CBIT-GSP
CLIMATE TRANSPARENCY

MINISTRY OF ENVIRONMENTAL PROTECTION AND AGRICULTURE OF GEORGIA

UN environment programme | copenhagen climate centre

Deep-dive into preparation and reporting of results of national GHG inventories under the ETF of the Paris Agreement

Training Workshop for the CBIT-GSP network countries
Tbilisi, Georgia 23 – 25 April 2024

Presentation:
Dealing with inventory data gaps and time-series consistency: issues with data availability; non-calendar year data; completion of missing data (splicing techniques) and recalculations.

Aiymgul Kerimray
Scientist - Energy
UNEP-CCC

UN environment programme

gef

Content

Introduction

- MPG requirements
- Time-series consistency

Resolving data gaps

- Issues with data availability
- Non-calendar year data
- Splicing techniques
  - Overlap
  - Surrogate
  - Interpolation
  - Extrapolation

MPG requirements for time-series consistency and recalculations

Decision 18/CMA.1

26. To ensure time-series consistency, each Party **should use the same methods and a consistent approach** to underlying activity data and emission factors for each reported year.

27. Each Party **should use surrogate data, extrapolation, interpolation and other methods** consistent with **splicing techniques contained in the IPCC guidelines** referred to in paragraph 20 above to estimate missing emission values resulting from lack of activity data, emission factors or other parameters in order to ensure a consistent time series.

28. Each Party **shall** perform **recalculations in accordance with the IPCC guidelines** referred to in paragraph 20 above, ensuring that changes in emission trends are not introduced as a result of changes in methods or assumptions across the time series.

Time-series consistency

Providing a consistent time series of GHG emissions and removals is a key component of a GHG inventory.

All emissions estimates in a time series should be estimated consistently, which means that as far as
possible, the time series should be calculated using;

1. **same method** and
2. **data sources** in all years.

- Using different methods and data in a time series could introduce bias because the estimated emission trend will reflect not only real changes in emissions or removals but also the pattern of methodological refinements.

Issues with data availability

Periodic Data

- **Some data may not be available on an annual basis.**
- Natural resource or environmental statistics, such as national forest inventories and waste statistics, may not cover the entire country on an annual basis.
- Instead, they may be carried out at intervals such as every fifth or tenth year, or region-by-region, implying that national level estimates can only be directly obtained once the inventory in every region has been completed.

Changes and gaps in data availability

- A change in data availability or a gap in data is **different from periodically available** data because there is unlikely to be an opportunity to recalculate the estimate at a later date using better data.
- In some cases, countries will improve their ability to collect data over time, so that higher tier methods can be applied for recent years, but not for earlier years.
- Some countries with economies in transition no longer collect certain data sets that were available in the base year, or if available these data sets may contain different definitions, classifications and levels of aggregation.

Non-calendar year data

It is good practice to **use calendar year data** whenever the data are available.

If calendar year data are unavailable:

Other types of annual year data (e.g., non-calendar fiscal year data e.g., April – March) can be used provided that it is used **consistently over the time series** and the collection period for the data is documented.

Similarly, different collection periods can be used for different emission and removal categories, again provided that the **collection periods are used consistently over time** and documented this is acceptable.

It is good practice to use the **same collection periods consistently** over the time series to avoid bias in the trend.

The data should be **corrected where possible to represent the calendar year**.

If uncorrected data are used, it is good practice for the inventory compiler to make consistent use of either calendar year data or fiscal year data for all years in the time series.

Data types and gaps

	Input Data Type	Examples of Input Data	Typical Sources of Data	Typical Gaps
Historical Data	- Economic indicators	GDP, unemployment rates, inflation rates	National statistical offices, World Bank, IMF	Inconsistent time series, Missing data for certain years
	- Demographic data	Population size, age distribution, urbanization rates	National census bureaus, UN Department of Economic & Social Affairs	Incomplete datasets, Lack of granular data (regional, age-group)
	- Energy usage trends	Energy consumption by sector, renewable energy usage	National energy agencies, IEA	Insufficient granularity, Under-reported sectors
	- Changes in land use	Land use changes, deforestation rates, urban expansion	National environmental agencies, remote sensing databases	Time-lags in reporting, Uncertainty in measurement techniques
Historical Emissions	- Activity data	Energy production and consumption, industrial activities	National environmental agencies, industry reports	Lack of source-specific data, Inconsistent methodologies
	- Emission factors	GHG emission factors for various sectors and activities	IPCC guidelines, national research institutions	Use of default factors, Lack of country-specific data
Non-emissions Data	- Environmental data	Deforestation rates, air and water quality data	National environmental agencies, satellite imagery	Sporadic data collection, Methodological inconsistencies
	- Socio-economic data	Income levels, educational attainment, health indicators	National statistical offices, World Bank	Insufficient data on vulnerable groups, Lag in data reporting
Projected Data	- Drivers	Predicted economic growth, demographic changes, energy price fluctuations	National economic agencies, international financial institutions	Uncertainties in projections, Over-reliance on historical trends
	- Policies	Upcoming government policies, industry standards for emission reductions	Government policy documents, industry reports	Uncertainty in policy implementation, Lack of detailed policies

