Understanding how to generate a consistent time series using agreed IPCC gap filling

5,6,7 April, 2022

Sekai Ngarize

Outline of Presentation

- Why is Time Series Consistency important in GHG inventories?
- Challenges when generating a consistent time series in land use
- IPCC Splicing or Gap filling techniques:
- Overlap
- Surrogate Data
- Interpolation
- Extrapolation
- Summary

Learning objectives

- An understanding of overview of how to address data gaps
- A general understanding of the methods and tools available, as well as of the main challenges of GHG inventory development in the AFOLU sector and how to overcome them using gap filling techniques
- An understating of how to determine which methods suits your country's situation best

Time series Consistency

- Inventories can help you understand emissions/removals trends.
- These trends should be neither over nor underestimated, as long as can be judged.
- An inventory is not just an estimate of a single year. It includes estimates for a number of years (time series of estimates)
- Annual estimates should be comparable and should reflect the real annual fluctuations in emissions and removals
- A time series should be calculated using the same method and same data sources in all years.
- However, it is not always possible to use the same method and data sets for the entire time series due to a lack of data

Challenges when generating a consistent time series in land use

- What to do when you have gaps in land area data?
- What if we have land cover maps data for 2003, 2008, 2013 and 2013 but no data for the intervening years?
- New activity data become available, but not for historical years
- A new source or sink category is added to the inventory, for which historical data are not available
- Department of Forestry stopped collecting data on National Forest Inventory or disaggregated data for wood removals/Timber data
- These problems can be especially a challenge for agriculture and LULUCF sectors.

Data availability and why data gaps occur

- **Changes and gaps in data**:

✓ More disaggregated or other improvements in data collection (e.g., better surveys in future years)

✓ Missing years or data no longer collected.

- **Periodic data:**

✓ Data collection only every few years or on regional rolling basis (i.e., each year a different region surveyed)

✓ Common for the LULUCF sector (e.g., forest inventory only done every five years).

- **No data?**

IPCC Splicing or Gap filling techniques

Splicing - Combining or joining of more than one method or data series to form a complete time series

When to use IPCC Splicing?

- Address a change in method and refinement (e.g., when Tier 2 method can only be applied to new data but Tier 1 is still used for historical data)

