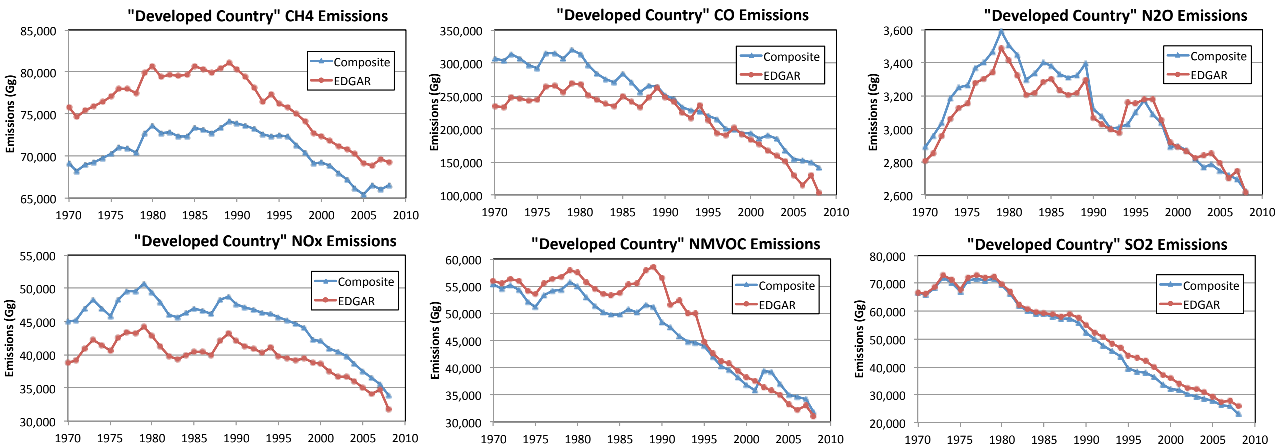

Figure 2. Comparison of emission estimates from EDGAR with emissions from country-level inventories for developed countries (e.g. largely countries that were members of the Organization for Economic Cooperation and Development as of 1990 including all of the current European Union). This figure shows that, even at this aggregate level, there are important differences in trends between the EDGAR emission estimates and country-level data.

The GAINS (**G**reenhouse Gas and **A**ir Pollution **In**teractions and **S**ynergies) model, while focusing on Europe and Asia, also produces estimates of global emissions in roughly five-year increments. GAINS estimates are available by sector and fuel for most medium to large countries of the world and some sub-country regions. GAINS estimates are informed by country-level data where those are considered reliable. There are also a number of recent studies focused on emissions from Asia such as the REAS (Regional Emission inventory in ASia) project (Kurokawa et al. 2013) and global estimates from the Peking University emissions group (Wang et al. 2013, Huang et al. 2014).

The current historical global emissions dataset used by the global atmospheric modeling community (Lamarque et al. 2010) was produced by combining data from a number of sources. For most anthropogenic emissions, country-level estimates since about 1990, including data reported to the UNFCCC and from the REAS project, were used with default emissions from EDGAR for other countries. Extrapolation back in time was based on a combination of the EDGAR-HYDE (van Aardenne et al., 2001) and RETRO (REanalysis of the TROpospheric chemical composition, Schultz et al., 2010) estimates. Estimates for sulfur dioxide (SO_2) emissions for all years were from Smith et al. (2011), who calibrated estimates to country-level inventories using methodologies that will be adapted in this project. Estimates of BC and OC emissions for all years were updated from Bond et al. (2007). Emissions from shipping and aircraft were drawn from bottom-up estimates (Erying et al. 2010, Lee et al. 2010). Emissions gridded to 0.5° by sector were provided at ten-year intervals from 1850-2000, with the emissions distribution based primarily on year 2000 emissions grids from EDGAR. Emissions from grassland and forest fires (Schultz et al. 2008, Mieville et al., 2010, van der Werf et al., 2006) were added to form a complete dataset of emissions to the atmosphere.

The Lamarque et al. (2010) dataset was a compilation of a set of "best available" estimates from a variety of sources. As with emissions data used for the Hemispheric Transport of Air Pollutants (HTAP) assessment (Janssens-Maenhout et al. 2012), it was judged to be important to use country-level emission estimates where these were found to be reliable and complete so that the resulting analysis was policy-relevant. While this effort was a major advance in terms of consistency and completeness, these data, however, have a number of shortcomings including limited temporal resolution, different methodologies between gases, lack of comprehensive uncertainty analysis, a most recent data point of year 2000, and a methodology that was not designed to be easily replicated. The current project stems from experience producing these previous datasets and, as described below, is designed to provide the comprehensive emissions data needed for Earth System Modeling in a manner that is readily updatable over time.

2.3 Emission Data Needs

While no one product or data system can meet all the needs of the global modeling community, drawing from discussions at a recent workshop (Granier et al. 2014), we aim here to develop a system that meets the following key criteria (discussed in more detail in following text):

- Annual temporal resolution with seasonal cycles (& diurnal cycle parameterizations)
- Resolution by broad sector and fuel
- Consistency between emissions species
- Facilitate improved VOC speciation
- Annual updates
- Uncertainty estimates
- Transparent assumptions
- Policy relevance

These criteria are inter-related, such as the requirements for more recent data (& therefore annual resolution) and estimates of uncertainty. Information on emissions controls, emission factors, and driver data (e.g. fuel use) takes time to collect, analyze, and validate. Even in the well-developed emissions data systems in the United States, for example, estimates do not stabilize for several years as data revisions take place (e.g. Table S-7 in Smith et al. 2011). Further, we will estimate emissions beyond the point in time where detailed country-level estimates are available. Emissions estimates for the most recent few years will, therefore, be more uncertain in general, meaning that it is important that uncertainty information be provided along with the emissions estimates (as detailed below).

The level of sectoral detail in the data system will be determined, in large part, by the need to meet the above criteria. The need for seasonal emissions estimates, for example, points to a need to consider building heating fuels separately (particularly biomass such as wood fuels, a major source of pollutant emissions in this sector). Not all needs can be met with one system. The need for VOC speciation points to at least a broad breakdown of emission by sector (transportation, industry, fossil-fuel tanker loading, etc.). However, more complex schemes for VOC speciation might require a more detailed breakdown. The EDGAR emissions estimates, for example, contain a very high degree of sectoral detail and could always be used for this purpose by potential users: there is no need to duplicate that higher level of detail in this work. We will also consider adding "additional PM2.5" as a category to represent PM2.5 emissions that are not already included in the BC and OC categories.

3. Proposed Community Emissions Data System

3.1 Overview

Bottom-up emissions estimates are obtained by multiplying a driver magnitude (fuel consumption, smelter output, number of cattle, etc.) by an emissions factor. The emissions factor can also incorporate the presence of emission control technology that reduces emissions of one or more substances. Accurate estimates of emissions require detailed data on technology shares, operating characteristics, and application of emission controls. Bottom-up estimates at varying levels of detail are regularly conducted in many countries by environmental agencies.

In some cases emissions are measured using Continuous Emissions Monitoring (CEM) systems, such as CO_2, SO_2 and NO_x emissions from most large electric power plants in the United States.[1] Bottom-up and CEM emission estimates have been found to disagree (Ackerman & Sundquist 2008), which is useful information to inform uncertainty analysis.

The goal of the CEDS system is to combine existing emissions estimates with driver data to be able to produce consistent estimates of emissions over time. The system will focus on emissions of anthropogenic aerosol (BC, OC) and aerosol and ozone precursor compounds (SO_2, NO_x, NH_3, CH_4, CO, NMVOC) over the entire industrial period (1750 – present). Emissions for CO_2, consistent with CDIAC assumptions, will also be provided as reference emissions.

