The Community Emissions Data System (CEDS)

RACHEL HOESLY
STEVEN J. SMITH

Joint Global Change Research Institute
College Park, MD

On behalf of the CEDS project team:

Leyang Feng	PREVIOUS	*Tyler Pitkanen*
Benjamin Goldstein	*Ryan Bolt*	*Patrick O'Rourke*
Rohit Mandavia	*Han Chen*	*Jonathan Seibert*
Presley Muwan	*Grace Duke*	*Linh Vu*
	Cecilia Moura	*Nazar Kholod*

And collaborators:

Robert Andres	Jun-ichi Kurokawa	Zifeng Lu
Tami Bond	Meng Li	Qiang Zhang
Laura Dawidowski	Liang Liu	
Zbigniew Klimont	Greet Maenhout	

EPA International Emissions Inventory Conference 2017

Baltimore, MD

Tools and GIS Session

August 17, 2017

Outline

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

1) CEDS Project Work To Date

- *Overview and Goals*
- *Methods and System Structure*
- *CMIP6 Data Product*

2) Upcoming Work (next ~12 months)

- *Open Source Release*
- *Sub-regional emission estimates*
- *Uncertainty Estimates:*
  - *Recent Years and Comprehensive Uncertainty Estimates*

Funding for this research Provided by the US Department of Energy office of Science

additional support:

The National Atmospheric and Space Administration's Atmospheric Composition: Modeling and Analysis Program (ACMAP)

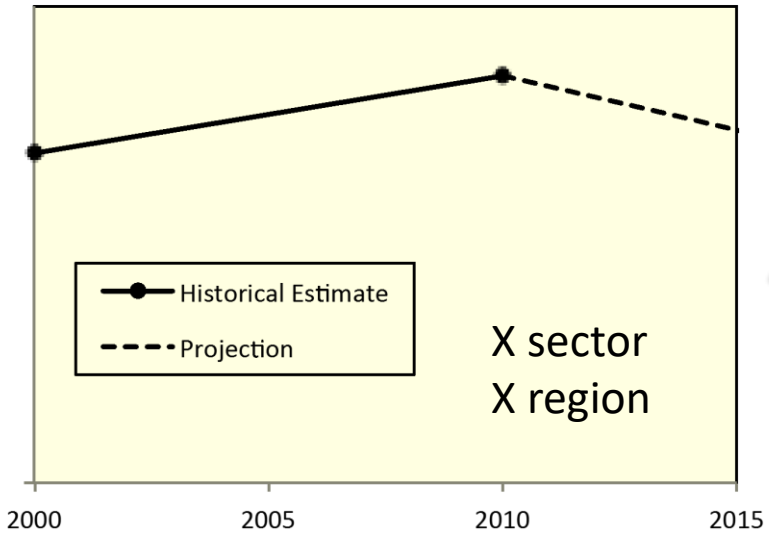
Community Emissions Data System

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Timely "research" estimates for anthropogenic emissions of aerosol (BC, OC) and aerosol precursor compounds (SO_2, NO_x, NH_3, CH_4, CO, NMVOC): key for aerosol research and Earth System Models. Also CO_2 and CH_4 (CH_4 not the full time series).

Needed for historical and future simulations, validation/comparisons with observations, historical attribution, uncertainty quantification, IAM calibration and validation, and economic/policy analysis.

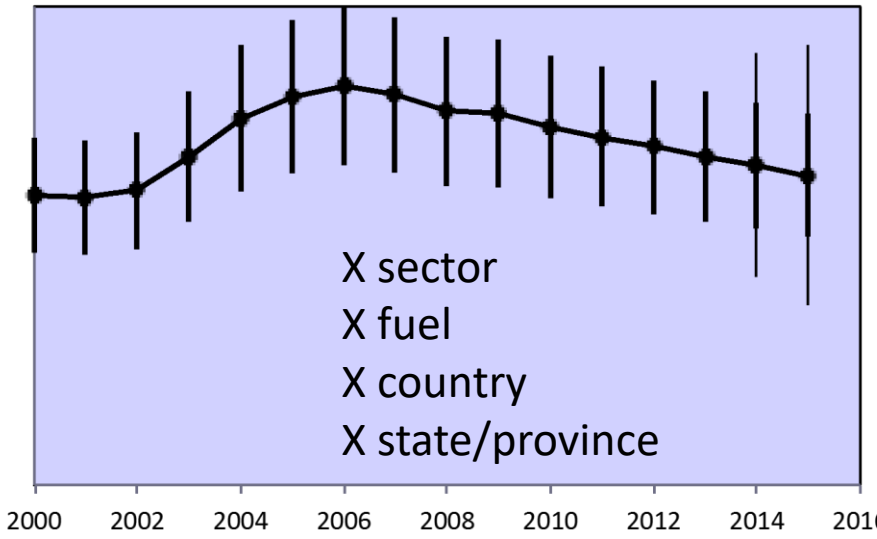
Instead of this



Produced using an open-source data system to increase data transparency and facilitate research advancements.

Produce

Uncertainty essential for estimates of more recent years.



Methodology

Approach: Hybrid of bottom-up emissions & inventory

- Default estimate (**GAINS** emission factors, **EDGAR** emissions, etc.)
- Calibrate to country-level inventories at the sectoral level where available and reliable. Similar to the approach for RCP/CMIP5 data & EDGAR-HTAP, except consistent activity data is also used in CEDS.

