CGE Greenhouse Gas Inventory Hands-on Training Workshop

Addressing Data Gaps

Target Audience and Expectation from Training Material

- This training material is suitable for persons with **beginner** to **intermediate level** knowledge of national GHG inventory development.

- After having read this Presentation, in combination with the related Handbook, the reader should:
  - have an **overview** of how to address data gaps;
  - have a **general understanding** of the methods and tools available, as well as of the main challenges of GHG inventory development in that particular area;
  - be able to **determine which methods** suits her/his country's situation best;
  - know where to **find more detailed informatio**n on the topic discussed.

- This training material **is developed primarily on the basis of methodologies developed, by the IPCC**; hence the reader is **always encouraged to refer to the original documents** to further detailed information on a particular issue.

Acronyms

- **EF** Emission Factor
- **LULUCF** Land Use, Land-Use Change and Forestry

Problems

- What do we do when we are missing data?
- We only have data for 1995 and 2000?
- We want to switch to a Tier 2 method, but we only have disaggregated livestock data starting last year
- The Energy Ministry stopped collecting data on natural gas flaring. What do we do?

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
- Time series should be calculated using the same method and same data sources in all years
  - **Reality**: not always possible to use exactly the same methods and data for entire time series

Dealing with Reality

- **Data gaps may occur because**:
  - A new emission factor (EF) or method is applied for which historical data is not available.
  - New activity data become available, but not for historical years
  - Change in how EF is developed or activity data are collected
  - Or activity data cease to be available
  - New source or sink category is added to inventory for which historical data are not available
  - Errors identified in historical data or calculations that cannot be easily corrected

- **Especially a challenge for Agriculture and LULUCF**

Emission Factors

- **Recognize**: using a constant emission factor does not ensure time series consistency.

- **For some emission processes, emission rates may vary over time due to technological or other changes**

  - No: For stoichiometric processes

  - Yes: For many biological and technology-specific processes

Data Availability

- **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 land use sector (e.g., forest inventory only done every five years)
- **No data?**

Splicing and Gap Filling Approaches

- **Splicing**: combining or joining more than one method or data series to form a complete time series
  - Addresses change in method (e.g., when Tier 2 method can only be applied to new data but Tier 1 is still used for historical data)
  - Fill gaps due to the collection of periodically collected data