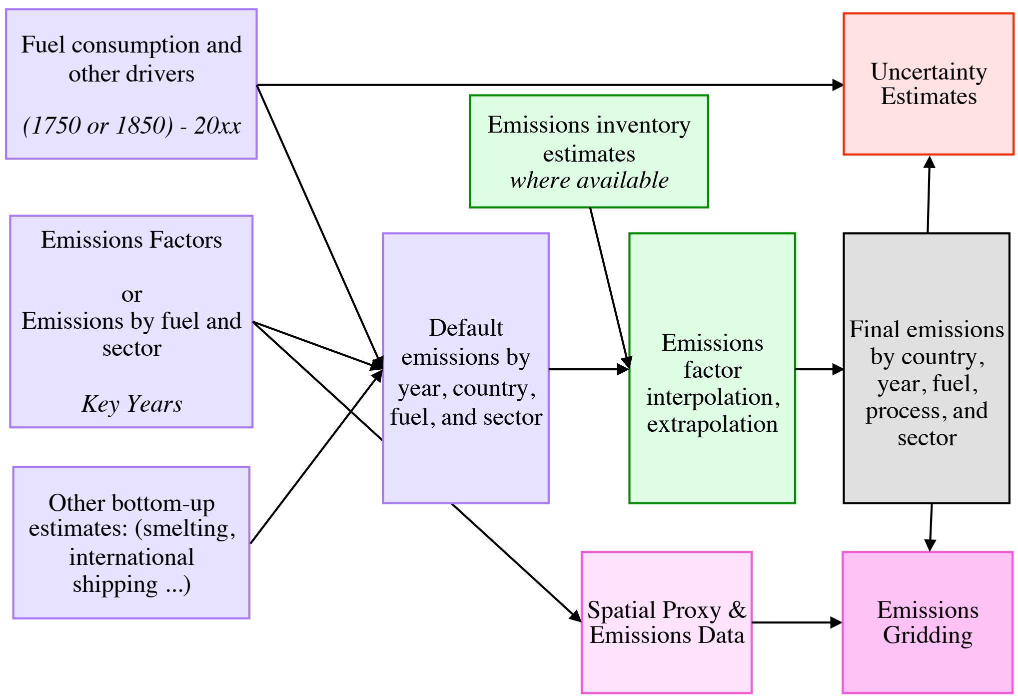

Figure 3. Simplified diagram of the emissions data system.

An overview of the system is given in Figure 3. Note that the proposed system focuses on anthropogenic sources and is not envisioned to cover forest and grassland fires, for which different methodologies are appropriate. The system is designed as follows:

- Where detailed emissions information is available and judged to be reliable (e.g. US EPA National Emission Inventory and US CEM measurement data, and data from most OECD countries), estimates from the data system will be calibrated to match this information. Emissions data from a variety of other sources such as (but not limited to) EDGAR (http://edgar.jrc.ec.europa.eu), GAINS (http://www.iiasa.ac.at), and REAS (Kurokawa et al 2013) will be used where country-level data is not available.
- Driver data, such as fuel use, will be collected and processed so that aggregate sector-level emissions factor trends can be estimated.

[1] http://www.epa.gov/airmarkets/emissions/continuous-factsheet.html

- Emissions estimates will be extended to recent years using trends in driver data (such as fuel use) and trends in emission factors at the sectoral level.
- Emissions estimates will be extended back in time using similar principles, with emissions factor trends set as appropriate for each sector and gas.
- Uncertainty will be estimated for all emissions as part of the output of the system.

The initial development of this data system will be conducted at JGCRI, but will be performed within a context of involvement by emissions experts from around the world as well as stakeholder groups (such as the CMIP community).

The stakeholder community will be closely involved in the design and deployment of the emissions data system as described further below. The data system will be released as open-source software, which will allow researchers, analysts, and policy-makers to access data by sector, fuel, and country that will allow an unprecedented community involvement in emissions research and data development (Frost et al. 2012). Unlike previous work where only emissions data are released, the data system will link emissions data and the associated driver data (such as energy consumption), which will facilitate research not only on trends of emissions, but also on the relationships between emissions and drivers across time, countries, and sectors. The data system may also be particularly valuable for decision-makers in developing countries where emissions estimation and analysis infrastructure may not otherwise exist.

3.2 Methodology

Central to this project is the calibration of emissions estimates to data from sources judged to be authoritative. This can be described as a "mosaic" approach, where we produce a "best estimate".[2] The first implementation of this system will use updated versions of data sources similar to those used in Lamarque et al. (2010) in order to produce an interim data product in time for use in the CMIP6 exercise, although new data sources will be used as needed and appropriate.

In order to accurately extrapolate to other time periods, and also to provide emission estimates at the level of fuel and sector, the system will require a set of default emissions factors (or, equivalently, emissions) at the level of fuel and sector. Because emissions data are commonly provided only at the sector level, we will, in essence, be downscaling the estimates to the level of fuel and sector by using the default emissions factor data. This methodology has been demonstrated by Smith et al (2011). There are a number of sources for the default emissions factor data. For Europe, China and India, the GAINS project, for example, supplies a publicly accessible database of detailed information on emission factors and emission control levels. These and other sources, such as Pulles et al. (2007), will be used to build up a default emissions factor database for key years (for example, at least each decade from 1960 through 2010). While in general, it may only be necessary to compile default emission factors for selected key years, it would require little additional effort to use annual estimates in the data system if these are readily available. The data system is flexible, allowing additional temporal or sectoral detail wherever needed.

The accuracy of this step to downscale emissions to the level of fuel and sector depends on having a reasonable set of default emissions factors. Based on our previous experience, we expect that we will have errors for this step that are well within the overall uncertainty of the emissions estimates. We will, however, estimate the errors involved as the analysis proceeds. For areas with high quality emissions data in the first place (e.g. United States, Europe, etc.), emissions factors con-

[2] The use of the indefinite article is deliberate. There are multiple possible "best" estimates, and this flexible data system will facilitate the examination of the implications of using alternative data sources and assumptions.

sistent with these emissions will generally also be available. This also will allow us to test the accuracy of this approach by comparing the "downscaled" emissions data using default emissions factors with the more detailed inventory information. In areas where emissions are more uncertain, such as developing countries, there will be larger uncertainty in this step; however, this is expected since emissions in these areas are also more uncertain in general.

For some countries, national emissions data will be sufficient to capture the key trends over recent decades. SO_2, NO_x, PM, and CO emissions for the United States, for example, extend back to 1970 before the imposition of emissions controls. This will enable us to capture the key trends in emissions factors over the time that when clean air regulations were implemented. For countries where such data do not exist we will use other data sources. For the road transportation sector, for example, estimates of fleet-average emissions factors of vehicles before emissions controls versus those in more recent years (e.g. Yan et al. 2011, Wang et al. 2011) could be used to guide trend estimates. There may, however, be trends over earlier periods that will need to be captured and literature reviews will be conducted to guide any needed adjustments in emission factor trends. In this project we will focus on capturing trends in the most important sectors with the anticipation that continued community involvement will result in future improvements.

Because both reported emissions and emissions driver data will be consistently included within the same data system, we can evaluate the consistency of implied emissions factors at the broad sectoral level and assure that trends are consistent over time. This will be done both across time and across countries. This addresses one of the known issues with the current historical emissions dataset (Lamarque et al. 2010), where different datasets were combined to form historical time series by simply smoothing over discontinuities. This analysis will both inform uncertainty analysis and also allow us to produce more consistent estimates by correcting for discrepancies in the existing data. We will not be able to definitively resolve all discrepancies. However, by documenting our findings, and by providing a community dataset and data system, we will provide tools for the emissions community to make progress toward resolving such issues.