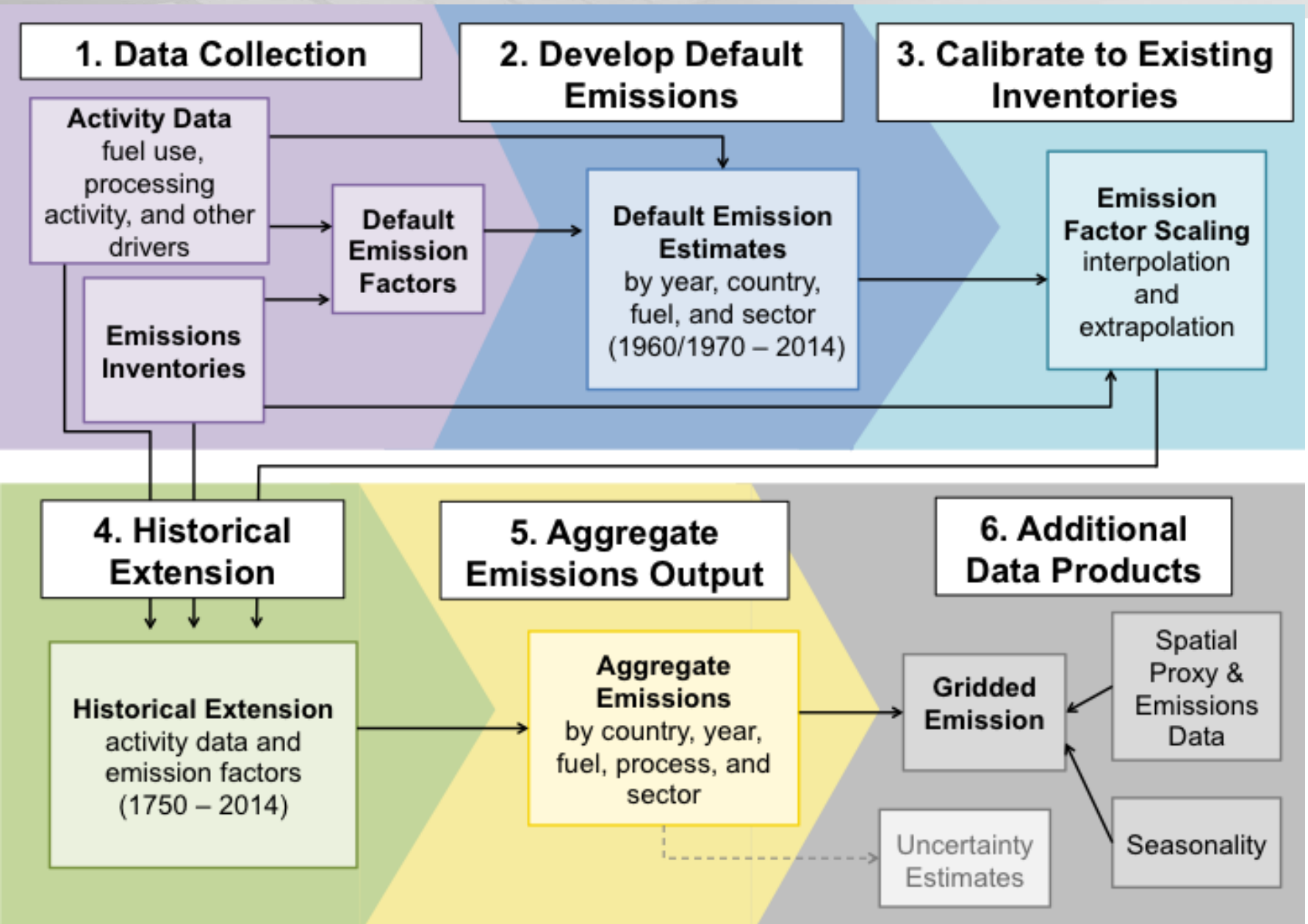
Produce "a" best estimate, not a fully independent estimate

Country-Level Inventory Data

- European Countries: Country data as reported to EMEP
- New Zealand, Belarus: UNFCCC reported data
- USA: US EPA Trends and GHG Inventory
- Canada: Environment Canada
- Argentina: Country submission to UNFCCC
- China: MEIC Inventory (2008, 2010, 2012)
- South Korea: http://airemiss.nier.go.kr/
- Japan: REAS historical inventory (preliminary updated)
- Australia: National Emissions Inventory
- Taiwan: National Emissions Inventory
- Other Asia: REAS 2.1 historical inventory

Plus many additional sector and country-specific data sources

System Diagram



Code Architecture

- Series of **R modules** (as individual code files), with common utility functions in header files
- Makefile system
- Most code files are agnostic to emissions species
  - Some emissions and emission factor information is species-specific.
- Output files labeled with emission species, so **parallel processing** by species is readily facilitated
- Modules currently pass data through plain-text csv files
  - Easy access to intermediate results
  - Maintains flexibility to insert modules
  - Separate diagnostic file outputs
- Using **gitHub** for version control and system release (code and most input data)
  - Integrates version control and wiki-based documentation