- Fill in Data gaps due to collection of period data

The 2006 IPCC Guidelines provide several splicing techniques

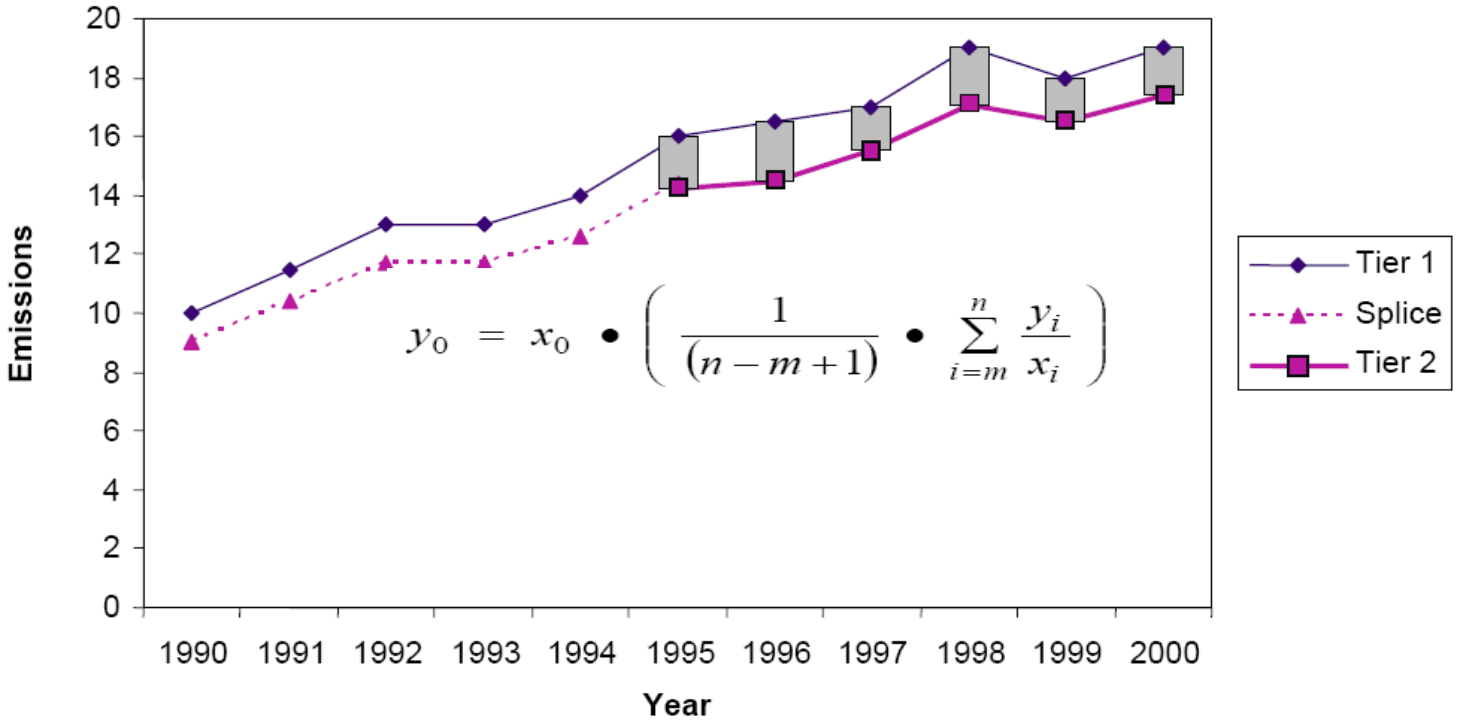
- Overlap
- Surrogate
- Interpolation
- Extrapolation