The second set of data central for this system is the driver data. This includes fuel consumption and fuel production, as well as production data for various processes that generate emissions such as metal smelting, cement production, wood pulp production, and agricultural and livestock production. The most central data are those for fuel consumption. For recent decades we will use International Energy Agency (IEA) energy statistics.[3] This is the most authoritative global data, is updated annually, and has the necessary level of sectoral detail.

We will extend this data to the most recent full year by using the annually released BP world energy statistics for relevant sectors as demonstrated in Klimont et al. (2013). The BP data provide only total energy consumption by fuel. The sectoral distribution of each fuel will likely be assumed equal to the most recent historical year available in the IEA dataset. This adds additional uncertainty to emissions estimates for more recent years, which will be taken into account when uncertainties are estimated, as described below.

IEA energy data is generally available from 1960 for OECD countries and from 1971 for other countries and three composite regions. Total fuel consumption by country can be obtained back to 1950 using UN data, and from 1750-1950 using Andres et al. (1999). These are the same data used to produce the CDIAC historical CO_2 estimates. A disaggregation of fuel consumption by sector has been estimated by Bond et al. (2007), with historical biomass consumption estimated by Fernandes

[3] Energy Statistics of OECD Countries and Energy Statistics of non-OECD Countries (http://www.iea.org). We will explore options with the IEA of making these data available for users in developing countries.

et al. (2007). We will use these and additional sources as historical driver data. Consistent with the open-source nature of this system we will use data that either 1) can be distributed with the data system or 2) can be obtained and inserted into the data system by the user.

Once driver data over the relevant years has been compiled, emissions will be consistently extended both forward and backward in time. In order to extend emissions to more recent years, trends in emission factors by fuel and sector will be used. For OECD regions, country-level inventory data are generally available with about a 2-3 year lag. Emissions in these cases will, therefore, only be extrapolated over an additional 1-2 years. For developing countries, emission estimates can lag much further, so the extrapolation will be over a longer period. For countries where emission control technologies are not being widely adopted, emission factors are not likely to change dramatically so additional errors are likely to be relatively low. The most difficult cases are the generally middle-income countries where emissions controls and/or combustion technologies are changing more rapidly. One particularly important example is China. Literature results and an international network of emissions stakeholders will be used to the guide assumptions for these cases, including assumptions for uncertainty. In the case of China specifically there is an increasing body of literature addressing pollution trends that can be used for this work (Lu & Streets 2011, Wang et al 2012, Zhao et al. 2013).

The stakeholder community (Granier et al. 2014) has indicated that CO_2 emissions consistent with the spatial and sectoral resolution provided in this project are also needed as CO_2 emissions are used as a reference emission in many applications (Pollack et al. 2012, Brioude et al. 2013, Peischl et al., 2013). Measurements of regional and also plant-specific emissions of aerosol and ozone precursors, for example, rely on ratios of the target pollutant to CO_2 emissions. We will collaborate with CDIAC to apply assumptions consistent with those used at CDIAC to produce CO_2 emissions at the same resolution (country, sector, fuel) as the other emissions in this project.

Emissions seasonality is an important feature that is needed in an updated data set (Granier et al. 2014). In at least one case, model experiments indicate that the largest gain through incorporating seasonality into emissions data is on the monthly level (Stohl et al 2013), with a much smaller gain due to the incorporation of daily emissions variations. Our initial approach will therefore be to use globally applicable methods to estimate emissions seasonality at the monthly level, while allowing for region-specific parameterizations in the future as such information becomes available. As mentioned above, data characteristics such as the choice of sectors and fuels, will be guided, in part, by the need to estimate the seasonal distribution of emissions. Drawing from current (http://aerocom.met.no) and past efforts (Bolscher et. al 2007), we will also provide parameterizations that can be used to produce diurnal variations.

Emissions from non-combustion processes are important for many species and these will also be included in the data system. Extrapolation of emissions estimates for these processes will generally be simpler than for combustion emissions, with emissions scaled with trends in the appropriate driving variables. For some specific emissions, such as SO_2 from metal smelting, specific bottom-up mass balance estimates are available (Smith et al. 2011) and will be incorporated into the data system.

We note that this scaling procedure can result in errors at levels of aggregation finer than in the supplied inventory. One of the tasks in the project will be to evaluate and document the magnitude of the errors in this step, which we will aim to minimize. Note that, because we are also estimating overall emissions uncertainty, we will have a natural metric to quantify if any errors in such extrapolation or estimation steps are important. If uncertainty and errors introduced by these extrapolations are much smaller than the overall uncertainty in emissions from a particular sector then we would consider this acceptable.

Note that for some countries, such as the USA, we can derive highly specific emission factors by sector and fuel from the US National Emission Inventory (NEI) data (and we will likely use these as default values for Canada, where technologies and procedures are broadly similar). For European and most Asian countries we will use country-specific default emissions factors from the GAINS project.

In general, where emissions data are of high quality, more detail will be available, minimizing errors in this extrapolation step. Where this detail is not available, emissions estimates are generally more fundamentally uncertain, which means that extrapolation errors of this type are unlikely to be a dominant cause of errors in the results. As noted, we will be able to evaluate if whether this hypothesis is correct by evaluating the uncertainty induced by sector scaling as compared to the overall uncertainty estimates.

3.3 Emissions Uncertainty

Few or no existing comprehensive global emissions datasets are produced with uncertainty estimates. Where uncertainty estimates are produced, they are often time-invariant and only general estimates by emission species (e.g. see assessment in Blanco et al. 2014 §5.2.3). In this work, it is particularly critical that time-dependent uncertainty estimates be produced given that the extrapolation procedures described above will produce greater uncertainty in the most recent years in both driver data and emission factors (see also Smith et al. 2011, §S.14 for an example).

Estimation of uncertainty for emissions is difficult because traditional approaches are generally not feasible at a global level due to the lack of well-defined statistical sampling for all the necessary variables, although Monte-Carlo uncertainty analysis has been applied at the country level (Winiwarter and Rypdal 2001, Lu et al. 2011, Joerss 2013). In all current approaches a substantial element of expert judgment is present, although as demonstrated in Smith et al. (2011), comparisons of different emissions estimates can be helpful to inform estimates of emissions uncertainty.

The uncertainty analysis for this project will start with a “tiered” approach (Smith et al. 2011, Andres et al. 2014), whereby a relatively small number of tiers of uncertainty for drivers and emission factors are defined, and these are applied by sector and fuel to determine country level uncertainty. The mapping between uncertainty levels and country and/or sectors is a judgment call, but this procedure is transparent and allows easy experimentation with different assumptions. While mathematically more sophisticated methods such as Monte-Carlo techniques can be used to generate uncertainty estimates, given the lack of the fundamental data (such as statistical distributions) to provide inputs to such methods, we feel a simpler and more transparent approach is more appropriate for long-term, global emissions data. The exact methodology for the uncertainty analysis will be determined after a literature review and consultation with our steering committee. We note, for example, that Bond et al. (2004) make a case for use of lognormal distributions for such estimates.

Comparisons among existing inventory datasets can be quite helpful in this regard. We will expand the uncertainty analysis by examining literature sources and emission factor databases (e.g., EPA’s AP-52 series) to assign uncertainties by sector and emission (Bond et al. 2004). In most work to date, while uncertainty assumptions are stated, the basis for these assumptions is often not well described. In this work we will focus on collating and documenting the data used to assess uncertainty values so that these results are accessible to users and other researchers, and can be used as a basis of future research and refinement. We will consider correlation in uncertainties across sectors (e.g. Bond et al. 2007), although we note that we will be conducting analysis at a moderate level of sectoral disaggregation which means that the cross-sectoral correlation between uncertainties will often be low (e.g. uncertainties in agricultural sector emissions will not be strongly correlated with uncertainties in the power sector).