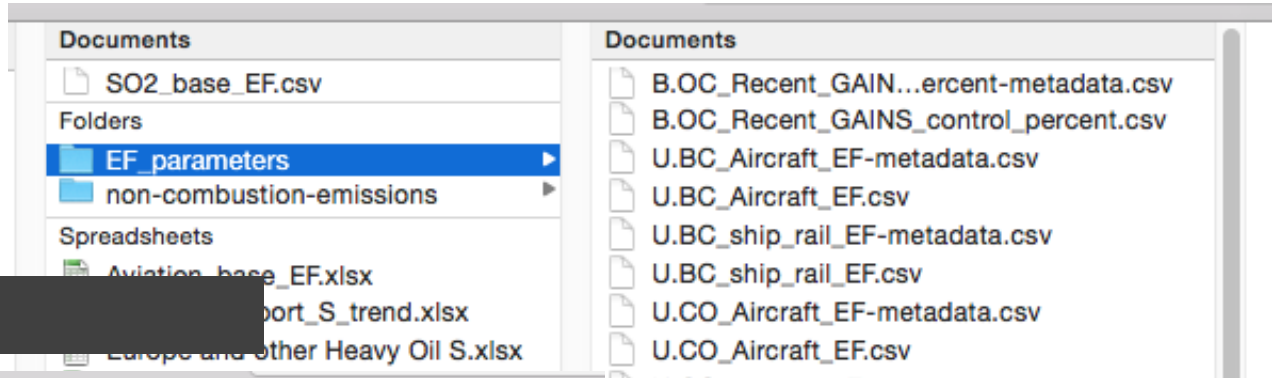
Automation and User Options

Predominantly automated processing, extending, checking.

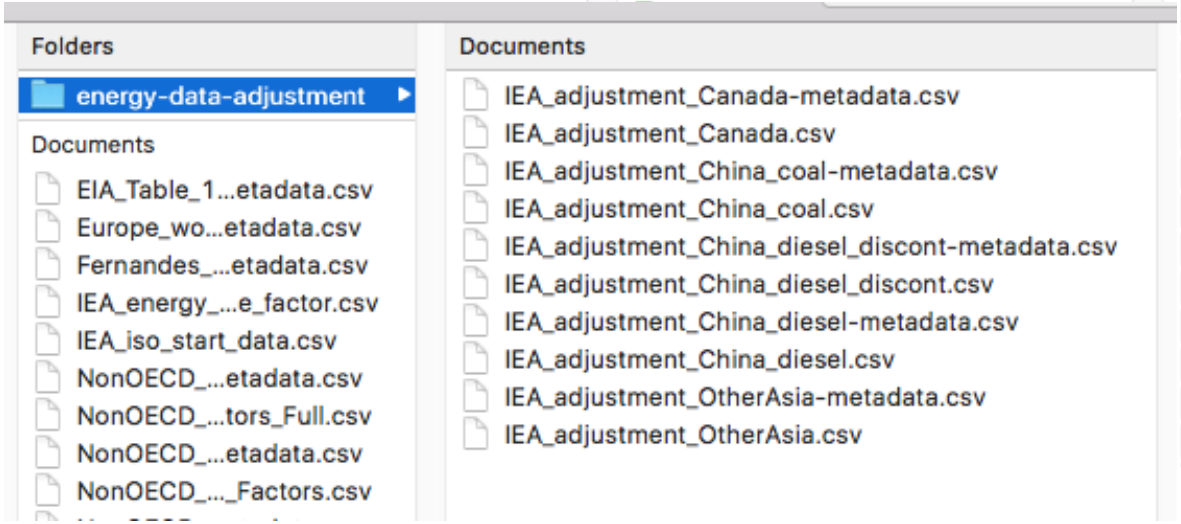
Assumptions and methods can be changed WITHOUT coding

- User drop folders

Default Emission Factor Parameters



Energy Data Adjustments



Sectoral Working Resolution

Operate at an intermediate resolution, by sector and fuel (L,M & H oil; Nat Gas; Hard Coal, Brown Coal, Coke, Biomass)

1A1a_Electricity-public
1A1a_Electricity-autoproducer
1A1a_Heat-production
1A2a_Ind-Comb-Iron-steel
1A2b_Ind-Comb-Non-ferrous-metals
1A2c_Ind-Comb-Chemicals
1A2d_Ind-Comb-Pulp-paper

- At this level, have combustion time series available from IEA
- This detail particularly useful for aggregating accurately to calibrate to inventory data releases

While important sub-sectoral trends **(LDV vs HDV, etc.)*:***

1A3b_Road
1A4a_Commercial-institutional
1A4b_Residential
1A4c_Agriculture-forestry-fishing

- No global driver data over time
- Rely on detailed inventories to capture sub-sectoral trends
- Some details (e.g., non-road mobile) not present in all inventories