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

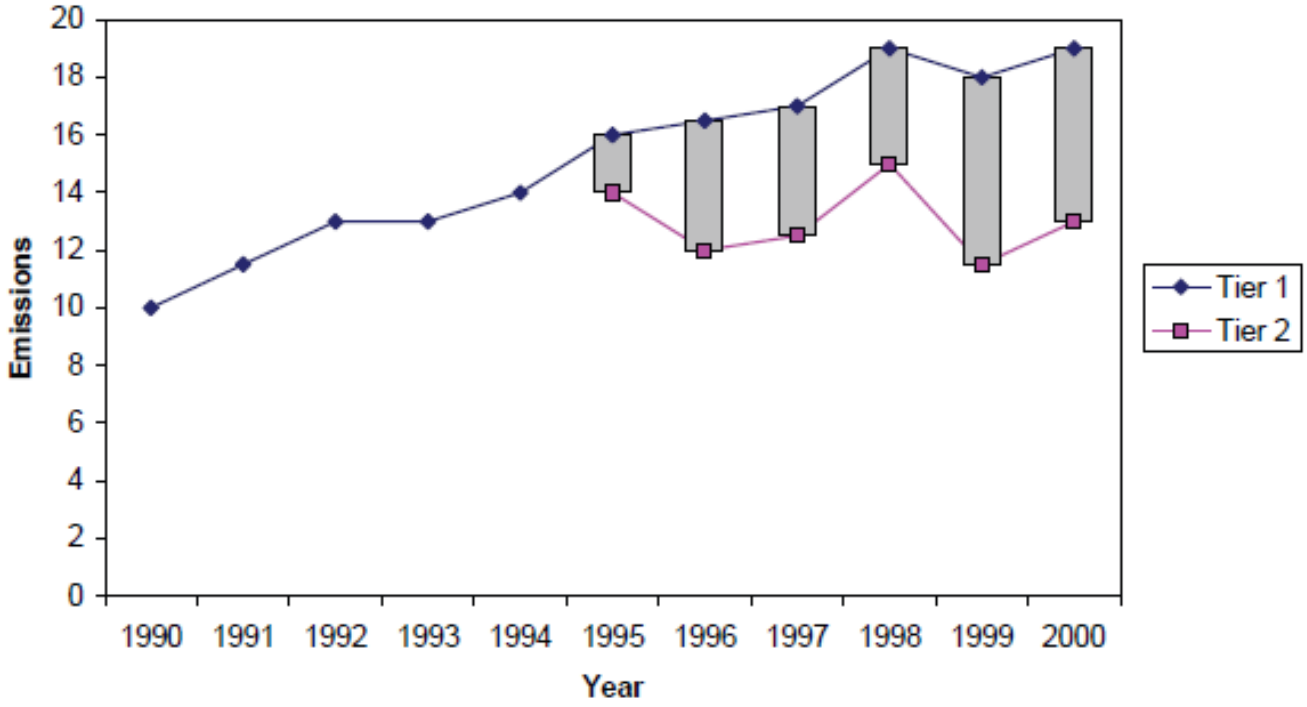
Overlap Approach

- When a new method is introduced but data are not available for early years in the time series (e.g. implementing a higher tier methodology)
- Develop a time series based on the relationship (or overlap) observed between the previously used and new method during the years when both can be used
- It is preferable to compare the overlap for multiple years to evaluate the relationship between the two methods
- If there is no consistent overlap between methods and it is not good practice to use the overlap technique

Overlap Approach: Consistent Relationship



Overlap: Inconsistent Relationship



Overlap Approach

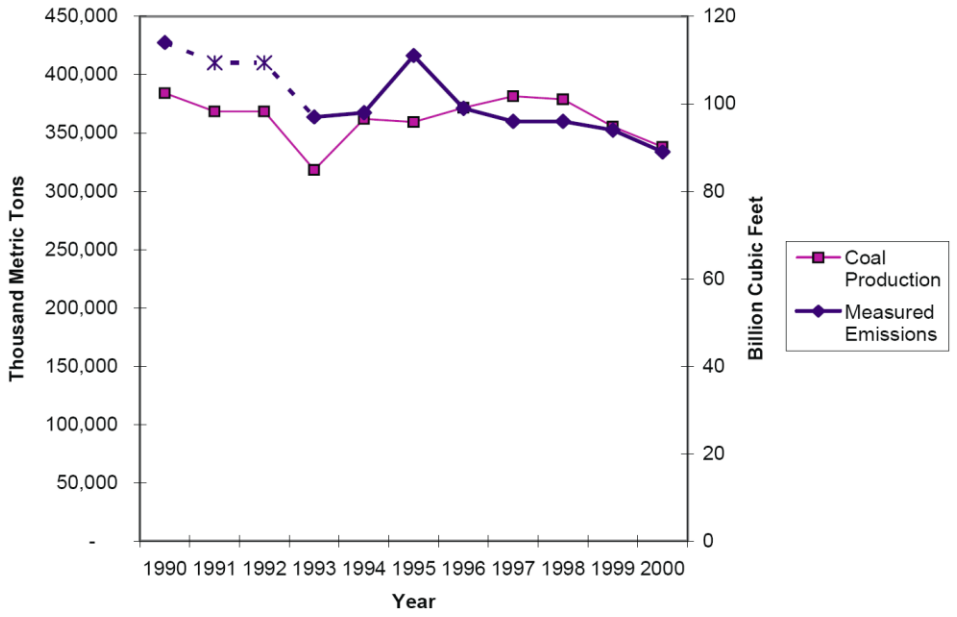
Example 1: Use the overlap approach to estimate GHG emissions for Tier 2 years 1990 – 2000, using the data below.

