

Approaches to Data collection

GHG Training Workshop, Naivasha, Kenya

8-12 April, 2024

Dr Sekai Ngarize,

Outline

- Why is data collection important in GHG compilation
- Data acquisition and confidentiality
- Collection of data
 - Sources of data
 - Collecting existing data
 - Generating new data
- Adapting data for inventory use
- Derivation or review of emission factors (EFs)
- Production or review of activity data (AD)
- Protocol for expert elicitation
- Summary

Why are approaches to data collection important in GHG compilation process


Data collection is an integral part of developing and updating a GHG inventory



Formalised data collection activities should be established, adapted to national circumstances and reviewed periodically as part of implementing *good practice*



Data collection procedures are necessary for finding and processing existing data, as well as for generating new data by surveys or measurement campaigns



Data collection should cover values and their uncertainties

IPCC provide guidance on collection of existing national/international data and new data on EF, AD and uncertainty data collection- this covers:

1

- Developing a data collection strategy to meet data quality objectives regarding timeliness, and also consistency, completeness, comparability, accuracy, and transparency

2

- Data acquisition activities include generating new source data, dealing with restricted data and confidentiality, and using expert judgement

3

- Turning the raw data into a form that is useful for the inventory

Data acquisition: Dealing with restricted data and confidentiality

- It is good practice to engage data suppliers in the process of inventory compilation and improvement



- Data providers might restrict access to information because it is confidential, unpublished, or not yet finalized.



- Find solutions to overcome these concerns by:
 - ✓ explaining the intended use of the data,
 - ✓ agreeing, in writing, to the level at which it will be made public,
 - ✓ identifying the increased accuracy that can be gained through its use in inventories,
 - ✓ offering cooperation to derive mutually acceptable data sets,
 - ✓ and/or giving credit/acknowledgement in the inventory to the data provided.

Gathering existing data: Possible sources of country-specific data

National

- National statistics Agencies
- Sectoral experts, stakeholder organisations
- Other national experts
- Reference libraries (National Libraries)

- National Inventory Reports from Parties to the United Nations Framework Convention on Climate Change

International

- IPCC Emission Factor Database (EFDB)
- International organisations publishing statistics e.g., United Nations, Eurostat or the International Energy Agency, OECD, FAO and the IMF (which maintains international activity as well as economic data)
- Other international experts

Other

- Scientific and technical articles in environmental books, journals and reports
- Universities
- Web search for organisations & specialists

Collection of data: General guidance for gathering existing data – Important Steps

1. Begin with screening of available data

- Iterative process where details of available data are built up
- Slow and requires questioning until final judgment about usefulness of a data set for the inventory can be made e.g. consider the original intent for data source

2. Refining data requirements

- Formal specification and data request (i.e. knowing what to ask for, from whom, and when)
- Specifications include the following:

Collection of data: Refining data requirements (specifications)...

Part:
1

- Definition of the data set (e.g., time series, sectors and sub-sector detail, national coverage, requirements for uncertainty data, emission factors and/or activity data units),
- Definition of the format (e.g., spreadsheet) and structure (e.g., what different tables are needed and their structure) of the data set

Part:2

- Description of any assumptions made regarding national coverage, the sectors included, representative year, technology/management level, and emission factors or uncertainty parameters
- Identification of the routines and timescales for data collection activities (e.g., how often is the data set updated and what elements are updated),

Part:
3

- Reference to documentation and QA/QC procedures,
- Contact name and organisation,
- Date of availability

Collection of data: General guidance to gathering existing data - National versus International data

Choosing between published national and international data



NATIONAL Data:
Preferable to use national data since sources are typically more up to date and provide better links to the originators of data



INTERNATIONAL Data:
Often international data have undergone additional checking and verification and may rely on nationally-derived data; adjusted to increase consistency

Cross-checking between data sets can help assess completeness and identify problems

Collection of data: General guidance to gathering existing data- Surrogate data

When to use and what is Surrogate data?