1B1_Fugitive-solid-fuels
1B2_Fugitive-petr-and-gas

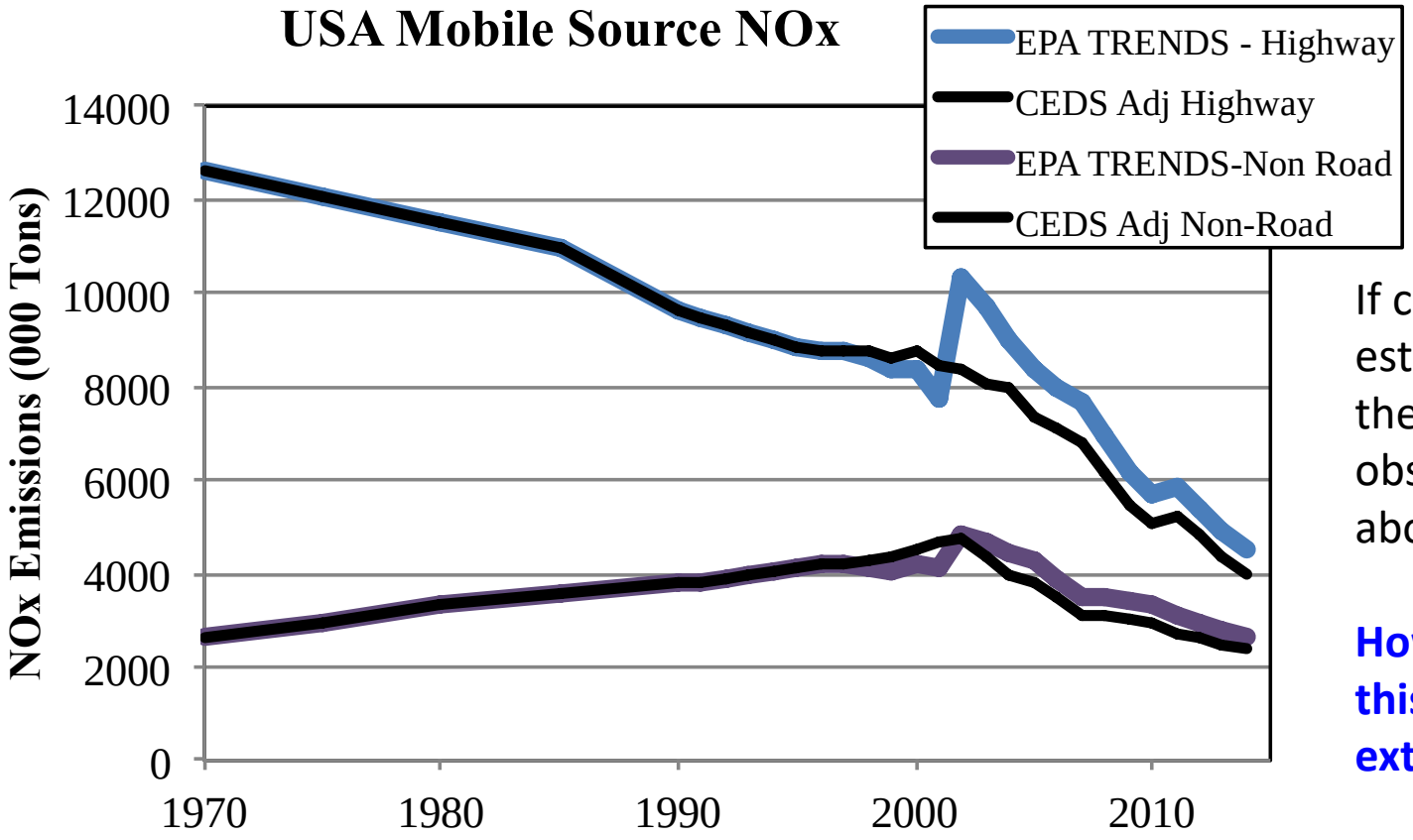
- Would like to break out some important emerging sectors to capture recent trends.
- However, inventories do not always report at this level!

Inconsistencies in Inventories

US NOx estimates

EPA TRENDS estimates have a discontinuity due to a methodology change in 2001. This is not physical, so we had to smooth this out.

Our choice motivated by observational studies indicate that NEI 2011 road NO_x estimates may be overestimated by 50%. (Anderson et al. 2014, Travis et al. 2016.)



If current emission estimates are too high, then what can observations tell us about the trend?

How far back would this overestimate extend?