- **Use surrogate or proxy data** to "create" data that are otherwise missing

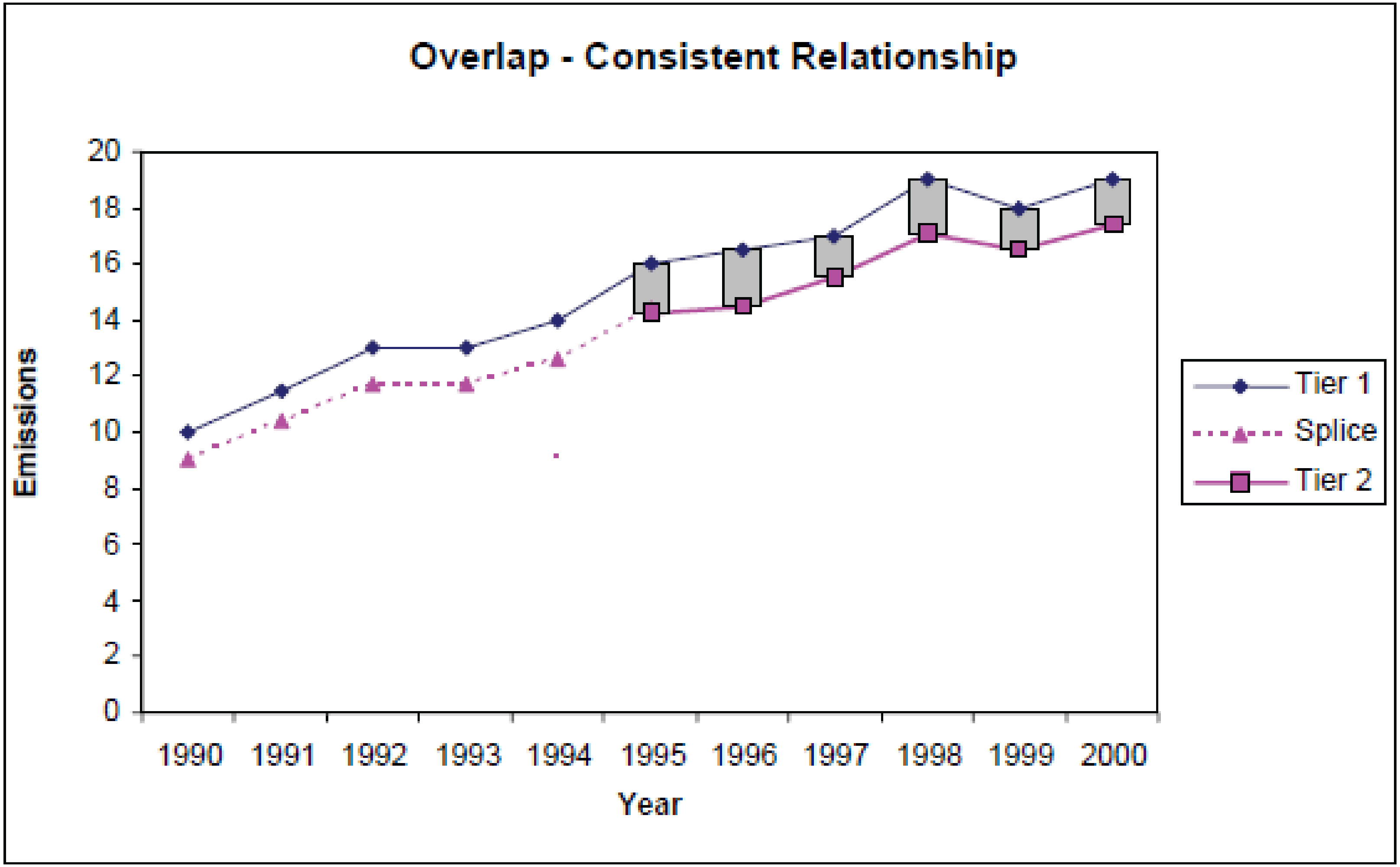
Splicing and Gap Filling Approaches

- Overlap
- Surrogate data (i.e., correlated proxy data)
- Interpolation/Extrapolation
- Trend extrapolation