We will take care to assign consistent relative uncertainties across emission species (within a given region or group of countries) within the same source category in order to provide consistent uncertainty estimates across emissions. In general, for example, we expect combustion sector uncertainties to be ordered approximately as CO_2 < SO_2 < CO, as CO_2 & SO_2 emissions depend largely on fuel properties while emissions from incomplete combustion (NO_x, CO, NMVOC, BC, OC) can have a strong dependence on combustion conditions and are more uncertain in general.

One particularly difficult issue is uncertainty correlations across different countries. We will experiment with methodologies based on spatial, sectoral, and other metrics (e.g. economic metrics) in order to develop realistic methodologies for representing correlations in uncertainty between countries. Similar issues may also apply spatially within countries, although we do not anticipate being able to examine uncertainty at the sub-country level in this project.

We also will estimate the additional uncertainty associated with extending emissions to more recent years (e.g. Figure 1). This uncertainty arises from several sources. First is the preliminary nature and lack of sectoral detail for energy and other driver data. Second, there is also a lack of detailed emissions factor, technology, and emissions control information for recent years, which means that emissions factor trends need to be used for estimation in many cases. This additional uncertainty for recent years is also reflected in country-level inventories (e.g. Smith et al. 2011). Both of these uncertainty components can be rigorously assessed by applying the data estimation methodologies to data released in previous years and comparing the resulting estimates with emissions estimates released at later dates.

Once uncertainty is estimated, ensembles of historical emission estimates can be produced. These can either be used in conjunction with simple models to evaluate, for example, historical forcing uncertainties (e.g. Smith and Bond 2014, who examined forcing but not emission uncertainty) or used in more complex models. One possible future experiment might evaluate the impact of emissions uncertainty by examining a large number of historical emission ensembles using an atmosphere and chemistry model coupled to a slab ocean.

The result of the uncertainty portion of the project will be uncertainty estimates that are consistent across emissions and countries. Uncertainties assigned to driver data, such as fossil fuel consumption, will consistently flow through to emissions uncertainty. The same methodologies and analysis will be used to assign uncertainty to emission factors, including uncertainty in emission control application/effectiveness.

3.4 System Architecture

The emissions data processing system is being built using the R open-source analysis platform (http://www.r-project.org/). R is a widely used platform for statistical analysis, data processing, and graphics with a rich set of available packages to extend the functionality of the application. Our choice of this platform is informed by our group's experience in using R for data processing, including the development of a comprehensive data processing system for producing input data files for the GCAM integrated assessment model. In preliminary work to date, we re-purposed some of the code from the GCAM data system for use in the emissions data system.

The overall data system architecture is illustrated in Figure 3 and data modules will mirror this overall structure. The system is designed to be data-driven, with code modules processing data at each step and producing output data that feeds into the next module. A makefile system is used to efficiently re-calculate only data that depends on input data, assumptions, or code that have changed. With this data-driven system, a variety of data sources and calculations can be accommodated with custom pre-processing modules as needed. Emission factors, for example, can be used either as aggregate emission factors, uncontrolled emission factors in conjunction with abatement efficiency, or gas or sector-specific emission factors (e.g. sulfur content, ash retention percentage, and control efficiency for SO_2 from coal-fired technologies). Alternatively, sector and fuel-specific base-year emissions can be used. Each input format is translated by specific modules into a common aggregate emissions factor format used in the core calculation modules.

A comprehensive, cumulative meta-data system is built into the data system. The meta-data allows specification of the geographical region/country, time period, and source for each input data file. This will include, for example, energy data (IEA, BP, or other by year), calibrated emissions, and default emission factors. Meta-data for each file read into the system is collected and accumulated along the calculation chain so that each final output file has an associated meta-data document. In this way the source of driver data and emissions assumptions will be automatically generated and archived along with the output data.

One issue associated with this "mosaic" approach for emissions calculation is that there could be temporal or spatial inconsistencies in emissions data. Data analysis modules will be constructed to examine breaks in temporal trends that will be flagged for further examination and possible resolution. Note that we have found such trend breaks even in ostensibly consistent emissions data, such as EPA emissions trends data, some of which are likely due to changing emission estimation methodologies over time. While we cannot resolve all such issues within this project, we can identify places where more work is needed. A second set of analysis will be to examine how emissions factors compare across countries. In some cases, different technology mixes or other local circumstances would be expected to produce different emission factors, while in other cases we might expect more similarity. In some cases, historical trends may be adjusted where existing data show large inconsistencies. Again, this analysis will also help identify where current data sets may not be consistent and where more work may be needed to improve historical emissions estimates. In both of these cases, these analyses will provide valuable data to inform emission factor uncertainty assumptions.

3.5 Emissions Gridding

The fundamental output of the data system is emissions by country, sector, and fuel. These emissions will be downscaled to a spatial grid, also referred to as "gridding" (Figure 4). The primary data source for the present-day distribution of emissions is likely to be existing high-resolution emission grids (such as those from EDGAR and EDGAR-HTAP), although other proxy data may be used for specific sectors (e.g. gas flaring, agricultural waste burning on fields, etc.).

The target resolution for the gridded emissions data will be 0.1° (Granier et al. 2014). There are two general approaches to global emissions gridding. The first is to use globally consistent sets of spatial data (power plant locations, road networks, etc.) to downscale emissions for all countries and regions. This is the approach used by EDGAR and Lamarque et al. (2010). A second approach is to combine default global gridded emissions (e.g. EDGAR) with regional high resolution emissions grids developed by regional air pollution regulators or research groups (e.g. TNO, Environment Europe, US EPA, REAS, etc.), an approach has been taken for the most recent round of TF HTAP experiments (HTAP_v2).

Both approaches have their strengths and weaknesses. Consistent global datasets hold more potential for consistent scaling over time and have fewer issues with data continuity and consistency at geo-political borders. Regionally developed emissions grids often can make use of more detailed spatial information collected at the country level. These gridded emission data are often used for detailed regional modeling work. Use of such grids would lead to a global dataset with regional gridded emissions that will tend to be more compatible with regional analyses, but will not necessarily be consistent across regions. We will examine experience with both methods and consult with stakeholder groups before deciding which approach to use.

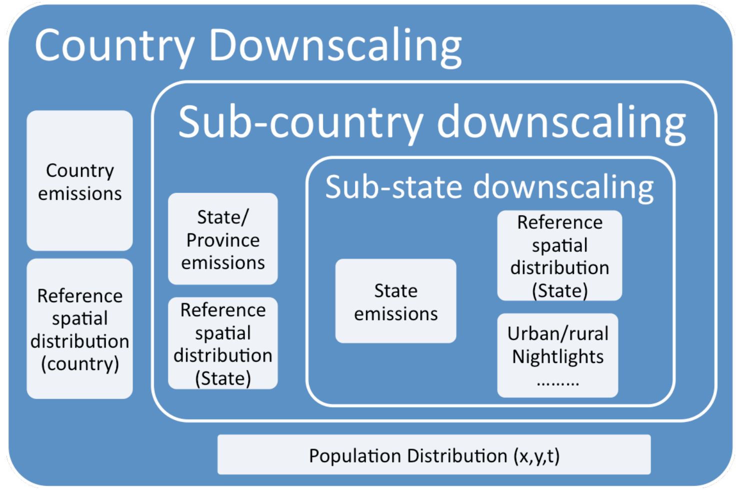

Figure 4. Schematic description of emissions downscaling.

Note that in either case, we will use emissions grids and other spatial proxy information that are publically available.