CMIP6 – Historical Anthropogenic Emissions

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

STATUS

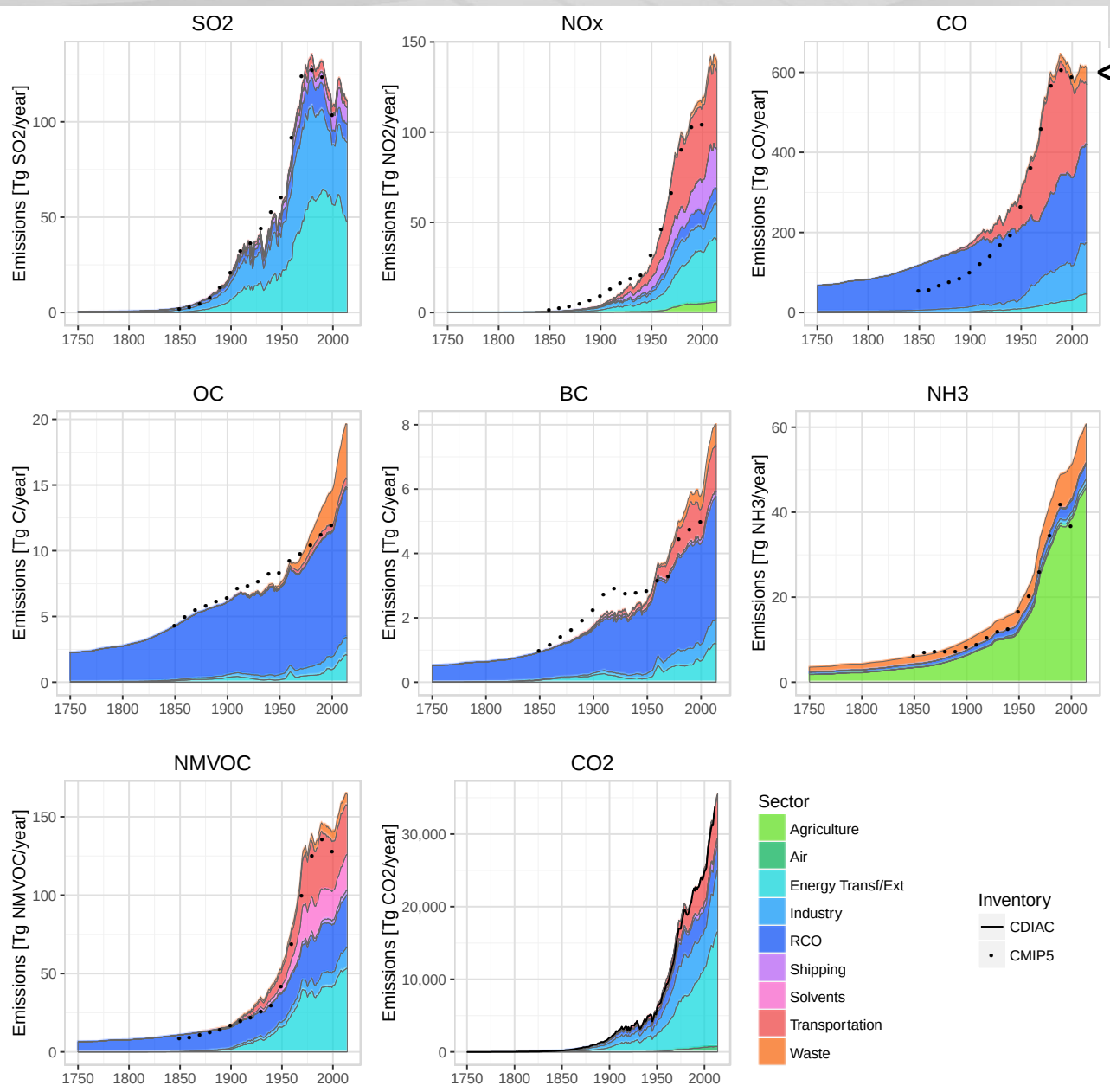
- Gridded emissions on Earth System Grid Federation (ESGF) with meta data:
  :activity_id = “input4MIPs”;
  :dataset_category = “emissions”;
  :realm = “atmos”;
  :institute = “PNNL-JGCRI”
  :product = “primary-emissions-data”
- Manuscript submitted to GMD (CMIP6 Special Issue) and undergoing revisions.

Key Features

- Annual estimates, 1750 – 2014
- CO, NO_X, SO_2, OC, BC, NH_3, NMVOC, CO_2 and CH_4
- All emissions estimated at the country level
- Mapped to 0.5 degree grids in 9 sectors for CMIP6
- Monthly Seasonality for gridded data in all sectors
- NMVOC Speciation

Original Data release
1750-1850: June 2016
1850-2014: July 2016

Global CMIP6 Emissions (& vs CMIP5)