- In some cases, directly applicable data may be unavailable or have gaps (e.g., if survey and sampling programmes may be infrequent).
- surrogate data – alternative data that have a correlation with the data that they are replacing should be physically and statistically related to the facility emissions:

What is Good Practice?:

When using surrogate data to estimate emissions or removals, it is **good practice** for countries to perform the following steps:

- 1) Confirm and document the physical relationship between emissions/removals and the surrogate activity data.
- 2) Confirm and document a statistically significant correlation between emissions/removals and the surrogate activity data.
- 3) Using regression analysis, develop a country-specific factor relating emissions/removals to the surrogate data

Collection of data: Generating new data

If representative emission factors, activity data or other estimation parameters do not exist, or cannot be estimated from existing sources.

1. By measurements

- representative sample
- suitable measurement method
- well-designed measurement programme defines (Table 2.1, Vol 1 2006GL):
 - measurement objective
 - methodology protocol
 - measurement plan with clear instructions to the measurement personnel
 - data processing and reporting procedures, and documentation

2. Using models

- Models can be used to generate data
- Models are a means of data transformation

Adapting data for inventory use

- Aim:** to ensure that the level of detail and coverage of the data, including sectors/process/abatement, match the location, land type, compound and years included
- Greenhouse gas inventories require consistent estimates across time series and between categories.
 - Adapting data includes filling gaps if data are missing for one or more years or the data do not represent the year or national coverage required.

(Techniques to address gaps in data sets will be explained in detail in the other presentation “Time Series Consistency”.)

Adapting data for inventory use - examples of approaches

Combining data sets numerically:

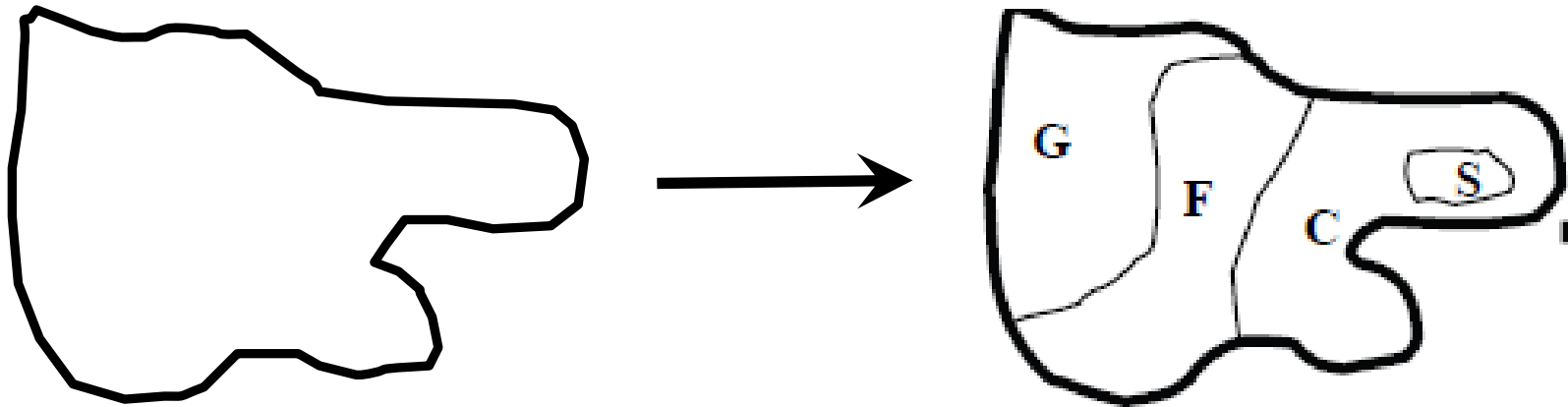
Presented with several potential datasets for the same estimates?

- combination can be achieved by pooling the raw data and re-estimating the mean and 95% confidence limit
- Also possible to combine measurements of a single quantity made using different methods that produce results with different underlying probability distributions.
- Method could be more complex; sufficient to use expert judgment to average the results or identify more reliable data set

Adapting data for inventory use

Data that are not homogenous?