In either method, additional assumptions and methodologies will need to be used to downscale emissions to the grid level for past years. In general, the necessary proxy data do not exist for much of the historical period examined here. The simplest approach is to use the emissions grids that were used for recent years for the recent past, scaling within each country and sector to match estimated emissions. This will map emissions generally to the correct geographic location. One potential advantage of the mosaic grid approach is that detailed historical emissions grids may be available for some regions that could be used directly (although experience to date indicates that this involves some additional work to resolve boundary issues).

In Lamarque et al. (2010) the year 2000 emissions grid was used for the most recent few decades, with the spatial distribution for most emissions sectors transitioning to a population-based grid by 1900. The use of a population-based emission distribution for periods before 1900 is also envisioned in the current project. We will explore available data sources for gridding the intermediate period.

As indicated in Figure 4, we will also examine methodologies for capturing changes in urbanization level at all spatial scales by splitting emissions into urban and rural components. One consideration in the development of the gridding algorithms within the data system is a design that allows community members to produce (and contribute to) enhanced emissions grids for specific regions.

3.5.1 Sub-Country Emissions Detail

For large countries, downscaling emissions using the same spatial grid over the last several decades could lead to substantial errors in the location of emissions, even when used by global models. This would introduce errors in any comparison between modeled and observed data that can be avoided by capturing the spatial shifts of emissions at a sub-country level.

In this task we will augment the data system to downscale emissions at the sub-country level, that is, by state or province. This task will entail the collection of driver and emissions information at the sub-national level, and the development of appropriate downscaling methodologies. We will first focus on the United States, where substantial historical data exists, to develop and implement state-level emissions downscaling methodologies. We will then apply these methodologies to other countries, focusing on Canada (in order to have consistent data over North America) and China, a country with a large geographic extent and large emissions changes over the last few decades. There is a substantial initial data collection effort necessary for this task, including digitizing historical data series (e.g. Smith et al. 2011), however once this data is collected, updates in subsequent years can be implemented with much less effort.

The focus here is to build on existing datasets to allow a more accurate downscaling of emissions trends. The level of detail will depend on data availability and the importance of each source to overall emissions. The first focus will be on fossil fuel emissions in general, where consumption data for the United States, for example, is available starting in 1960 from the EIA state energy data system. We will research historical data sources to examine if, at least for the largest sources and regions, some of this data can be extended further back in time. We expect that, at some point, we will follow the procedures used in previous work and transition to a population-based distribution by state (for example by 1900, as in Smith et al. 2011). We will use spatially detailed emissions data available from the US EPA and EIA for recent years. For some sources, for example metal smelters or other specific process-level emissions, it may not be feasible to extend the data in detail at the state level in the initial construction of the dataset, particularly for earlier years, and we will use appropriate proxies, such as industrial energy consumption, to extend these data to earlier years.[4] Note that, in all cases, we will assure that emissions match total US emissions estimates. We also will focus on provincial level detail for China. The community review and input process within this project may allow sub-national detail to be added for additional countries.

Note that the focus of this effort is "getting emissions in the correct sub-region" to improve global and regional atmospheric chemistry and climate modeling efforts by using data that are readily available. We are not proposing to develop new high-resolution emission and proxy datasets (e.g. Gurney et al. 2009), which requires a more substantial effort to collect high-resolution proxy data that are continuous across jurisdictional boundaries. As new higher resolution emissions data and data proxies are developed these could be incorporated in the future.

3.6 Comparison to Satellite Datasets

In this portion of the project, we will conduct a third level of evaluation: comparison with satellite data. We will use gridded emissions that capture changes at the sub-country level for large countries described above). These data will be used to compare aerosol trends over recent years as modeled in the Community Atmosphere Model version 5 (CAM5) in off-line mode, with satellite aerosol measurements. This will provide additional testing for the emissions dataset and better constrain emission uncertainties.

3.6.1 Comparison Methodology

Satellite data have begun to show promise in examining regional trends in air pollutant concentrations and as a method of validating emission inventories for recent years (Chin et al. 2014). Itahashi et al. (2012), for example, compare MODIS trends in aerosol optical depth (AOD) over China to model simulations over 2000-2005. Using relationships between modeled and observed AOD over this period, they inferred trends for SO_2 emissions over 2006-2010. While this work demonstrates the potential for use of satellite data, this also shows the limits of current approach: the simulations were limited to this time period because more up to date emission inventories at the

[4] As with the historical data system in general, we expect that data coverage will expand over time as researchers contribute to the open-source effort. There is also substantial potential for improvement over time through undergraduate research projects.

necessary level of detail were not available. The impacts of inter-annual variability in meteorology on emissions for 2006-2010 (estimated 3-10% of AOD) could not be directly simulated in that project, nor could any nonlinearities in emission-to-concentration relationships (for example, due to oxidant limitations). Another limitation of using satellite-observed total AOD to constrain anthropogenic emissions is the interference of natural aerosols (such as dust, sea-salt), which could be substantial in some regions. In this project, we will utilize satellite observed total AOD and particle properties.

We will focus on this portion of the project to determine if satellite data can be used to constrain trends in emission estimates over the most recent years in the dataset (where emissions are more uncertain), focusing on trends in aerosol and aerosol precursor compounds. Our goal is to research and test methodologies that could be used on a more regular basis in the future to validate emission inventories and/or better quantify emissions uncertainty.

Our procedure will be to conduct simulations using the CAM5, the atmospheric component of the Community Earth System Model (CESM), in offline mode (Ma et al., 2013; also called specified-dynamics mode), driven with MERRA reanalysis data, along with some recent improvements on global aerosol simulations (H. Wang et al., 2013) and a unique aerosol source tagging capability recently developed for the model (Wang et al., 2014). We will conduct CAM5 simulations from about year 2000 through to the most recent year available from the emissions dataset. This allows overlap with current studies in the literature for comparison, but also extending the evaluation to the most recent full year. We will determine the extent to which we can constrain emission trends using satellite data by (a) comparing model simulated SO_2 and CO with corresponding dataset from satellite observations, and (b) comparing modeled total and component AOD with satellite observed total AOD and particle properties (e.g. particle size and shape, which can be used to derive component AOD for specific aerosol types, such as dust and combustion aerosol).

The emissions project will produce uncertainty estimates for all emissions. Given this information, and the emission-AOD comparisons and relationships conducted by the modeling work in this project, we can explicitly examine how uncertainties in emissions could impact the model-observation comparison. A central outcome of this work is to determine if satellite observations can better constrain emission estimates. We first will examine if it is possible to improve the emission uncertainty estimates, and, if differences between model and observed trends are sufficiently large, we will investigate potential changes to the underlying inventory assumptions.

3.7 Data System Release

The emission data system will be released by JGCRI as open-source software. Included in the system release will be system code, documentation, tutorials, and all input data that is publically available. Some data, such as the IEA energy data are copyrighted and cannot be distributed with the data system. However, once the IEA energy data is purchased by a data system user, it is relatively simple to process the IEA data into a form that can be used by the data system and those instructions will be distributed with the system. Other than this key dataset, we will aim to use only data that are either currently publicly available, or for which we can obtain permission for public release for the portions of data used for this project, so that input data can be released along with the data system as open source software.

The satellite AOD trend data analyzed for this project will be processed into the same spatial units (e.g. countries) as used by the emission data system and will be released along with the emissions data. We will consult with our stakeholder network the exact format that would be most useful (e.g. annual average AOD, monthly average, population weighted). This step will make satellite

AOD data accessible to a wide range of potential users that might otherwise not have ready access to this information.

Emissions at the level of country and sector, and associated gridded emissions from this project will be publically available. Note that the IEA dataset is needed only if a user wants to conduct work directly using the CEDS system. We will investigate if emissions data can also be supplied, at some level, also by fuel and sector to the extent allowed by data agreement that we reach with IEA.