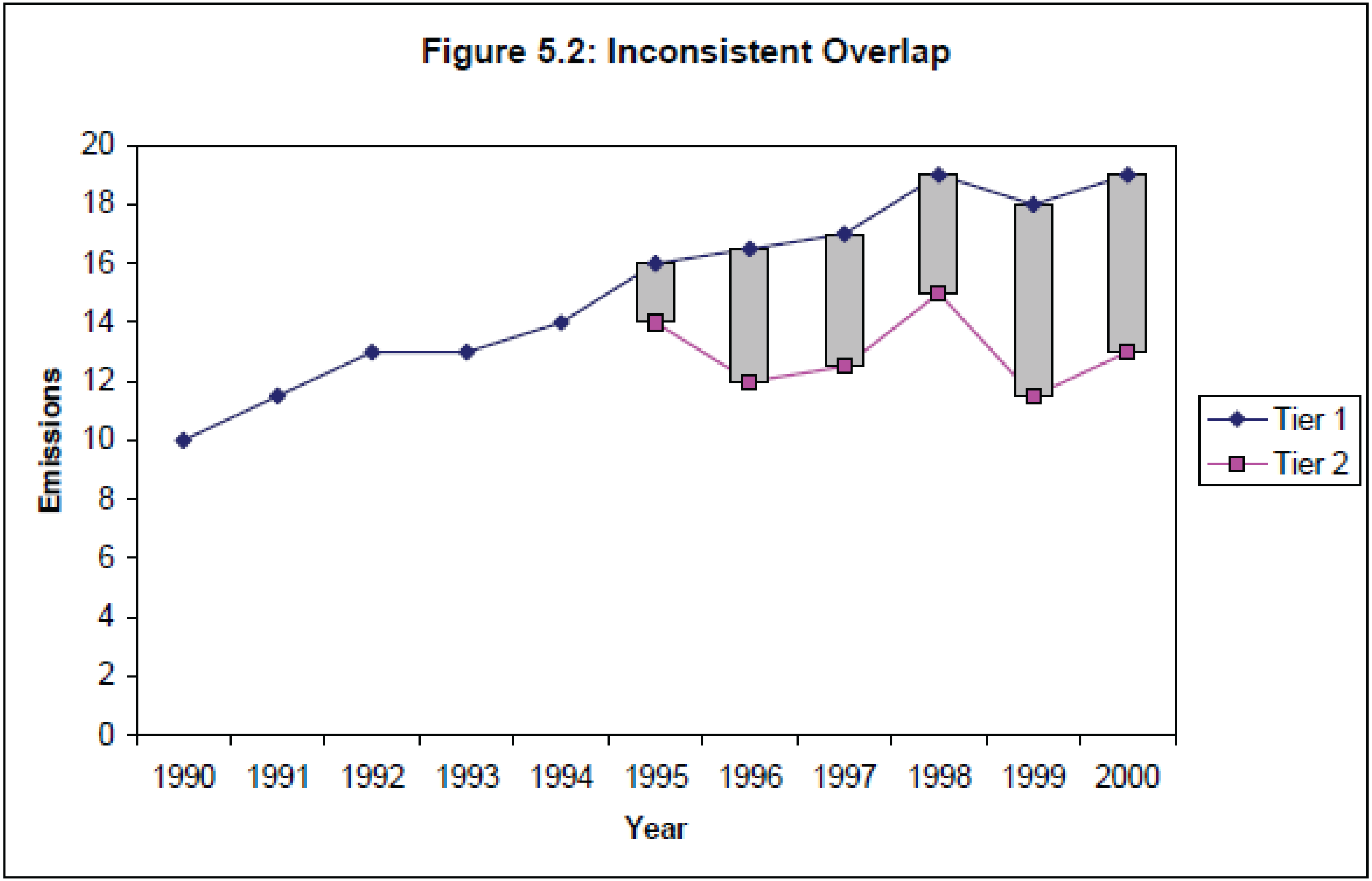
Overlap Approach

- Calculate emissions or collect data using both old and new method/system for several years
  - Can be used with 1 year overlap but should be done with great caution
- Investigate relationship between old and new time series for years of overlap
- Develop mathematical relationship and use to recalculate historical data to be consistent with new method/system

Consistent Overlap Relationship



Inconsistent Overlap Relationship



Overlap Approach

EQUATION 5.1
RECALCULATED EMISSION OR REMOVAL ESTIMATE COMPUTED USING THE OVERLAP METHOD

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

Where:

y_0 = the recalculated emission or removal estimate computed using the overlap method

x_0 = the estimate developed using the previously used method

y_i and x_i are the estimates prepared using the new and previously used methods during the period of overlap, as denoted by years m through n

- Where you have a consistent relationship, default is to use a proportional adjustment of old estimates/data to be consistent with new

Overlap Approach

- **Example 1: Use the overlap approach to estimate GHG emissions for years 2001 – 2003, using the data below**

	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

- **Example 1: Step 1**

	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
Tier 2 - tier 1 ratio				0.96	0.96	0.90	0.90	0.90	0.92	0.96

For each year, calculate the ratio between Tier 2 and Tier 1

E.g. for year 2010:
4790/5000 = 0.96

Overlap Approach

- **Example 1: Step 2**

	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
Tier 2 - tier 1 ratio				0.96	0.96	0.90	0.90	0.90	0.92	0.96

Calculate average and Standard deviation

Average = 0.93

Standard deviation = 0.027

Low variability → Overlap approach seems appropriate

Overlap Approach

- **Example 1: Step 3**

	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	**3,713**	**3,713**	**3,806**	4,035	4,598	4,410	4,500	4,320	4,513	4,790
Tier 2 - tier 1 ratio				0.96	0.96	0.90	0.90	0.90	0.92	0.96

Apply average to calculate missing data:

Year 2001: 4,000 * 0.93 = 3,713

Year 2002: 4,000 * 0.93 = 3,713

Year 2003: 4,100 * 0.93 = 3,806

Overlap Considerations

- Again, should have **multiple years of overlap to apply properly**
- **Should not be applied blindly**. You should do your best to understand the relationship between old and new method
  - E.g., why does old method consistently give results that are 10 to 15% less than new method?
- **If you cannot explain the difference then you are not sure that new method is actually better!**
- Just because a method/model is **more complicated does not mean it is more accurate**

Surrogate Data Approach

- Find a **surrogate** (i.e., **proxy**) variable that is well-correlated with missing data
  - Can be used for missing activity data, for emission factors (that change each year) or for emission estimates
    - Example: Automobile license payments may be well-correlated with petrol use. So license data may serve as a surrogate for petrol consumption
- Approach builds on techniques used in **statistical** (e.g., econometric) **analysis**
  - Regression techniques are valuable to identify potential surrogate parameter
  - Correlation analysis