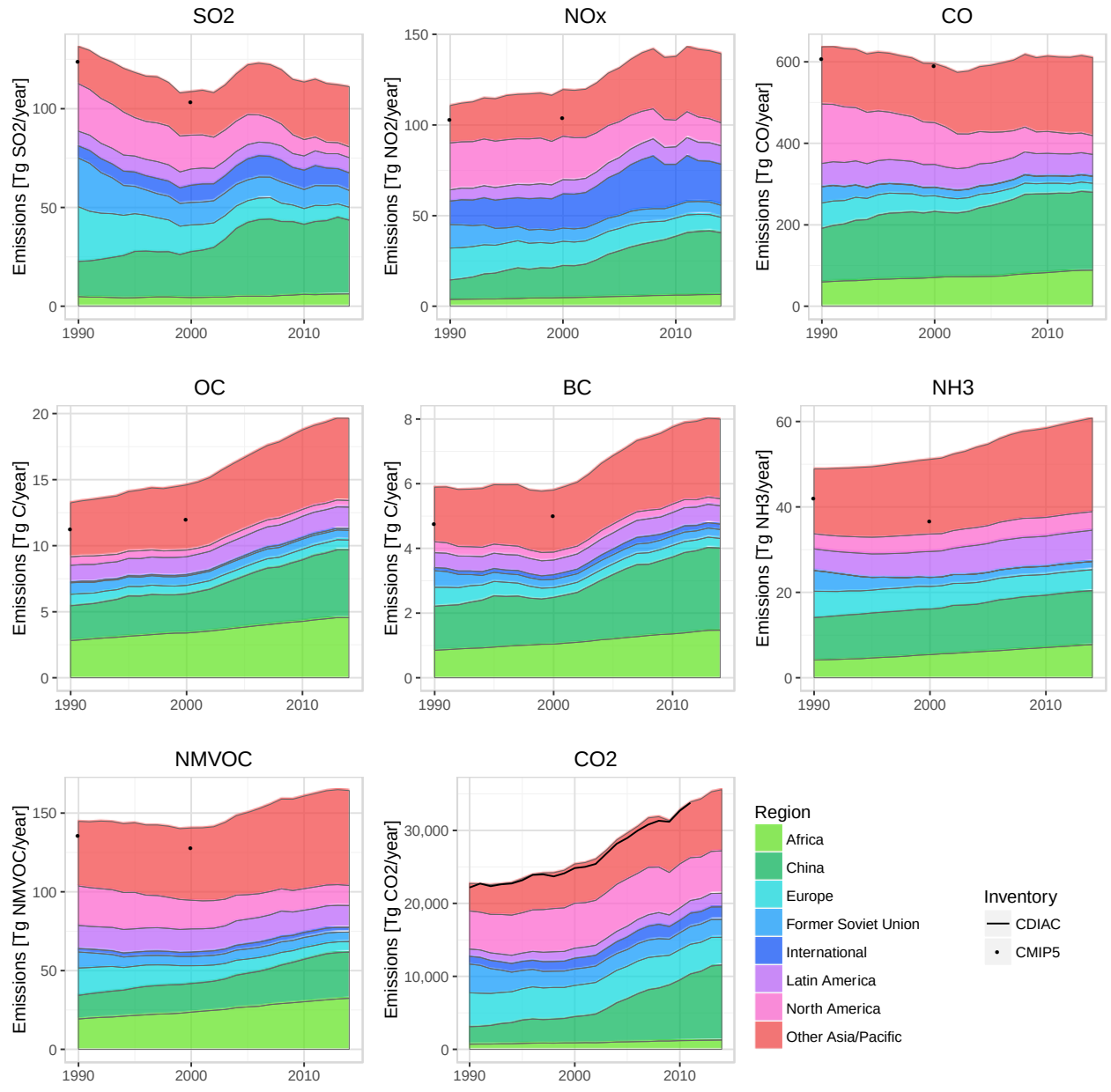


<- CEDS by sector & CMIP5 totals

- Residential biomass are the dominant emissions in early years, e.g. 1850. (Except NH3, where manure emissions are dominant, and NO_X , which is relatively small at this point.)
- Residential biomass has large contribution to BC and CO even to current day
- Transportation sector has large contribution to recent trends for NO_X and CO

Like with like comparison does not include aviation or agricultural waste burning.
** CO2 comparison includes aviation*

Recent Trends in Global CMIP6 Emissions



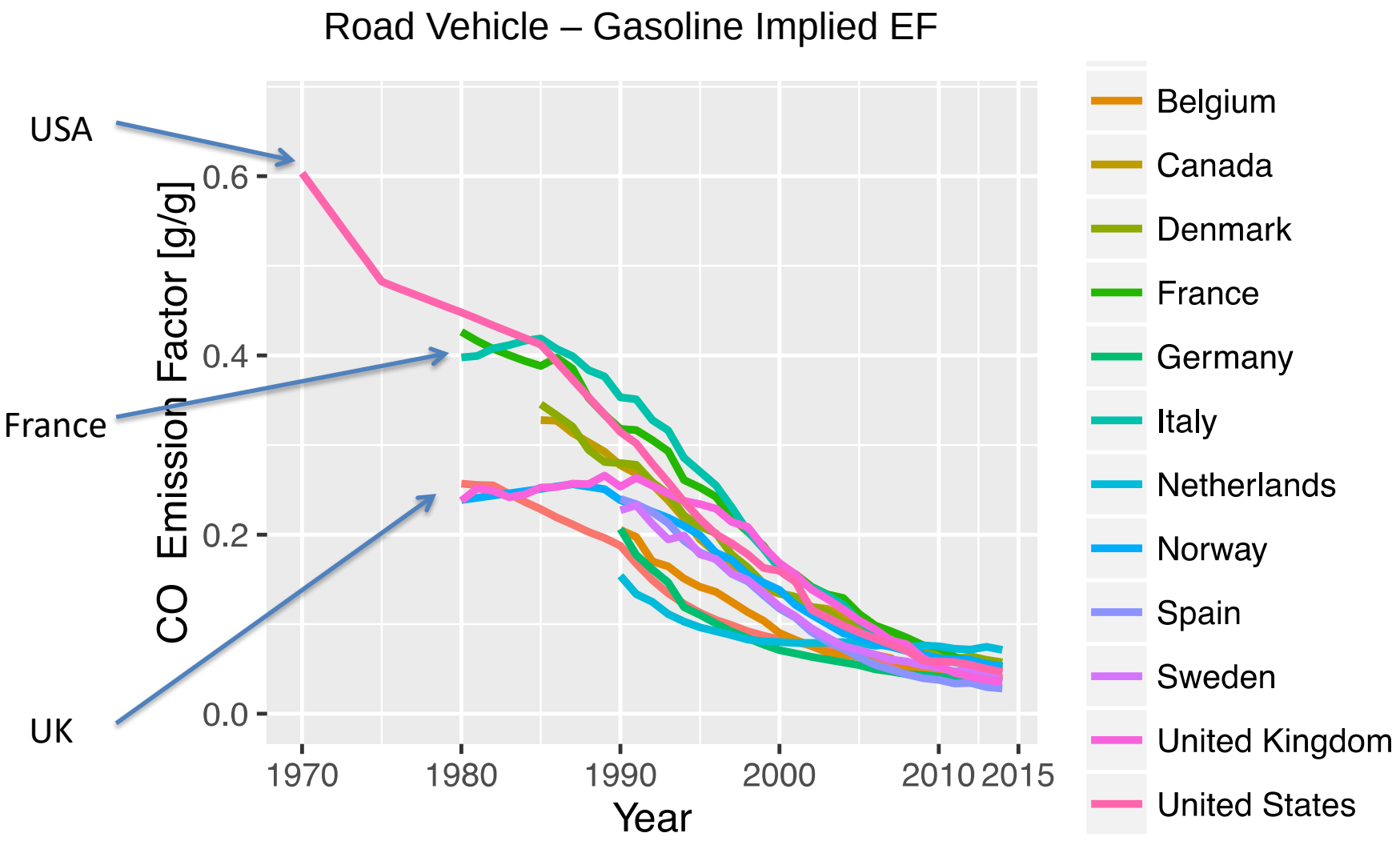
- Growth in BC and OC in Asia and Africa from residential biomass. BC growth in China due to energy transformation
- Declining NO_X and CO in North America and Europe is offset by growing NO_X and CO in Asia and China
- Recent trends are more uncertain – they rely on new activity data and are not scaled to inventories.
- Notable trends are from uncertain sources: population driven sectors (waste, agriculture), BC/OC, and China