4. Project Timeline

The overall project is organized in three phases: I (initial system implementation and release of CMIP6 interim data product), II (data development and implementation for full system, data system "beta" release, including seasonality and emissions gridding), III (uncertainty analysis, final documentation, data system release). The timing of this project is strongly influenced by the need to produce emissions data for use in the CMIP6 model inter-comparison project. The three phases of the project will overlap in time, but are separately described below to provide a sense of the overall flow of the project.

4.1 Phase I: CMIP6 Data Product (CEDS interim data product for CMIP6)

The first data product from this project will be a historical emissions dataset to be used for historical GCM model experiments as part of CMIP6 and as the starting point for the future emission scenarios from IAMs. The CMIP6 timeline calls for historical data to be available by fall 2015. At a planning meeting in 2014,[5] CMIP6 researchers discussed options and determined that, given the short timeline by which this data is needed, the best path was to use much of the existing historical emissions inventory (Lamarque et al. 2010), while updating emissions for at least the last two decades (through the CEDS project). Emissions gridding for recent years would likely be performed using the same base-year grids and codes as used in the last round, eliminating one potential bottleneck for data production for this phase.

This data product is referred to as the *interim CEDS data product for CMIP6* to emphasize that this project will then proceed to produce an updated emissions dataset, as described in this proposal. The work needed to produce the interim CMIP6 data product is a sub-set of the work needed for the full CEDS emissions data, so this is a natural first step in the project. The work for this interim product will focus on gathering, processing, evaluating, and extending emissions data for recent decades. The last year in the RCP historical dataset (Lamarque et al. 2010) was 2000. Because much of this data was preliminary, we will collect and process data from 1990 to the most recent full year, with the aim of releasing a new dataset for the post-1990 period that is consistent with the RCP historical data (although earlier data will also be updated if this is feasible). This allows us to evaluate how well newer data compare with the previous dataset in 1990. If necessary the new data will be scaled to ensure consistent trends with the previous data, and any such scaling will be documented.

The project timeline is aimed at producing a review release of the new emissions data in summer of 2015. With allowance for several months for community input and review, and modifications based on that review, the interim CEDS data product for CMIP6 would be released in fall 2015 for use by chemistry models that will produce fields needed for CMIP6 beginning January 2016. This allowance for testing and review is one of the "lessons learned" from the previous exer-

[5] Experimental design for CIMP6: Aerosol, land-use, and future scenarios, August 3-8 2014, Aspen Global Change Institute

cise to construct historical emissions data as part of the RCP process (Lamarque et al. 2010). At least one journal paper will be published on the interim data product.

One portion of the *interim CEDS data product for CMIP6* will be a consistent base-year dataset (gridded and by country/sector/fuel) for use by the Integrated Assessment Models that will produce new future scenarios. It is essential to provide data that is identical to that used by GCM models at the grid scale for use in developing future gridded projections. Recent emissions trends are also necessary to assure that future scenarios are consistent with our current knowledge of these trends. Timing for this handoff will be coordinated with the Integrated Assessment Modeling Consortium (IAMC), which organizes the development of future IAM scenarios. Note that while it is critical that gridded emissions estimates for future scenarios are consistent at the grid cell level with historical data, harmonization procedures can be used to blend native IAM emissions with the historical inventory. We expect that the fuel and sector-level detail in the CEDS emission system will be useful also for base-year calibration for IAMs.

4.2 Phase II: Full CEDS Emissions Data Development

Once the interim data product for CMIP6 is in production, the project will begin to shift to performing the additional work needed to produce the full CEDS historical emissions product. Most of this work will be concentrated on developing emissions trends over the 20th century, where large changes in technologies and air pollution controls occurred. The project timeline is constructed to allow participation from the community. We will actively seek out feedback from emission inventory scientists as well as atmospheric modelers through our stakeholders network. In addition to developing consistent emission trends over the 20th century, work in this phase will also address adding additional emission detail such as seasonality and diurnal variation, adding any additional sectoral driver and emissions information, incorporating emissions gridding, and examining the consistency of current data (both over time and between countries).

We anticipate that it will be useful to identify "beta test" users for the data system in this phase of the project. These users can test the system and available documentation and suggest improvements.

4.3 Phase III: Emission Data System Development, Testing, and Release

The major scientific focus of the phase III of the project is the development of emission uncertainty estimates and releasing the first annual emissions update. Research in this phase will focus on assessment of emissions uncertainty as discussed above, including workshops (likely virtual) with emissions inventory experts around the world to discuss uncertainty assumptions. Given the time line for producing the interim CMIP6 data product, it will not be possible to produce comprehensive uncertainty estimates for all emissions at the time of the interim data release. However, the uncertainty analysis will become available for use in interpreting the previously-released interim CMIP6 data product.

We will also release two annual emissions updates, which will use data released over the previous year and any updates to emissions factors or other information that has become available. This data release will also be useful as reference data for the calibration of integrated assessment model simulations. It may also be useful to construct a CMIP6 data continuity release, which is a data release that considers data updates over the previous few years but constructed "as if" all information for years before about 2010-2012 was frozen at the values used to produce the CMIP6 interim data release. This data continuity release can, therefore, be compared directly to the CIMP6 data release in terms of absolute and relative trends.

5. Acknowledgements

The development of this project would not have been possible without the support and discussions with a number of individuals and also through several meetings organized by the Global Emissions Inventory Activity (GEIA). We would like to specifically thank Laura Dawidowski , Greg Frost, Claire Granier, Greet Janssens-Maenhout, Terry Keating, Zbigniew Klimont, Jean-Francois Lamarque, Zifeng Lu, and Min Shao. We gratefully acknowledge funding from US Department of Energy, Office of Science for the overall CEDS project. Funding for comparisons with satellite data, related uncertainty assessment, and sub-regional gridding supplied by the NASA Atmospheric Composition Modeling and Analysis program under grant number NNH15AZ64I. The Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830.

6. References

Ackerman, K.V. and E.T. Sundquist (2008) Comparison of Two U.S. Power-Plant Carbon Dioxide Emission Data Sets *Environ. Sci. Technol.* ***42**, pp*. 5688–5693.

Andres, R. J., Boden, T. A. & Higdon, D. 2014. A new evaluation of the uncertainty associated with CDIAC estimates of fossil fuel carbon dioxide emission. *Tellus B,* 66, 23616, http://dx.doi.org/10.3402/tellusb.v66.23616

Andres R.J., Fielding D.J., Marland, G., Boden, T.A., Kumar, N.: Carbon dioxide emissions from fossil-fuel use, 1751-1950 (1999) Tellus, 51, 759-765.

Blanco et al. Chapter 5: Drivers, Trends and Mitigation. In Climate Change 2014: Mitigation of Climate Change. Working Group III Contribution to the IPCC 5th Assessment Report (http://www.ipcc.ch/).

Bolscher et. al (2007) Emission datasets and methodologies for estimating emissions, available at: http://retro.enes.org, RETRO Report D1-6.

Bond, T. C., Bhardwaj, E., Dong, R., Jogani, R., Jung, S., Roden, C., Streets, D. G. and Trautmann, N. M.: Historical emissions of black and organic carbon aerosol from energy-related combustion, Global Biogeochem Cyc, 21, GB2018, doi:10.1029/2006GB002840, 1850–2000, 2007.

Brioude, J., Angevine, W. M., Ahmadov, R., Kim, S.-W., Evan, S., McKeen, S. A., Hsie, E.-Y., Frost, G. J., Neuman, J. A., Pollack, I. B., Peischl, J., Ryerson, T. B., Holloway, J., Brown, S. S., Nowak, J. B., Roberts, J. M., Wofsy, S. C., Santoni, G. W., Oda, T., and Trainer, M.: Top-down estimate of surface flux in the Los Angeles Basin using a mesoscale inverse modeling technique: assessing anthropogenic emissions of CO, NOx and CO2 and their impacts, Atmos. Chem. Phys., 13, 3661-3677, doi:10.5194/acp-13-3661-2013, 2013.