(See excel files for practical session)

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Tier 1	10.0	12.0	12.5	13.0	13.5	13.9	15.0	15.1	15.0	16.1	17.0	17.9	18.6	19.9	20.5	21.0
Tier 2											15	16	16.8	17.7	18.8	19.1
Estimated/Overlap	9.0	10.8	11.2	11.7	12.1	12.5	13.5	13.6	13.5	14.5	15.3					

Surrogate Data

- The surrogate method relates emissions or removals to underlying activity or other indicative data
- Data (statistical) that is related to the emission (emissions may be proportional to production, vehicle distances travelled and population etc.)
- 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



Surrogate Approach Steps

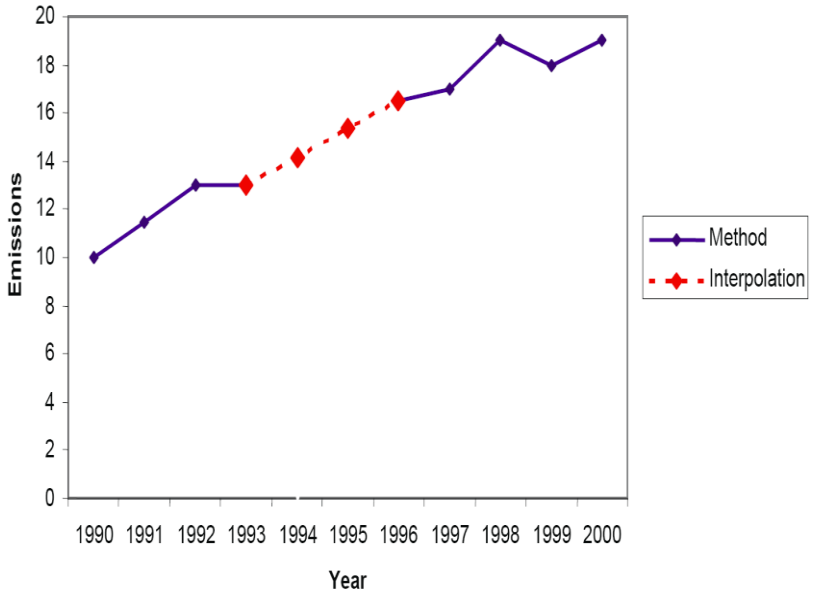
- Identify potential surrogate/proxy variables.
- If you have some actual data, calculate simple **correlation coefficients**:
  - ✓ You should have more than one year of actual data to establish a relationship with the surrogate parameter.
- If the correlation is not obvious, then consider more sophisticated **regression techniques** to see if a relationship between actual and surrogate parameter can be found.
- If you have no **actual data**, then you will **need to justify why the surrogate parameter is a legitimate proxy for actual variable(s).**

Example 2: Use the surrogate approach to estimate GHG emissions years 1990 – 2005, using the data below. (See excel files for practical session)

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Measured Emissions (Yt)	y0	12.0	12.5	13.0	13.5	13.9	15.0	15.1	15.0	16.1	17.0	17.9	18.6	19.9	20.5	21.2
Surrogate data (St)	25.2	27.6	29.3	28.7	31.2	32.1	34.8	33.6	35	37.1	37.4	40.5	42.6	45.4	45.9	46.4
Estimated/Surrogate (	10.95652															
	y / s	0.43	0.43	0.45	0.43	0.43	0.43	0.45	0.43	0.43	0.45	0.44	0.44	0.44	0.45	0.46

Interpolation

- When detailed statistics is collected every few years, or it is impractical to conduct detailed surveys on an annual basis
- Estimates for the intermediate years in the time series can be developed by interpolating between the detailed estimates when:
- Overall trend appears stable
- Actual emissions are not substantially different from the values estimated by interpolation
- If information on the general trends or underlying parameters is available, then the surrogate data can be used
- It is good practice to compare interpolated estimates with surrogate data as a QA/QC check



Interpolation

Interpolation: Filling gaps in existing time series.

Techniques:

- ✓ Linear or nonlinear, justify choice
- ✓ Should not be used for variables that have large variability from year to year.