Splicing Techniques

Splicing: combining or joining of more than one method or data series to form a complete time series

- Methodological change and refinement
- Data gaps

Several splicing techniques are available (the 2006 IPCC Guideline):

- Overlap
- Surrogate
- Interpolation
- Extrapolation

Selecting a technique requires an evaluation of the specific circumstances and a determination of the best option for the particular case

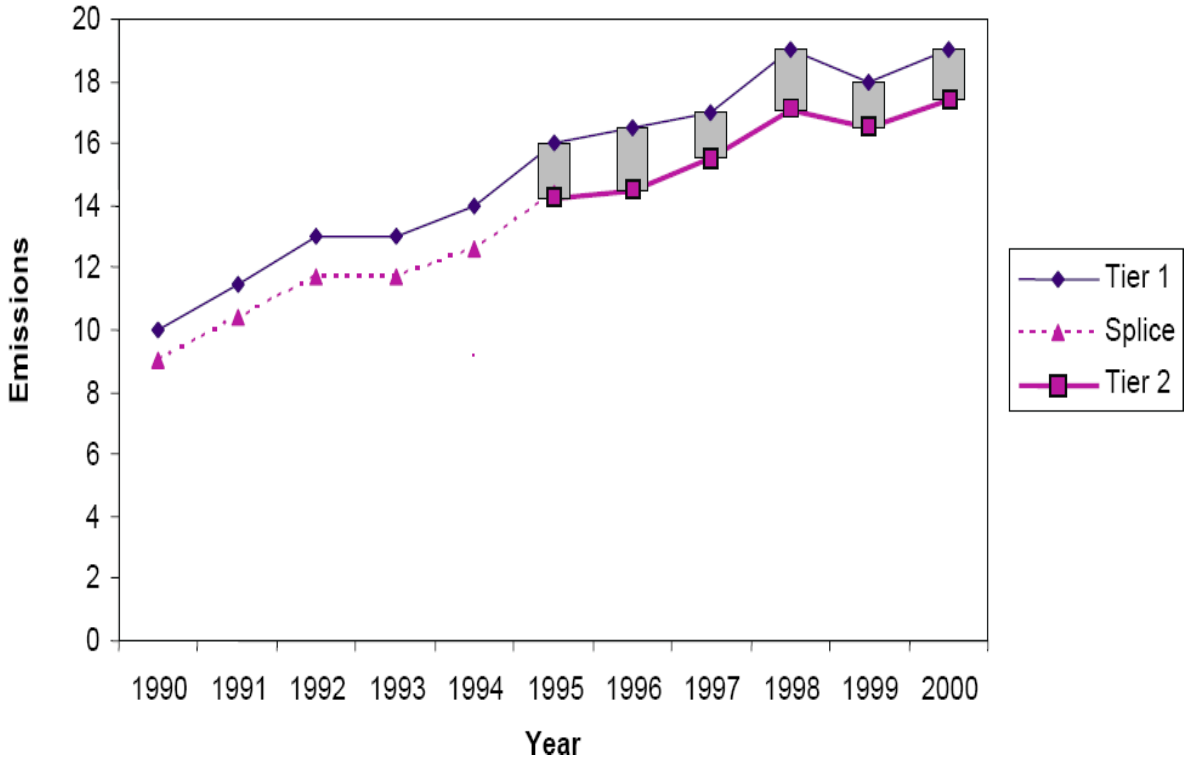
It is good practice to perform the splicing using **more than one technique** before making a final decision and to document why a particular method was chosen.