Surrogate Approach Steps

- ❑ Identify potential surrogate/proxy variables
- ❑ If you have some actual data, calculate simple **correlation coefficients**
  - ❖ Should have more than one year of actual data to establish relationship with surrogate parameter
- ❑ If 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 surrogate parameter is a legitimate proxy for actual variable**

Surrogate Approach

EQUATION 5.2

EMISSION/REMOVALS TREND ESTIMATES USING SURROGATE PARAMETERS

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

Where:

y = the emission/removal estimate in years 0 and *t*

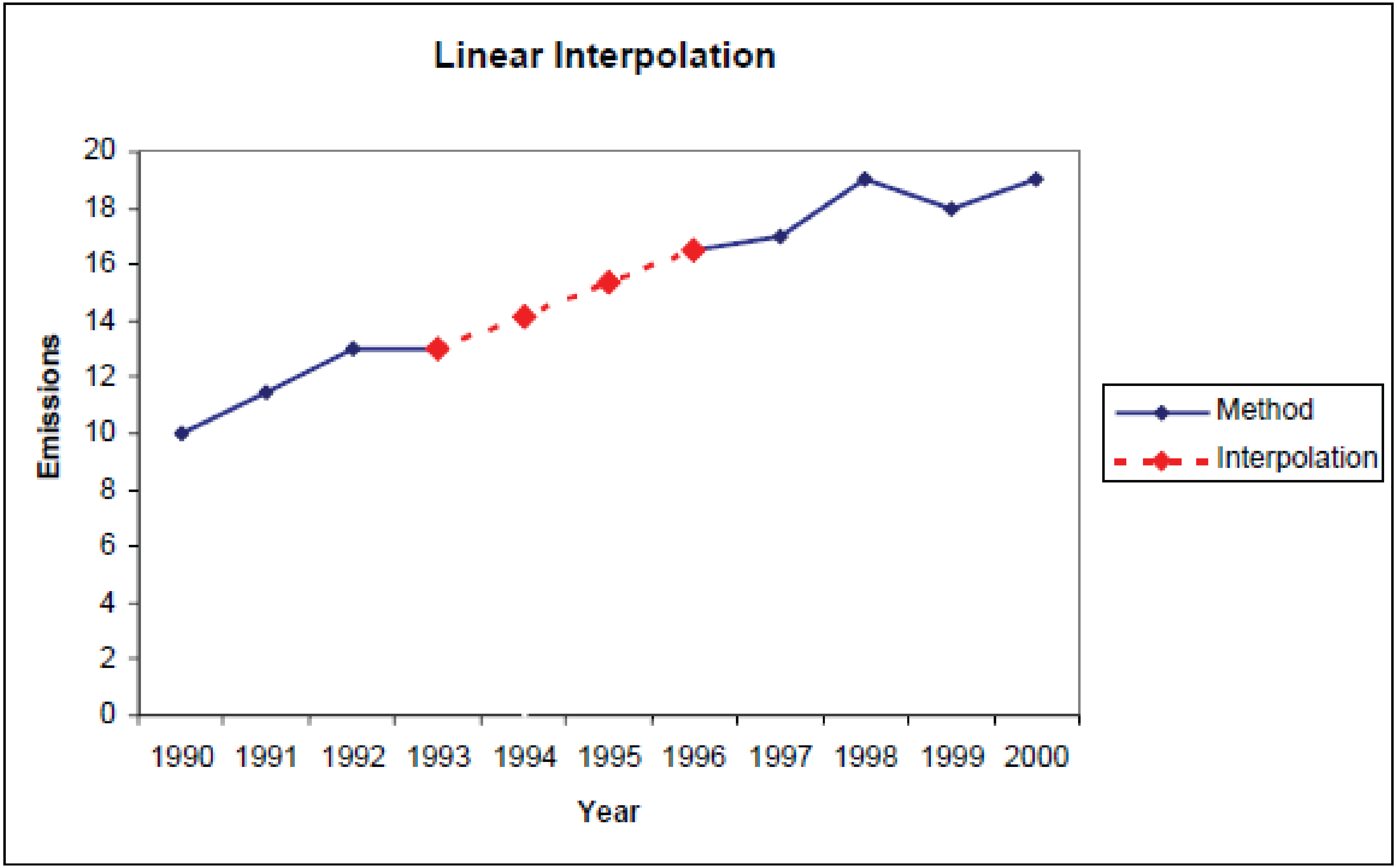
s = the surrogate statistical parameter in years 0 and *t*

- This formula assumes a simple proportional relationship between surrogate and actual variable.

