Addressing time-series consistency/recalculations within the AFOLU sector of the U.S. GHG Inventory

The 18th Workshop on Greenhouse Gas Inventories in Asia (WGIA 18)

July 14, 2021

John Steller

U.S. Environmental Protection Agency

GHG Inventory Time-Series

Compiling time-series greenhouse gas (GHG) estimates enhances utility of data, but it is important to ensure consistency over time (i.e., IPCC good practice)

Enables inventory users to understand:

- Historical emissions trends
- Trends are real and not only reflecting method changes
- Effects of emission reduction strategies/measures at the national level

GHG Inventory Time-Series

- A time-series helps tell a user how anthropogenic emissions and sinks are changing over time
- Under the Enhanced Transparency Framework (ETF), countries will be reporting a time-series (e.g., 1990-2022) in their first Biennial Transparency Report (BTR) and may also include a stand-alone inventory
- Emissions and removals across the time-series should be estimated with consistent methods and data
- *In some cases*, it may not be possible to use the same method and/or data sets for the entire time-series
    - IPCC Guidelines includes techniques to help address data gaps and ensure a consistent time series

Resource: Technical handbook for developing country Parties on Preparing for implementation of the enhanced transparency framework under the Paris Agreement: https://unfccc.int/sites/default/files/resource/ETF%20Handbook-first%20edition%20June%202020-for%20costing.pdf

Agriculture and LULUCF Sectors – Unique Data Circumstances

- Most countries have access to agriculture, land use, land use change and forestry (LULUCF) data through ministries and other data providers
- National data is also available through other organizations e.g., FAOSTAT
- While data is generally available to conduct an inventory at a Tier 1 level, the agriculture and LULUCF sectors can have some unique data circumstances which require extra attention
  - Key data sources such as natural resource surveys are often conducted and published on a periodic basis resulting in data gaps.
  - Some data sources, such as forest inventories, may not be available annually because of resource constraints.
  - Other key data sources may not be available for the early years of your time-series, such as for newer remote sensing products, or more recent years

Splicing Techniques

- For addressing data gaps, *2006 IPCC Guidelines* provide methods to overcome/address data gaps in time series to ensure consistency
- IPCC authors had foresight → they understood the need to provide methods that facilitate incorporation of newer, improved data
- <u>Splicing</u>: combining or joining of more than one method or data series to form a complete time-series
  - Overlap
  - Surrogate
  - Interpolation
  - Extrapolation
- Splicing techniques help:
  - Methodological change and refinement (including incorporating new data and science)
  - Address data gaps

Splicing Techniques

TABLE 5.1
SUMMARY OF SPLICING TECHNIQUES

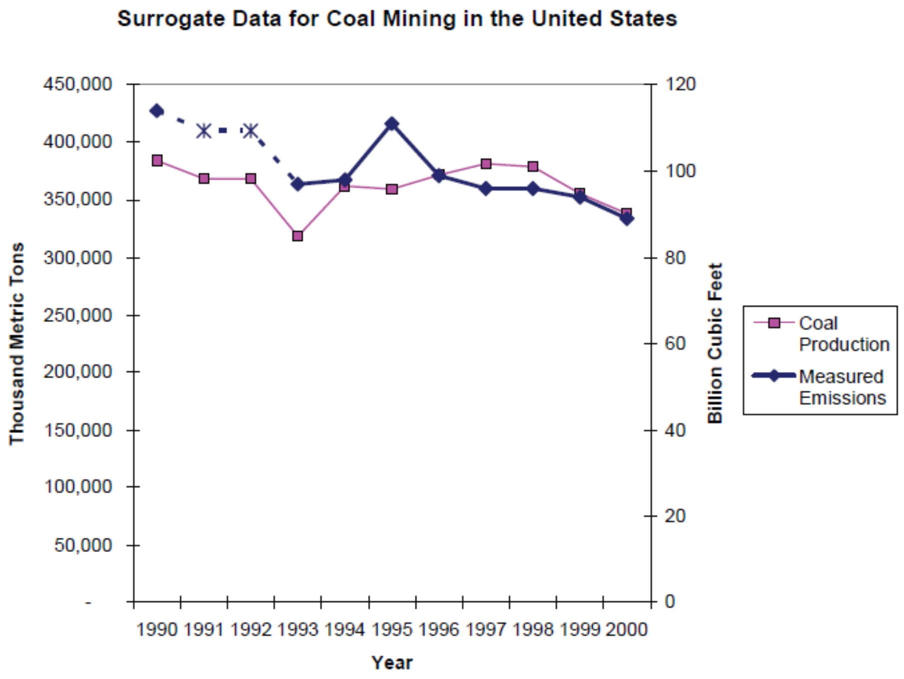
Approach	Applicability	Comments
Overlap	Data necessary to apply both the previously used and the new method must be available for at least one year, preferably more.	• Most reliable when the overlap between two or more sets of annual estimates can be assessed. • If the trends observed using the previously used and new methods are inconsistent, this approach is not *good practice*.
Surrogate Data	Emission factors, activity data or other estimation parameters used in the new method are strongly correlated with other well-known and more readily available indicative data.	• Multiple indicative data sets (singly or in combination) should be tested in order to determine the most strongly correlated. • Should not be done for long periods.
Interpolation	Data needed for recalculation using the new method are available for intermittent years during the time series.	• Estimates can be linearly interpolated for the periods when the new method cannot be applied. • The method is not applicable in the case of large annual fluctuations.
Trend Extrapolation	Data for the new method are not collected annually and are not available at the beginning or the end of the time series.	• Most reliable if the trend over time is constant. • Should not be used if the trend is changing (in this case, the surrogate method may be more appropriate). • Should not be done for long periods.
Other Techniques	The standard alternatives are not valid when technical conditions are changing throughout the time series (e.g., due to the introduction of mitigation technology).	• Document customised approaches thoroughly. • Compare results with standard techniques.

