Addressing time-series consistency/recalculations

April/May 2024

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) so trends reflect real changes rather than method changes

Time series facilitates understanding of

- Emission trends over time
- Effects of emission reduction strategies/measures at the national level

GHG Inventory Time-Series

- 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
  - *Noting flexibility in light of capacities para 57 of MPGs : Parties may report data covering, at a minimum, the reference year/period for their NDC under Article 4 of the Paris Agreement and, in addition, a consistent annual time series from at least 2020 onward*
- 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 authors had foresight, i.e., they understood gaps would exist, along with the need to provide methods that facilitate **incorporation of new and improved data**
- 2006 IPCC Guidelines (and its 2019 Refinement) include guidance on techniques to help address data gaps and recalculate data to ensure a consistent time series

Reasons to Recalculate

- Available data have changed
- A category has become key (e.g., move to higher tiers)
- The previously used method is not consistent with the 2006 IPCC guidelines for that category
- The previously used method is insufficient to reflect mitigation activities in a transparent manner
- The capacity for inventory preparation has increased
- Correction of errors

COMMON DATA GAPS

- Data availability issues
  - Periodic data
  - Available data covers only part of time series
- Non-calendar year data

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 (Tier 1)
  - Can also be used for QA if using higher tiers
- 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 earlier years of your time-series, such as for newer remote sensing products, or more recent years

See also in 2006 IPCC GL and 2019 Refinement to 2006 IPCC Guidelines (below)

Box 5.1 (Updated)
Recalculation in the Agriculture, Forestry and Other Land Use (AFOLU) Sector

It is anticipated that the use of recalculation techniques in the AFOLU Sector will be particularly important. The development of inventory methods and interpolation/extrapolation tools (models) for this sector is ongoing and it is anticipated that changes to the methods of many countries will occur over time due to the complexity of the processes involved. In simple cases, sampling or experimentation may provide country-specific emission factors, which might require a time series recalculation. Situations that are more complicated can also arise. For example:

- The instruments used to collect activity data may change through time, and it is impossible to go back in time to apply the new instrument. For example, land-clearing events can be estimated by the use of satellite imagery, but the satellites available for this work change or degrade through time. In this case, the overlap method described in Section 5.3.3.1 is most applicable.
- Some data sources such as forest inventories required for AFOLU categories may not be available annually because of resource constraints. In this case, interpolation between years or extrapolation for years external to the available times series of data may be the most appropriate method to apply, possibly using a proxy. Extrapolated data may be recalculated when final data become available (see Sections 5.3.3.3 and 5.3.3.4 on interpolation and extrapolation).
- Emissions and removals in AFOLU typically depend on past land use activity (*GPG-LULUCF 2003*). Thus, data must cover a large historical period (20-100 years), and the quality of such data will often vary through time. Overlap, interpolation or extrapolation techniques may be necessary in these cases.
- The calculation of emission factors and other parameters in AFOLU may require a combination of sampling and modelling work. Time series consistency must be applied to the modelling work as well. Models can be viewed as a way of transforming input data to produce output results. In most cases where changes are made to the data inputs or mathematical relationships in a model, the entire time series of estimates should be recalculated. In circumstances where this is not feasible due to available data, variations of the overlap method could be applied.

ENERGY, IPPU, AND WASTE SECTORS DATA CIRCUMSTANCES

- Energy
  - Revisions in data collection where data provider doesn't recalculate
- IPPU
  - New plant-specific data, but challenges obtaining historical data
  - Data confidentiality can change over time
  - Surveys vs. census, so may be incomplete, and have to consider activity potentially excluded from survey
- Waste
  - General lack of data

Splicing Techniques

- Splicing refers to 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

2019 Refinement to 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 1: General Guidance and Reporting, Chapter 5: Time Series Consistency

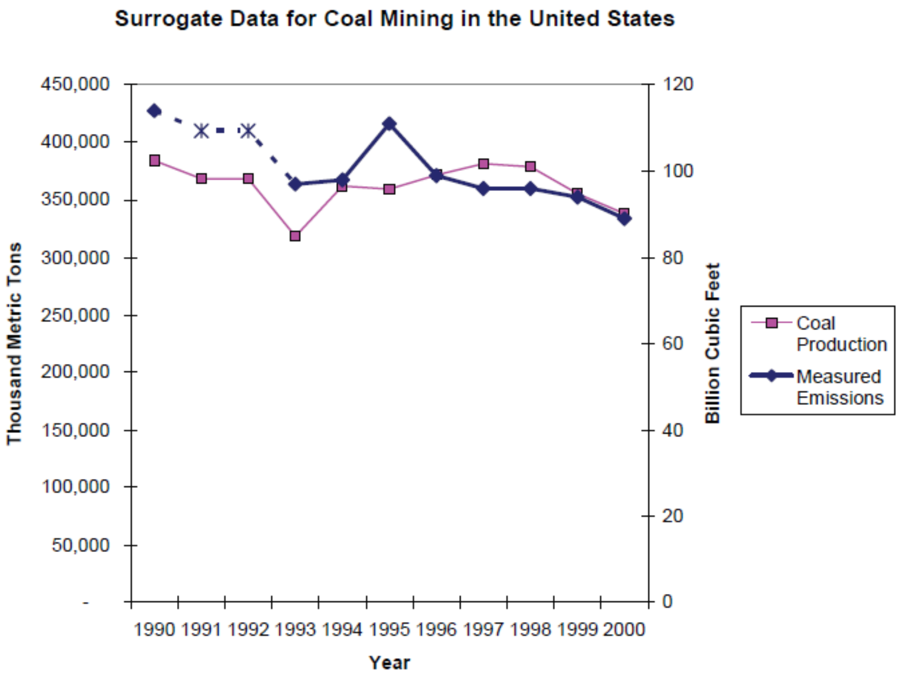
(Additional examples are also provided in the *2019 Refinement to the 2006 IPCC Guidelines*)

Table 5.1 (Updated)
Summary of splicing techniques

Splicing Technique	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 applied for long periods.
Non-Linear Trend Analysis (Interpolation/ Extrapolation)	In cases where time series consistency is best represented by multiplicative (exponential) rather than additive (linear) relationships	• Most reliable for trend analysis of model outputs. • Applicable in the case of large annual fluctuations with observed high standard deviations (see Box 3.0a, Chapter 3, Volume 1 of the *2019 Refinement* for guidance on standard deviation values.
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.