Overlap

Use case: The overlap method is used when implementing a new data collection method without historical data.

Method: It creates a consistent time series by aligning older estimates with new methods during years where data overlap.

Outcome: This adjusted series ensures a smooth transition to the new method without losing historical continuity.



Overlap

$$y_0 = x_0 \bullet \left(\frac{1}{(n-m+1)} \bullet \sum_{i=m}^{n} \frac{y_i}{x_i} \right)$$

y_0: recalculated estimate using the overlap method
x_0 : estimate developed using the previously used method
m,n : overlapping years

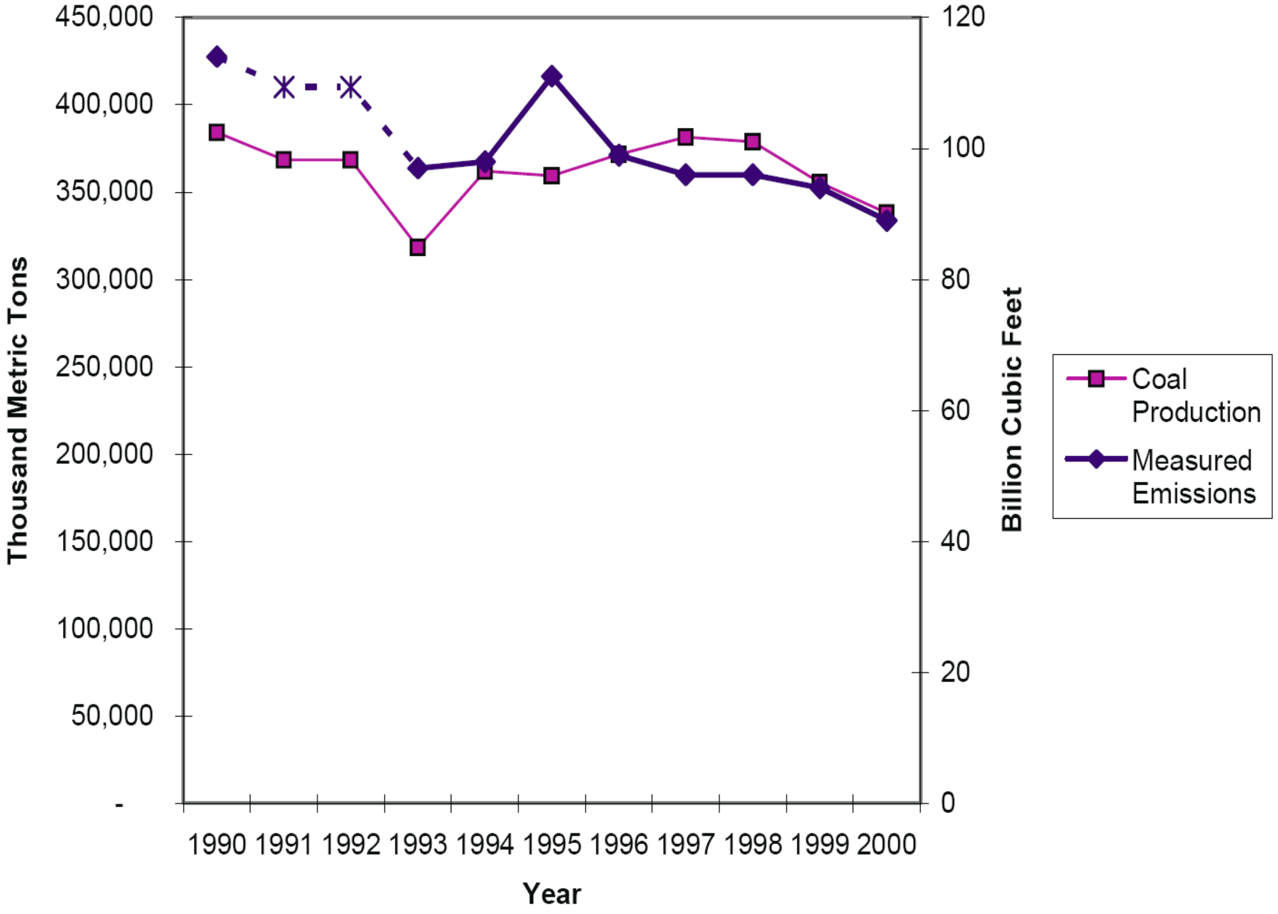
The formula adjusts original emissions estimates to align with a new method by averaging the ratios of new to old estimates during years where both data sets overlap.