2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 1: General Guidance and Reporting, Chapter 5: Time Series Consistency

(Additional examples are also provided in the *2019 Refinement to the 2006 IPCC Guidelines*)

Splicing Techniques – e.g., Surrogate Data

- The surrogate method relates emissions/removals to underlying activity or other related data indicative of emissions
- Changes in these data are used to estimate the trend in emissions or removals.
- The estimate should be related to the data source that best explains the time variations of the category
- Such as:
  - Crop sales data (taking into account import/export)
  - Crop productivity and harvested area
  - Milk production data
  - Animals slaughtered
  - Gross-domestic product of each specific category



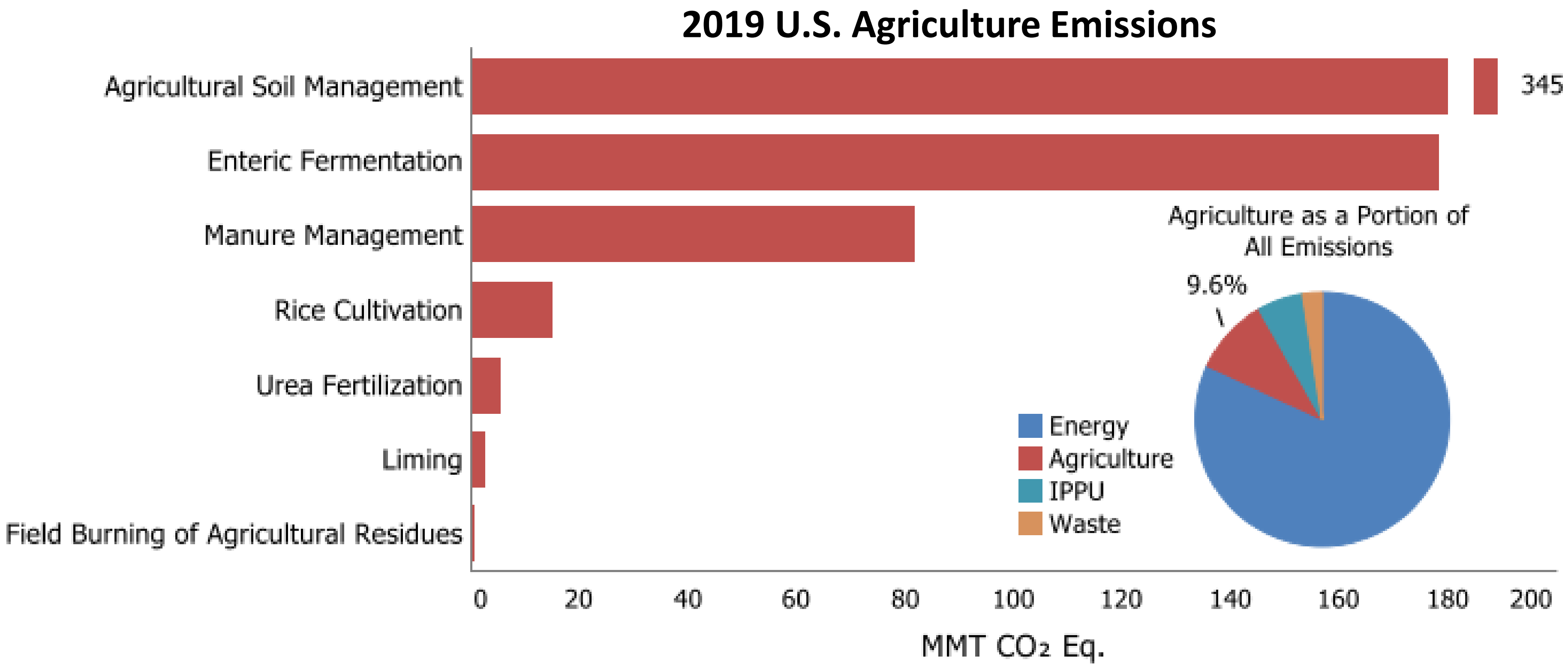
E.g., Coal mining emissions tied to underground coal production