Bond, T.C., Streets, D.G. Yarber, K.F., Nelson, S.M., Woo, J.-H., and Klimont, Z.: A technology-based global inventory of black and organic carbon emissions from combustion, J. Geophys. Res., 109, D14203, doi:10.1029/2003JD003697, D14203, 2004.

Chin, M., Diehl, T., Tan, Q., Prospero, J. M., Kahn, R. A., Remer, L. A., Yu, H., Sayer, A. M., Bian, H., Geogdzhayev, I. V., Holben, B. N., Howell, S. G., Huebert, B. J., Hsu, N. C., Kim, D., Kucsera, T. L., Levy, R. C., Mishchenko, M. I., Pan, X., Quinn, P. K., Schuster, G. L., Streets, D. G., Strode, S. A., Torres, O., and Zhao, X.-P. (2014) Multi-decadal aerosol variations from 1980 to 2009: a perspective from observations and a global model, Atmos. Chem. Phys., 14, 3657-3690, doi:10.5194/acp-14-3657-2014.

Eyring, V., Isaksen, I. S. A., Berntsen, T., Collins, W. J., Corbett, J. J., Endresen, O., Grainger, R. G., Moldanova, J., Schlager, H., and Stevenson, D. S. (2010) Transport impacts on atmosphere and climate: Shipping, Atm. Env., 44, 4735-4771.

Fernandes, S. D., Trautmann, N. M., Streets, D. G., Roden, C. A., and Bond, T. C. (2007) Global biofuel use, 1850–2000, Global Biogeochem. Cyc, 21, GB2019, doi:10.1029/2006GB002836.

Frost GJ, S Falke, C Granier, T Keating, JF Lamarque, M Melamed, P Middleton, G Petron, and SJ Smith (2012) New Directions: Toward a community emissions approach *Atmospheric Environment* **51** 333–334.

Granier, C., G. Frost, J.-F. Lamarque (2014) Workshop on the Development of a Community Historical Emission Inventory, IGAC News, **51**, pp 9.

Giglio, L., J. T. Randerson, and G. R. van der Werf (2013) Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4) Volume 118, Issue 1, pages 317–328.

Gurney, K., D. Mendoza, Y. Zhou, M. Fischer, C. Miller, S. Geethakumar, and S. de la Rue du Can (2009). High resolution fossil fuel combustion CO_2 emission fluxes for the United States. *Environmental Science & Technology*, 43 (14): 5535-5541.

Holben, B. N., et al., 1998. AERONET-A federated instrument network and data archive for aerosol characterization. Remote Sens. Environ., 66(1), 1–16.

Huang, Y., Shen, H., Chen, H., Wang, R., Zhang, Y., Su, S., Chen, Y., Lin, N., Zhuo, S., Zhong, Q., Wang, X., Liu, J., Li, B., Liu, W. & Tao, S. 2014. Quantification of Global Primary Emissions of PM2.5, PM10, and TSP from Combustion and Industrial Process Sources. *Environmental Science & Technology*, 10.1021/es503696k.

Huneeus, N., Boucher, O., and Chevallier, F. (2013) Atmospheric inversion of SO_2 and primary aerosol emissions for the year 2010, Atmos. Chem. Phys., 13, 6555-6573, doi:10.5194/acp-13-6555-2013.

Itahashi, S., Uno, I., Yumimoto, K., Irie, H., Osada, K., Ogata, K., Fukushima, H., Wang, Z., and Ohara, T. (2012) Interannual variation in the fine-mode MODIS aerosol optical depth and its relationship to the changes in sulfur dioxide emissions in China between 2000 and 2010, Atmos. Chem. Phys., 12, 2631-2640, doi:10.5194/acp-12-2631-2012.

Janssens-Maenhout, G., F. Dentener, J. van Aardenne, S. Monni, V. Pagliari, L. Orlandini, Z. Klimont, J. Kurokawa, H. Akimoto, T. Ohara, R. Wankmüller, B. Battye, D. Grano, A. Zuber, and T. Keating (2013) EDGAR-HTAP: a harmonized gridded air pollution emission dataset based on national inventories (European Commission, Joint Research Centre ISBN 978-92-79-23123-0).

Joerss, W. (3013) Determination of the uncertainties of the German emission inventories for particulate matter and aerosol precursors using Monte-Carlo analysis. Climatic Change. doi: 10.1007/s10584-013-1028-5

Klimont, Z, S J Smith and J Cofala (2013) The last decade of global anthropogenic sulfur dioxide: 2000-2011 emissions *Environmental Research Letters* **8** 014003. doi:10.1088/1748-9326/8/1/014003.

Kuenen, J. J. P., Visschedijk, A. J. H., Jozwicka, M. & Denier Van Der Gon, H. a. C. 2014. TNO-MACC_II emission inventory; a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling. *Atmos. Chem. Phys.,* 14**,** 10963-10976, 10.5194/acp-14-10963-2014.

Kurokawa, J., Ohara, T., Morikawa, T., Hanayama, S., Janssens-Maenhout, G., Fukui, T., Kawashima, K., and Akimoto, H.: Emissions of air pollutants and greenhouse gases over Asian regions during 2000–2008: Regional Emission inventory in ASia (REAS) version 2, Atmos. Chem. Phys., 13, 11019-11058, doi:10.5194/acp-13-11019-2013, 2013.

Lamarque, J. F; Bond, Tami C; Eyring, Veronika; Granier, Claire; Heil, Angelika; Klimont, Z; Lee, David S; Liousse, Catherine; Mieville, Aude; Owen, Bethan; Schultz, Martin; Shindell, Drew; Smith, Steven J; Stehfest, Eike; van Aardenne, John; Cooper, Owen; Kainuma, M; Mahowald, Natalie; McConnell, J.R.; Riahi, Keywan; Van Vuuren, Detlef (2010) Historical (1850-2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application *Atmospheric Chemistry and Physics* **10** pp. 7017–7039. doi:10.5194/acp-10-7017-2010

Lee, D.S., G. Pitari, V. Grewe, K. Gierens, J.E. Penner, A. Petzold, M.J. Prather, U. Schumann, A. Bais, T. Berntsen, D. Iachetti, L.L. Lim, R. Sausen (2010) Transport impacts on atmosphere and climate: Aviation. Atmos. Env. **44** (37) pp. 4678–4734.

Levy, R., S. Mattoo, L.A. Munchak, L.A. Remer, A.M. Sayer, Patadia, F., and N. Hsu (2013). The Collection 6 MODIS Aerosol Products over Land and Ocean *Atmos. Meas. Tech*, *6*, 2989-3034
doi:10.5194/amt-6-2989-2013

Li, C., J. Joiner, N. A. Krotkov, and P. K. Bhartia (2013), A fast and sensitive new satellite SO_2 retrieval algorithm based on principal component analysis: Application to the ozone monitoring instrument, *Geophys. Res. Lett.*, **40**, doi:10.1002/2013GL058134.

Liu, X., Easter, R. C., Ghan, S. J., Zaveri, R., Rasch, P., Shi, X., Lamarque, J.-F., Gettelman, A., Morrison, H., Vitt, F., Conley, A., Park, S., Neale, R., Hannay, C., Ekman, A. M. L., Hess, P., Mahowald, N., Collins, W., Iacono, M. J., Bretherton, C. S., Flanner, M. G., and Mitchell, D.: Toward a minimal representation of aerosols in climate models: description and evaluation in the Community Atmosphere Model CAM5, Geosci. Model Dev., 5, 709-739, doi:10.5194/gmd-5-709-2012, 2012.