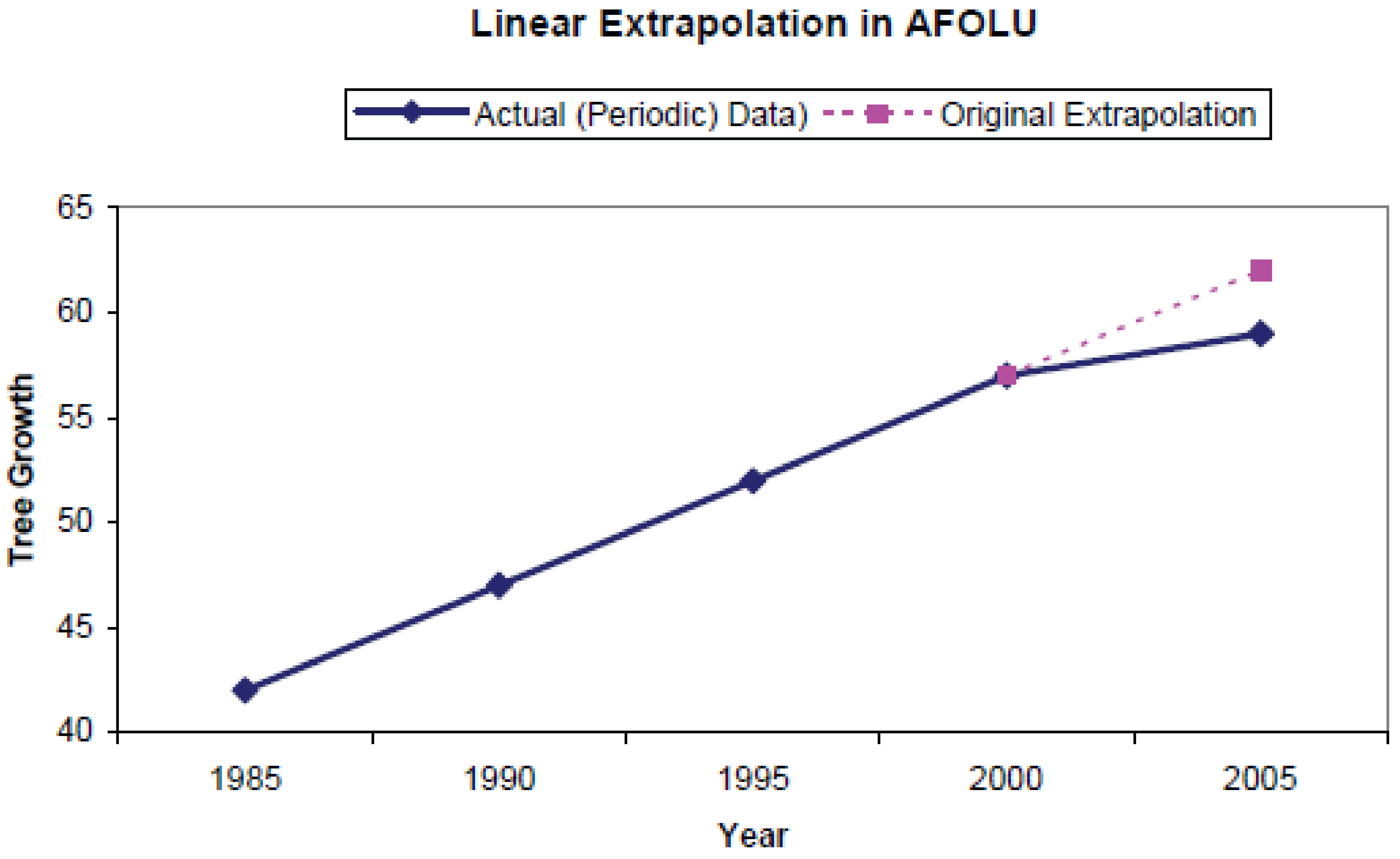
Interpolation and Extrapolation

- **Interpolation**: Filling gaps in existing time series
- **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

Interpolation Example



Extrapolation Example



- **Exercise 3: Using the interpolation technique, estimate the GHG emissions for years 2004 - 2006**