2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 1: General Guidance and Reporting, Chapter 5: Time Series Consistency;

Additional Guidance: 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories

Agriculture Emissions in the U.S.

- Agriculture represents ~9.6% of 2019 emissions
- Primarily from managing soils (N_2O) and livestock (CH_4)



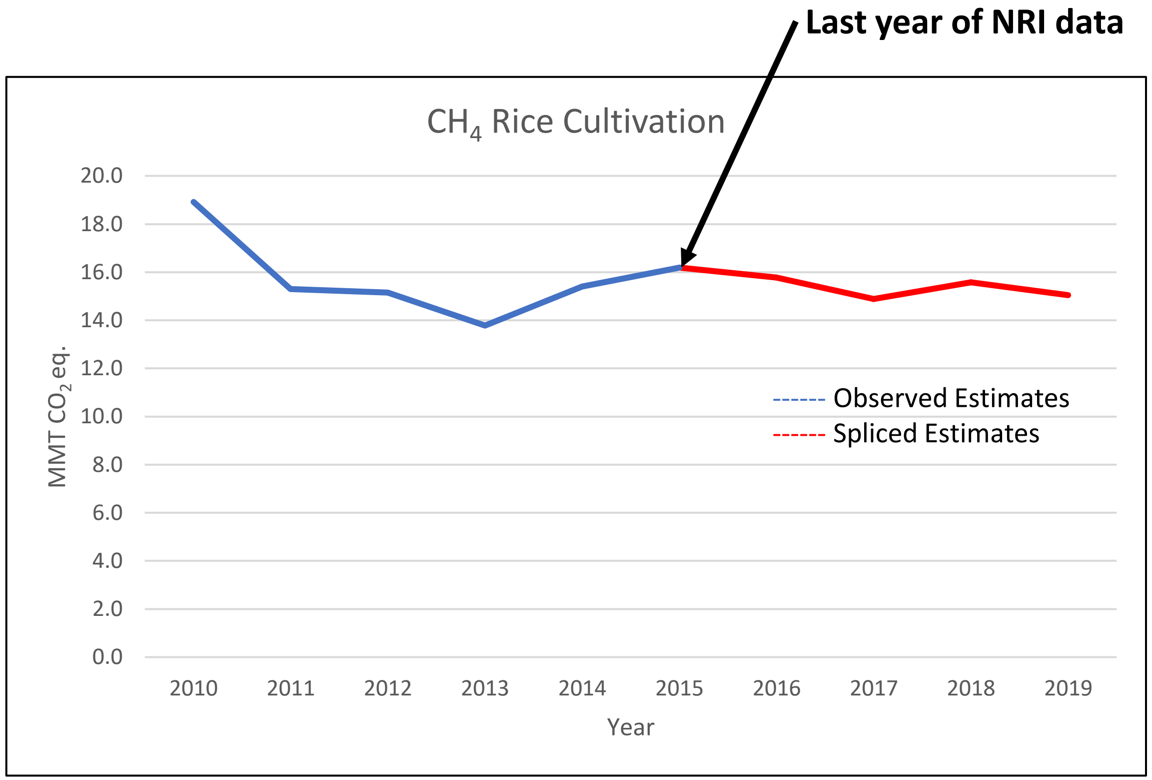
Rice Cultivation – CH_4 Emissions

- Key activity data = rice cultivation areas
- U.S. Department of Agriculture, National Resources Inventory (NRI):
  - Statistically-based sampling (survey) of all U.S. non-federal land
  - Includes key data such as rice cropping and land use histories
  - Published every 3 to 5 years
  - The NRI generally lags 2 to 3 years behind the latest year of the time-series, resulting in the need to extend time-series in a consistent manner
- Current U.S. GHG Inventory includes NRI data through 2015
- Therefore, each Inventory without new NRI data has gap at the end of the time-series
  - We need an approach to extend the time-series