Like with like comparison does not include aviation or agricultural waste burning.
** CO2 comparison includes aviation*

Implied Emission Factors

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Because we are consistently processing inventory and driver data, we can examine implied emission factors.



Uncertainty Estimates

Uncertainty For Most Recent Years (~ next 4 months)

It is critical that emissions for recent years are coupled with uncertainty estimates

- The additional uncertainty in the most recent years can be rigorously assessed by applying the extension methodologies to past data

  Although "past uncertainty does not guarantee future uncertainty"

Comprehensive Uncertainty Estimates (~next year)

All bottom-up emission uncertainty estimates contain a substantial element of expert judgment

- Guide assumptions with literature (including observational studies), comparisons between inventories, and comparisons within CEDS
- Reduce dimensionality by a "tiered" approach to group assumptions

  Otherwise: ~40 sectors X 200+ countries X 5 fuels X ~10 emissions
- Consider correlations across sectors and countries (spatially)
- Result: **uncertainty ensembles** consistent across species, region, time

Sub Regional Estimates

Downscaling Estimates for the US, then China (~ next year)

Goal - Better informed spatial emissions data to incorporate into Gridded Data Products

- Currently – aggregate US estimates are spatially distributed using proxy data (EDGAR grids)
- Goal – State level emission time series
  - Capture historical spatial emission shifts
  - Capture recent trends at the state level (in between NEI editions)
- State level data sources
  - EIA State Energy Data System (Driver data)
  - US NEI (Emissions)
- Similar Methodology – use existing data – not an independent estimate
  - Historical driver data
  - Scale emissions/emission factors to state level

Engagement

Feedback and participation has been facilitated through several means

- http://www.globalchange.umd.edu/CEDS/
- A listserv for updates
  - Send an email to listserv@listserv.umd.edu with the email body: “subscribe cedsinfo”. (You will get a return e-mail asking you to verify your subscription.)

The data system and input data will be released as open source software

- Scheduled release: Early Fall 2017
- Update to 2016: Early Fall 2017
- Including capability of producing gridded emission datasets

  You will have to purchase a license for the IEA energy data

  (Although we are exploring a fully open source option that would cover the post-1990 period.)

END

Additional Slides

Comparisons to Observations

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Trends in Observed and Inventory NO_X/CO emission ratios

City/Country	Years	Observed*	MAACity*	CEDS (road)	CEDS (total)
USA (various cities)	1989-2013	4.1	1.45	3.86	2.37
UK (London)	1989-2015	7.2	1.88	6.90	5.90
France (Paris)	1995-2014	8.8	1.59	7.47	3.39
Values shown in log linear trends in units of $\%yr^{-1}$					*(Hassler et al., 2016)

Observed and Inventory BC/CO emission ratios

Country	Observed* 2009 - 2015	CEDS 2009 - 2014	CEDS 2008	REAS2.1 2008
Japan	5.9 ± 3.4	9.7	9.5	6.5
Korea	6.7 ± 3.7	89.8	82.3	23
China (North East)	6.0 ± 2.8	14.3	12.8	
China (North Central East)	5.3 ± 2.1			8.3
China (South Central East)	6.4 ± 2.2			9.9
China (South)	6.9 ± 1.2			
Values shown in $ngm^{-3}ppb^{-1}$				* (Kanaya et al., 2016)

Hassler et al.: Analysis of long-term observations of NOx and CO in megacities and application to constraining emissions inventories, Geophys. Res. Lett., 43(18), 2016GL069894, doi:10.1002/2016GL069894, 2016.
Kanaya et al.: Long-term observations of black carbon mass concentrations at Fukue Island, western Japan, during 2009–2015: constraining wet removal rates and emission strengths from East Asia, Atmospheric Chem. Phys., 16(16), 10689–10705, doi:10.5194/acp-16-10689-2016, 2016.