Splicing technique: Surrogate

Use Case: The surrogate method is applied when direct data is unavailable.

Method: It uses related activity or indicators to estimate emissions, drawing on statistical correlations to fill data gaps.

Outcome: This approach enhances time series reliability by allowing for more accurate estimations without direct data.



Splicing technique: Surrogate

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

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

Although the relationship between emissions/removals and surrogate can be developed on the basis of data for a single year, the use of **multiple years** might provide a better estimate.

The estimate should be related to the statistical data source that best explains the time variations of the category. For example:

- mobile source emissions may be related to trends in vehicle distances travelled
- emissions from domestic wastewater may be related to population
- industrial emissions may be related to production levels in the relevant industry

Examples of Surrogate Data

TABLE 5.0 (NEW)
EXAMPLES OF SURROGATE DATA BY SECTOR

Energy	IPPU	AFOLU	Waste
• Gross-domestic product • Population statistics • Vehicle fleet • Fuel sales data (taking into account import/export) • Annual Income	• Commodity Production statistics • Gross-domestic product (of each specific category where available) • Plant-specific parameters	• 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 • Fuelwood consumption data (taking into account import/export)	• Gross-domestic product • Population statistics • Annual Income • Protein intake data • Ratio of domestic/ industrial wastewater

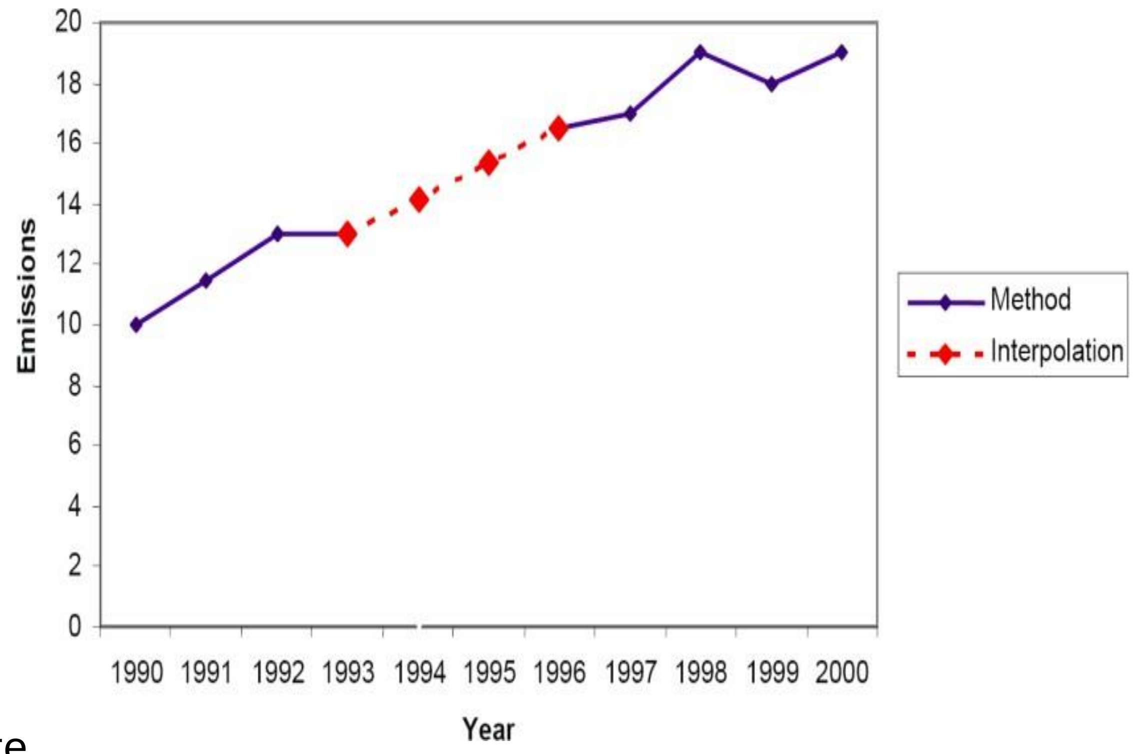
Source: IPCC 2019 Refinement

Interpolation

Use case: The interpolation technique is used when you have emissions data for some years but not for others.