Rice Cultivation – Approach and Outcome

Approach – Surrogate Data Method

- Used a linear regression:
  - Establish relationship between surrogate data and observed data
- Surrogate data used:
  - Rice commodity data
- Observed data: 1990-2015 (i.e., available National Resources Inventory (NRI) data)
- Model predicted/estimated missing data for 2016-2019
- Time-series will be updated when new data is available



$$y_0 = y_t \bullet \left(s_0 / s_t \right)$$

y : emission/removal estimate in years 0 and *t*
s : surrogate statistical parameter in years 0 and *t*

Time-Series Documentation and QA/QC

- Check approach for outliers, inconsistencies (i.e. implied EFs)
- All recalculations and methods/techniques used to improve the time-series consistency and incorporated new data, methods, etc. should be well documented and reported
- When using splicing techniques, it is best practice to document the:
  - Reason for the recalculation
  - Effect of the recalculation on the time series
  - Splicing techniques used
  - Other information that a reader would need to know to understand how you addressed time-series consistency

Example documentation: U.S. GHG Inventory

Box 5-2: Surrogate Data Method

An approach to extend the time series is needed to estimate emissions from Rice Cultivation because there are gaps in activity data at the end of the time series. This is mainly due to the fact that the National Resources Inventory (NRI) does not release data every year, and the NRI is a key data source for estimating greenhouse gas emissions.

A surrogate data method has been selected to impute missing emissions at the end of the time series. A linear regression model with autoregressive moving-average (ARMA) errors (Brockwell and Davis 2016) is used to estimate the relationship between the surrogate data and the observed 1990 to 2015 emissions data that has been compiled using the inventory methods described in this section. The model to extend the time series is given by

$$Y = X\beta + \varepsilon,$$

where Y is the response variable (e.g., CH_4 emissions), $X\beta$ is the surrogate data that is used to predict the missing emissions data, and ε is the remaining unexplained error. Models with a variety of surrogate data were tested, including commodity statistics, weather data, or other relevant information. Parameters are estimated from the observed data for 1990 to 2015 using standard statistical techniques, and these estimates are used to predict the missing emissions data for 2016 to 2019.

A critical issue in using splicing methods is to adequately account for the additional uncertainty introduced by predicting emissions with related information without compiling the full inventory. For example, predicting CH_4 emissions will increase the total variation in the emission estimates for these specific years, compared to those years in which the full inventory is compiled. This added uncertainty is quantified within the model framework using a Monte Carlo approach. The approach requires estimating parameters for results in each Monte Carlo simulation for the full inventory (i.e., the surrogate data model is refit with the emissions estimated in each Monte Carlo iteration from the full inventory analysis with data from 1990 to 2015).

Recalculations Discussion

The major improvements to the current Inventory were (1) incorporating new land use and crop histories from the NRI survey; and (2) modeling SOC stock changes to 30 cm depth with the Tier 3 approach (previously modeled to 20 cm depth), which impacts the simulation of methanogenesis in DayCent. The surrogate data method was also applied to re-estimate stock changes from 2016 to 2017. These changes resulted in an average increase in rice cultivation CH_4 emissions of 1.2 MMT CO_2 Eq. from 1990 to 2017, which is an average of 9 percent larger compared to the previous Inventory.

Outcomes

- Developing time series data is useful for understanding trends, meets ETF requirements
  - An Inventory should reflect a time-series, rather than a single year (If this statement is “IPCC good practice” state on slide)
- Where there may be data gaps, splicing techniques are available for developing consistent estimates of emissions/removals for all years
- You can have confidence in the quality of your estimates by using these widely accepted techniques
- You can improve over time with each subsequent inventory through documentation!

Resources

- 2006 IPCC Guidelines for National Greenhouse Gas Inventories
- 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories
    - Provides case studies of using splicing techniques
- Technical handbook for developing country Parties on Preparing for implementation of the enhanced transparency framework under the Paris Agreement
- Brockwell, Peter J., and Richard A. Davis. *Introduction to time series and forecasting*. Springer, 2016.
- U.S. EPA Toolkit for Building National GHG Inventory Systems
    - Includes pre-defined National System Templates which help provide comprehensive documentation of each critical component of managing the Inventory process

Thank You for Your Attention!

John Steller

U.S. Greenhouse Gas Inventory Team – Agriculture/LULUCF

U.S. Environmental Protection Agency

Steller.John@epa.gov