CEDS CMIP6 – Improvements Relative To CMIP5

- More robust emission trends
  - Consistent methodology across all years
  - All emission species use same driver data
  - Consistently calibrated to country-level inventories where available
- Annual data resolves important socio-economic events
- Emissions estimates out to 2014 to capture recent trends as best as possible
  - *Albeit with additional uncertainty (which is now being estimated)*
- 1850 Emissions – CO higher and NO_X lower, due to explicit representation of biomass and coal emissions for all species.
- New sectors included
  - Residential waste burning
  - Flaring (from ECLIPSE project)
  - Fossil-fuel Fires (from EDGAR)
- Reproducible emissions generation process
  - CEDS data system and most input data will be released as open source software
  - Updated data such as new country inventories and energy driver data can be readily incorporated to allow annual updates
  - Modular system facilitates data updates (e.g., “drag and drop” capability)

CMIP6 – Historical Anthropogenic Emissions

Issues and Limitations

- Extrapolation before 1970/1960 can be improved
  - Collection of additional driver data and consideration of changes in emission factors
  - Focus on residential sector, however, captures major transitions over this time
- Process emissions
  - Some process sectors aggregated together: e.g., 1A1bc other transformation, 2C_metal
  - Time series would be improved by using additional sector-specific driver data
- A number of gridding proxies are static over time
  - Residential (and related) emissions distributed using population distribution, which changes over time. These are dominant in earlier years, so much of the major shifts in spatial distribution with a country is being captured.
  - Other sectors have mix of proxies, few of which are newer than 2010, and most were kept static over time

Potential Future Directions

Potential Future Directions

Since we are releasing this as an open-source data system, it is our goal to facilitate others in the community to contribute to improving and extending the data and data system.

- General improvements
- Better constraining emissions, and emissions uncertainty, using observational data
- Regular updates to extend emission time series to most recent full year
- Additions (sub-regions, sectors, emission species) as might be needed for specific projects
- Facilitating community involvement

General Improvements

There are may areas in which the system can be improved.

Some specific points

- Adding regionally-specific emissions grids as proxy data (e.g. HTAP_V2)

  Coupled with sub-country resolution for large countries (e.g. USA, China) this will improve the spatial pattern of emissions and their changes over time for use in global/regional models.

- Add additional sectoral resolution (and associated gridding proxies if available)
- Developing further algorithms to detect data inconsistences so they can be resolved (e.g. inventory emissions inconsistent with driver data)

Improving inventories with observational data

How do we come to a synthesis of bottom up and top down methods?

The methods and data used, e.g. bottom-up inventories, atmospheric models, surface and aircraft measurements, and satellite data products all have errors and biases that must be considered.

- **Comparing surface point measurements with estimates averaged over large grid cells** is problematic, as demonstrated by Wang et al. (2014)
- Important to note that **uncertainty in inventory estimates varies** by country, sector, and species. (e.g., any inversion technique that doesn't match SO_2 emissions in countries such as USA, Canada, and Western Europe is likely to be incorrect!)
- Important to **compare to both gridded and aggregate inventory data** where possible
  - Gridded inventory data are generally first produced at an aggregate level, and these aggregate data are then mapped to a spatial grid. Spatial mapping introduces additional error.
- Global and regional vs local applications
  - Increasing number of urban- and regional-scale studies, but we now need more synthesis in order to draw more general conclusions
  - Use of concentration ratios is promising

Some limited use of observations in CEDS

Missing SO_2 emissions identified from satellite data by McLinden et al (2016) were added to six countries where their "missing" estimate was greater than default totals.

- This was a conservative approach – greater disaggregation of the CEDS inventory in the future (explicit breakout of petroleum refining sector) would allow greater use of such estimates.
  - Important not to double count emissions (e.g., emissions may be in the inventory but just not properly assigned to the grid)
- Explicit incorporation of point source information (location, source strength) into CEDS would also allow use of spatial information from satellite data.
  - Satellite data are best for estimating "large" sources. For some emissions more diffuse sources are substantial.
  - Point source data in some existing inventories are derived from proprietary facility datasets. Which could be a tradeoff with our general goal of having an open-source data system.

Annual Emission Time Series Updates

The CEDS system is designed to facilitate regular updates.

There are two main categories of data that need to be updated to produce emissions estimates to a more recent year

- Driver Data
  - BP Energy statistics - required out to most recent year
  - IEA energy statistics – Useful to update to get more recent sectoral data
  - Other driver data (metal smelting, pulp and paper production, etc.)
- Country-level Emission Inventories
  - USA, Canada, Europe, Australia, Korea, etc.
  - Useful to get the most recent trends

Within CEDS, each of these their own read-in routine so little effort to update in as long as data format hasn't changed too much.

We will be implementing an update to extend the estimate to 2016 over this summer and will use this to track the time required to do this.