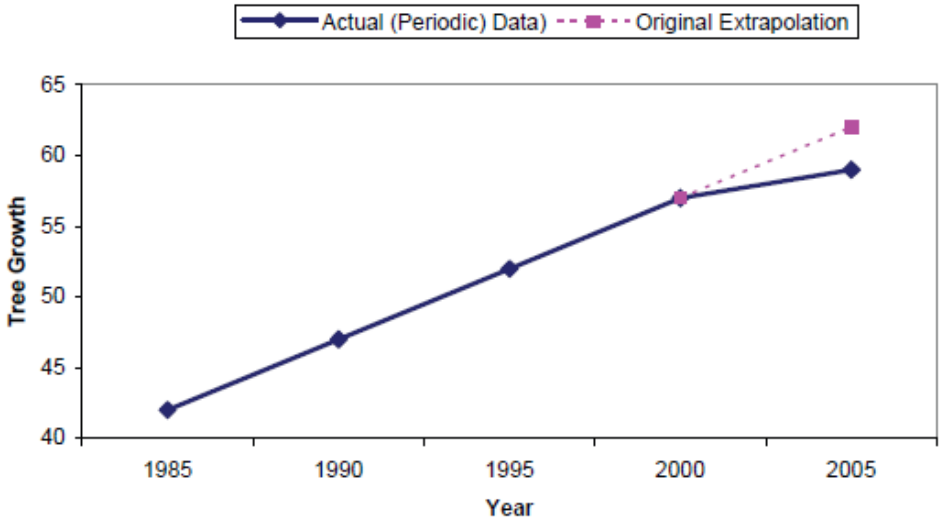
$$Y_t = Y_{start} + \frac{(T_t - T_{start})}{(T_{end} - T_{start})} * (Y_{end} - Y_{start})$$

Example 3: Use the overlap approach to estimate forest land area for years 1994 – 1998, using the data below.
(See excel files for practical session)

					T_{start}	T_1	T_2	T_3	T_4	T_{end}						
	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Total forest area	946837.1	941190.6	935544.1	929897.7	924251.2					896018.9	890372.5	878217.7	866062.9	853908.2	841753.4	829598.6
Forestland remaining forestland	910620.8	904974.3	899327.9	893681.4	888035.0					859802.7	854156.2	838430.2	822704.2	806978.2	791252.2	775526.2
Land converted to Forest land	36216.3	36216.3	36216.3	36216.3	36216.3					36216.3	36216.3	39787.5	43358.7	46930.0	50501.2	54072.4
Interpolated - Total forest area						918604.76	912958.307	907311.8497	901665.3925							
Interpolated - Forestland remaining forestland						882388.5	876742.0461	871095.5889	865449.1316							
Interpolated - Land converted Forest land						36216.261	36216.26084	36216.26084	36216.26084							
					Y_{start}	Y_1	Y_2	Y_3	Y_4	Y_{end}						

Extrapolation

- When data for the base year or the most recent year are not available
- The data can be extrapolated assuming that the trend in emissions/removals remains constant over the period of extrapolation
  - ✓ Should not be used if the trend is not constant over time.
- Analyse the character of trend – e.g. linear or more complex



Extrapolation

- Filling gaps at end or beginning of time series.
- **Techniques**:
  - ✓ Linear or nonlinear, justify choice
  - ✓ Should not be used for variables that have large variability from year to year.

Example 4: Use the extrapolation approach to amount of waste for years 2003 – 2005, using the data below.

(See excel files for practical session)

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Amount of waste, 1000 ton	10.0	12.0	12.5	13.0	13.5	13.9	15.0	15.1	15.0	16.1	17.0	17.9	18.6	19.9		
Extrapolated (simple)	simple extrapolation													19.9	21.2	22.5
Extrapolated (since 1998)	forecast function													19.9	20.7	21.6

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

Summary

- We need consistent estimates of emissions/ removals for all years
  - ✓ Same method and data sources should be applied to all years, if possible
- Where this is not possible, inventory compilers should follow the time series consistency guidance to provide consistent estimates for all years
  - ✓ Overlap/ Surrogate / Interpolation / Extrapolation /etc.
- We need to ensure quality of time series
  - ✓ Quality checks are applied to entire time series
- All decisions, methods and reasons should be documented