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

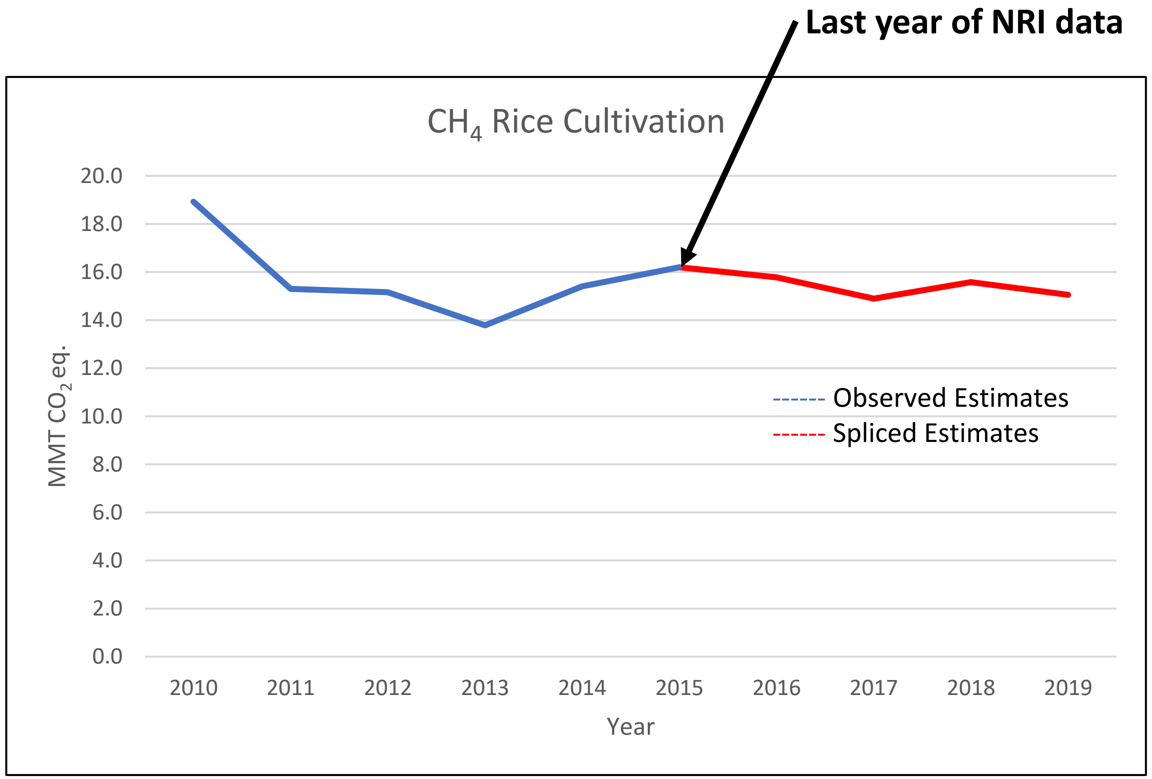
Rice Cultivation – CH_4 Emissions

- 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 2020
- 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: 1990-2020 (i.e., available National Resources Inventory (NRI) data)
- Model predicted/estimated missing data for 2021-2022
- Time-series will be updated when new data is available



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

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

Time-Series Documentation, Reporting 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 (also for consistency with MPGs):
  - 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 because 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 2020 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), Xβ 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 2020 using standard statistical techniques, and these estimates are used to predict the missing emissions data for 2021 to 2022.

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 2020).

Recalculations Discussion

Several improvements have been implemented in this *Inventory* leading to recalculations, including a) updated time series of land representation data that identifies which points and years were sown with rice (Nelson et al 2020), b) extending the time-series of crop history with CDL data, c) imputing ratooning and winter flooding onto individual NRI survey points, d) updated fertilizer and organic amendment additions, and e) revisions to the approach for assigning organic matter amendments and crop residue inputs. As a result of these changes, CO_2-equivalent emissions changed annually with an average annual increase of 0.97 MMT CO_2 Eq., or 5.5 percent, over the time series from 1990 to 2021 compared to the previous *Inventory*.

Outcomes

- Developing time series data is useful for understanding trends and meeting ETF requirements
  - An Inventory should reflect a time-series, rather than a single year (more useful)
  - 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 as new science/data are available and documenting/reporting approach along the way to reflect incremental progress!
- Transparent reporting on recalculations in NID (i.e., explaining reason for recalculation, changes over time series, and impact on emissions trends

Exercise

Consider the example below where we should evaluate the application of the overlap approach to estimate GHG emissions for the years 2001–2003.

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Tier 1 quantified	4,000	4,000	4,100	4,200	4,800	4,900	5,000	4,800	4,900	5,000
Tier 2 quantified				4,035	4,598	4,410	4,500	4,320	4,513	4,790

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