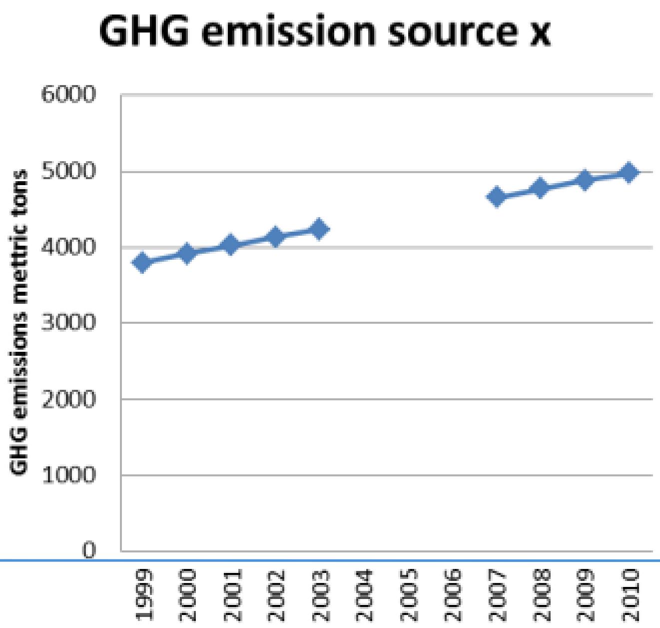
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
GHG emission source x	3,800	3,920	4,030	4,135	4,235				4,655	4,770	4,880	4,975

Interpolation Example

- Exercise 3: Step 1

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
GHG emission source x	3,800	3,920	4,030	4,135	4,235				4,655	4,770	4,880	4,975

Analyze data and assess applicability and type of interpolation technique desired



Linear interpolation seems appropriate for this data set

- Exercise 3: Step 2

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
GHG emission source x	3,800	3,920	4,030	4,135	4,235				4,655	4,770	4,880	4,975

Calculate difference in GHG emissions between last year before the gap and first year after the gap
→ 4,655 – 4,235 = 420

■ Exercise 3: Step 3

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
GHG emission source x	3,800	3,920	4,030	4,135	4,235				4,655	4,770	4,880	4,975

Calculate difference in GHG emissions between last year before the gap and first year after the gap
→ 4655 – 4235 = 420

Calculate length of the gap
→ 2007 – 2003 = 4 years

Exercise 3: Step 3

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
GHG emission source x	3,800	3,920	4,030	4,135	4,235				4,655	4,770	4,880	4,975

Calculate difference in GHG emissions between last year before the gap and first year after the gap
→ 4655 – 4235 = 420

Calculate length of the gap
→ 2007 – 2003 = 4 years

Calculate average change in emissions per gap year
→ 420 / 4 = 105

■ Exercise 3: Step 3

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
GHG emission source x	3,800	3,920	4,030	4,135	4,235	**4,340**	**4,445**	**4,550**	4,655	4,770	4,880	4,975

Calculate difference in GHG emissions between last year before the gap and first year after the gap
→ 4655 – 4235 = 420

Calculate length of the gap
→ 2007 – 2003 = 4 years

Calculate average change in emissions per gap year
→ 420 / 4 = 105

Calculate total emissions for gap year by adding the average change per year

2004 emissions = 4235 + 105 = 4340

2005 emissions = 4340 = 105 = 4445

2006 emissions = 4445 + 105 = 4550

Splicing and Gap Filling Summary

Approach	Applicability	Comments
Overlap	Data necessary to apply both old and new method must be available for at least one year, preferably more.	Only use when overlap shows pattern that appears reliable.
Surrogate Data	Missing date is strongly correlated with proxy data	Should test multiple potential proxy data variables.
Interpolation	For periodic data or gap in time series	Linear or non-linear interpolation. Only use where data shows steady trend.
Trend Extrapolation	Beginning or the end of the time series is missing data	Only use where trend is steady and likely to be reliable. Should only be used for a very few years.

Summary Comments

- **Preferred approaches** are **overlap** and **surrogate**
  - Because they are based on actual data
  - Interpolation and extrapolation are effectively projections that assume certain trends in the absence of data
- Similar in research, **it is not good practice to simply apply a gap filling method blindly**
  - You should understand why your approach is justified and transparently explain why it is.
  - Ask yourself, will what I am doing stand up to peer review in a technical journal?

Thank you