Lu, Z., Zhang, Q., and Streets, D. G.: Sulfur dioxide and primary carbonaceous aerosol emissions in China and India, 1996–2010, Atmos. Chem. Phys., 11, 9839-9864, doi:10.5194/acp-11-9839-2011, 2011

Ma, P.-L., Rasch, P. J., Wang, H., Zhang, K., Easter, R. C., Tilmes, S., Fast, J. D., Liu, X., Yoon, J.-H., Lamarque, J.-F., 2013. The Role of Circulation Features on Black Carbon Transport into the Arctic in the Community Atmosphere Model Version 5 (CAM5). Journal of Geophysical Research. D. (Atmospheres) 118(10), 4657-4669.

Mieville, A., Granier, C., Liousse, C., B. Guillaume, Mouillot, F., Lamarque, J.-F.,. Grégoire, J.-M, and Pétron, G.: Emissions of gases and particles from biomass burning using satellite data and an historical reconstruction, Atmos Environ, 10.1016/j.atmosenv.2010.01.011, 2010.

Peischl, J., Ryerson, T. B., Brioude, J., et al. 2013. Quantifying sources of methane using light alkanes in the Los Angeles basin, California. Journal of Geophysical Research-Atmospheres, 118, 4974-4990. doi:10.1002/jgrd.50413

Pollack, I. B., Ryerson, T. B., Trainer, M., et al. 2012. Airborne and ground-based observations of a weekend effect in ozone, precursors, and oxidation products in the California South Coast Air Basin. Journal of Geophysical Research-Atmospheres, 117.

Pulles, T., M. van het Bolscher, R. Brand, and A. Visschedijk (2007), Assessment of global emissions from fuel combustion in the final decades of the 20th Century, TNO Rep. 2007-A-R0132B, Ned. Org. voor toegepast Natuurwet. Onderzoek, Apeldoorn, Netherlands.

Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R.D., Lucchesi, R., Molod, A., 2011. MERRA - NASA's Modern-Era Retrospective Analysis for Research and Applications. Journal of Climate 24, 3624-3648.

Schultz, M. and Rast, S.: Emission datasets and methodologies for estimating emissions, available at: http://retro.enes.org (last ac- cess: 16 February 2010), RETRO Report D1-6, 2007.

Schultz, M.G., Heil, A., Hoelzemann, J.J., Spessa, A., Thonicke, K., Goldammer, J., Held, A.C., Pereira, J.M., van het Bolscher, M.: Global Wildland Fire Emissions from 1960 to 2000, Global Biogeochem. Cyc, 22, GB2002,doi:10.1029/2007GB003031, 2008.

Smith SJ and Bond, T. C. (2014) Two hundred fifty years of aerosols and climate: the end of the age of aerosols, *Atmos. Chem. Phys.* **14** 537–549. doi:10.5194/acp-14-537-2014.

Smith, SJ, J van Aardenne, Z Klimont, R Andres, AC Volke, and S Delgado Arias (2011) Anthropogenic Sulfur Dioxide Emissions: 1850-2005 *Atmos. Chem. Phys.*, **11**, 1101–1116.

Stohl, A., Klimont, Z., Eckhardt, S., Kupiainen, K., Shevchenko, V. P., Kopeikin, V. M. & Novigatsky, A. N. 2013. Black carbon in the Arctic: the underestimated role of gas flaring and residential combustion emissions. *Atmospheric Chemistry and Physics,* 13**,** 8833-8855.

Tanré, D., Bréon, F. M., Deuzé, J. L., Dubovik, O., Ducos, F., François, P., Goloub, P., Herman, M., Lifermann, A., Waquet, F., 2011. Remote sensing of aerosols by using polarized, directional and spectral measurements within the A-Train: the PARASOL mission. Atmos. Meas. Tech., 4, 1383-1395, doi:10.5194/amt-4-1383-2011.

van Aardenne, J., Cooper, O., Kainuma, M., Mahowald, N., McConnell, J.R., Riahi, K., Van Vuuren, D. (2010) Historical (1850-2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application, Atmospheric Chemistry and Physics, 10, 7017–7039.

Van der Werf, G.R., Randerson, J.T., Giglio, L., Collatz, G.J., Kasibhatla, P.S., Collatz, G.J., Kasibhatla, P.S., Arellano, A.F.: Interannual variability in global biomass burning emissions from 1997 to 2004, Atmos Chem Phys, 6, 3423-3441, 2006.

Wang, R., Tao, S., Shen, H., Wang, X., Li, B., Shen, G., Wang, B., Li, W., Liu, X., Huang, Y., Zhang, Y., Lu, Y. & Ouyang, H. (2011). Global Emission of Black Carbon from Motor Vehicles from 1960 to 2006. *Environmental Science & Technology,* 46**,** 1278-1284, 10.1021/es2032218.

Wang, Rong, Shu Tao, Wentao Wang, Junfeng Liu, Huizhong Shen, Guofeng Shen, Bin Wang, Xiaopeng Liu, Wei Li, Ye Huang, Yanyan Zhang, Yan Lu, Han Chen, Yuanchen Chen, Chen Wang, Dan Zhu, Xilong Wang, Bengang Li, Wenxin Liu, and Jianmin Ma (2012) Black Carbon Emissions in China from 1949 to 2050 *Environmental Science & Technology 46* (14), 7595-7603.

Wang, R., Tao, S., Ciais, P., Shen, H. Z., Huang, Y., Chen, H., Shen, G. F., Wang, B., Li, W., Zhang, Y. Y., Lu, Y., Zhu, D., Chen, Y. C., Liu, X. P., Wang, W. T., Wang, X. L., Liu, W. X., Li, B. G. & Piao, S. L. 2013. High-resolution mapping of combustion processes and implications for CO2 emissions. *Atmos. Chem. Phys.,* 13**,** 5189-5203, 10.5194/acp-13-5189-2013.

Wang, H., Easter, R., Rasch, P., Wang, M., Liu, X., Ghan, S., Qian, Y., Yoon, J.-H., Ma, P.-L., and Vinoj, V., 2013. Sensitivity of Remote Aerosol Distributions to Representation of Cloud-Aerosol Interactions in a Global Climate Model. Geoscientific Model Development 6, 765–782.

Wang, H., P. J. Rasch, R. C. Easter, B. Singh, R. Zhang, P.-L. Ma, Y. Qian, S. J. Ghan, and N. Beagley (2014), Using an explicit emission tagging method in global modeling of source-receptor relationships for black carbon in the Arctic: Variations, sources, and transport pathways, J. Geophys. Res. Atmos., 119, 12,888–12,909, doi:10.1002/2014JD022297.

Winiwarter, W. & B. Muik (2010) Statistical dependence in input data of national greenhouse gas inventories: effects on the overall inventory uncertainty *Climatic Change* 103:19–36 DOI 10.1007/s10584-010-9921-7

Yan, F., E. Winijkul, S. Jung, T.C. Bond, and D.G. Streets (2011) Global emission projections of particulate matter (PM): I. Exhaust emissions from on-road vehicles. *Atmospheric Environment*, 45, 28, 4830-4844, doi:10.1016/j.atmosenv.2011.06.018.

Yu, H., M. Chin, L.A. Remer, R.G. Kleidman, N. Bellouin, H. Bian, and T. Diehl, Variability of marine aerosol fine-mode fraction and estimates of anthropogenic aerosol component over cloud-free oceans from MODIS, Journal of Geophysical Research – Atmospheres, 114, D10206, doi:10.1029/2008JD010648, 2009.

Zhao, Y., Zhang, J., and Nielsen, C. P.: The effects of recent control policies on trends in emissions of anthropogenic atmospheric pollutants and CO_2 in China, Atmos. Chem. Phys., 13, 487-508, doi:10.5194/acp-13-487-2013, 2013.