Method: It estimates the missing data by drawing a straight line — linear interpolation — from known data points, assuming emissions change at a consistent rate.

Outcome: This helps to create a smoother, more complete emissions trend line.

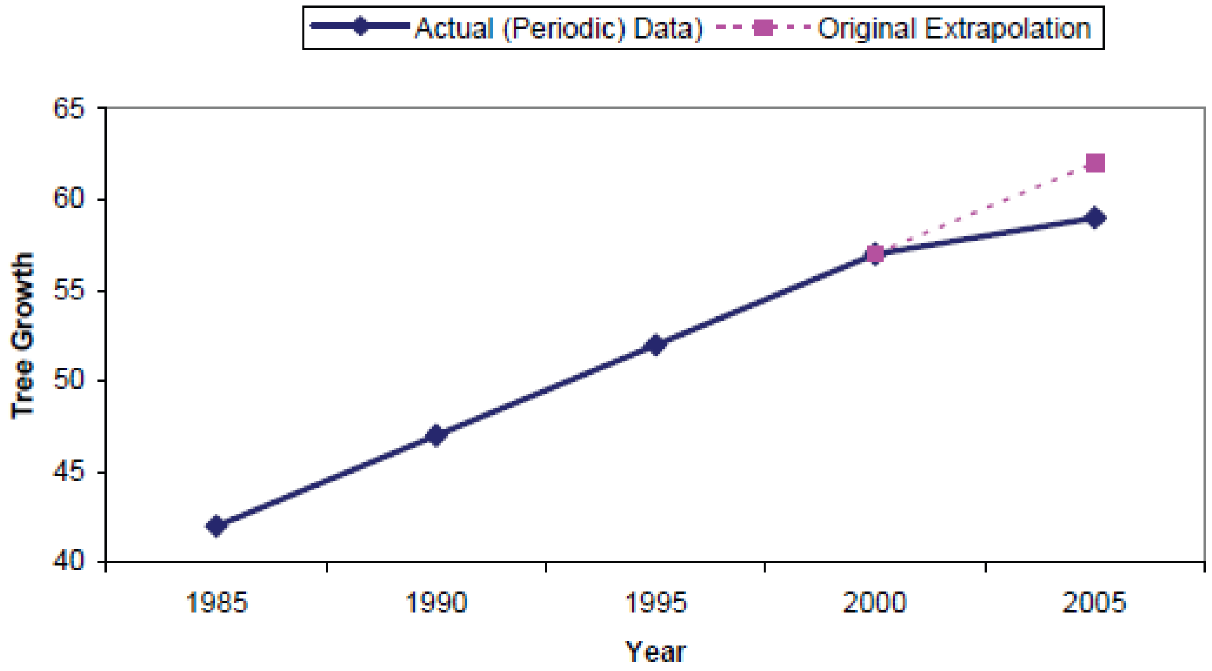


For categories that have volatile emission trends surrogate data will be a better option

It is good practice to compare interpolated estimates with surrogate data as a QA/QC check.

Extrapolation

- **Use case:** Extrapolation is used when data at the start or end of a series is missing.
- **Method:** It extends the trend from known data to estimate these points, using a consistent pattern seen in the data.
- **Outcomes:** This method fills in gaps temporarily, acknowledging that these estimates grow less certain the farther they extend from known data, and are best updated when more information becomes available.



Extrapolation should not be used if the change in trend is not constant over time and should also not be used over long periods of time.

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.

Source: IPCC 2019 Refinement

Quality of Time Series and Documentation

Comparison of the results of multiple approaches where it is possible to use more than one approach

- Plotting and comparing the results of splicing techniques on a graph is useful
- If alternative splicing methods produce different results, should consider which result is most realistic

Comparison of recalculated estimates with previous estimates can be a useful check on the quality of a recalculation

- However, higher tier methods may produce different trends than lower tier methods because they more accurately reflect actual conditions

All recalculations and measures taken to improve time series consistency should be documented and reported

- Reason of the recalculation
- Effect of the recalculation on the time series
- Splicing techniques used

Thank you for your attention!

For more information:

https://climate-transparency-platform.org/

Aiymgul Kerimray
Scientist-Energy
UNEP-Copenhagen Climate Centre
Aiymgul.Kerimray@un.org

CLIMATE TRANSPARENCY PLATFORM

CBIT GSP