- inventory estimate should be stratified (subdivided) so that each stratum is homogeneous and the national total for the source category will then be the sum of the strata.



Adapting data for inventory use

Empirical datasets contain outliers?

- rule: lying more than 3x standard deviation
- but exercise caution: anomalous data may indicate some other set of circumstances that may need to be separately estimated (e.g. plant in start-up conditions; drainage leading to sudden release of CH₄ in rice field)

Multi-year averaging?

- countries should report annual inventory estimates
- avoid using multi-year averaging of data:
 - over- or under-estimates of emissions over time
 - increased uncertainty
 - reduced transparency, comparability, or time-series consistency of estimates

Adapting data for inventory use

Non-calendar year data?

- **Good Practice** to use calendar year (CY) data whenever available
- if not, use other types; but use consistently over time-series and document
- data should be corrected where possible to represent CY

Regional inventory data?

- in some cases, regional data are more detailed and up-to-date than national data
- can be used provided that:
 - each regional component is compiled consistent with good practice QA/QC, choice of tiers, Time Series Consistency, and completeness
 - The approach to aggregate is transparent
 - The final inventory complies with good practice requirements

How are EFs derived?

It is good practice to follow a stepwise approach to data collection by:

1. setting priorities
2. developing a strategy for accessing the data, and
3. collecting and processing data

The derivation or review of emission factors and other estimation parameters include use of:

1. literature sources
2. data obtained by measurements
3. derived datasets

How are EFs derived?

1. Literature sources

- a useful approach for gathering and selecting from among a variety of possible data
- maybe time-consuming because many lead to old data
- should be fully documented to be transparent
- It is *good practice*, for countries to use their own, peer-reviewed, published literature because this should provide the most accurate representation of their country's practices and activities.
- If no country-specific peer-reviewed studies available, the inventory compiler can use IPCC default factors and Tier 1 methods as indicated by the decision trees in Volumes 2 to 5, or Tier 2 methods with data from Emission Factor Database (EFDB), or other literature values

Table 2.2 , Vol 1, 2006 GL
Potential sources of literature data

Literature Type	Where to find it	Comments
IPCC Guidelines	IPCC website	Provide agreed default factors for Tier 1 methods but may not be representative of national circumstances.
IPCC Emission Factor Database (EFDB)	IPCC website	Described in more detail below. May not be representative of processes in your country or appropriate for <i>key category</i> estimates.
EMEP/EEA Emission Inventory Guidebook (Formerly EMEP/CORINAIR)	EEA (European Environment Agency website)	Useful defaults or for cross-checking. May not be representative of processes in your country or appropriate for <i>key category</i> estimates.
International Emission Factor Databases: USEPA	USEPA website	Useful defaults or for cross-checking. May not be representative of processes in your country or appropriate for <i>key category</i> estimates.
Country-specific data from international or national peer reviewed journals	National reference libraries, environmental press, environmental news journals	Reliable if representative. Can take time to be published.
National testing facilities (e.g., road traffic testing facilities)	National laboratories	Reliable. Need to make sure the factors are representative and that standard methods are used.

How are EFs derived?

2. Data obtained by measurements (applies concepts presented in 2006 GL, Volume 1, Section 2.2.2 to assess quality of measurement data)

- emissions can be determined directly (e.g., using continuous emission monitoring systems) or calculated
- require a well-designed measurement programme
- require suitable and well-documented measurement methods

How are EFs derived?

- 2. Data obtained by measurements** (applies concepts presented in Volume 1, Section 2.2.2 to assess quality of measurement data)
- use instruments with known performance characteristics (i.e. regularly calibrated, maintained, and inspected) or perform relative accuracy audits against established standard reference methods
 - be representative of the sector-specific activity or practice and appropriate for national circumstances
 - not double-count or omit the emissions data that are intended to be measured

How are Activity Data produced?

It is good practice to follow a stepwise approach to data collection by:

1. setting priorities
2. developing a strategy for accessing the data, and
3. collecting and processing data

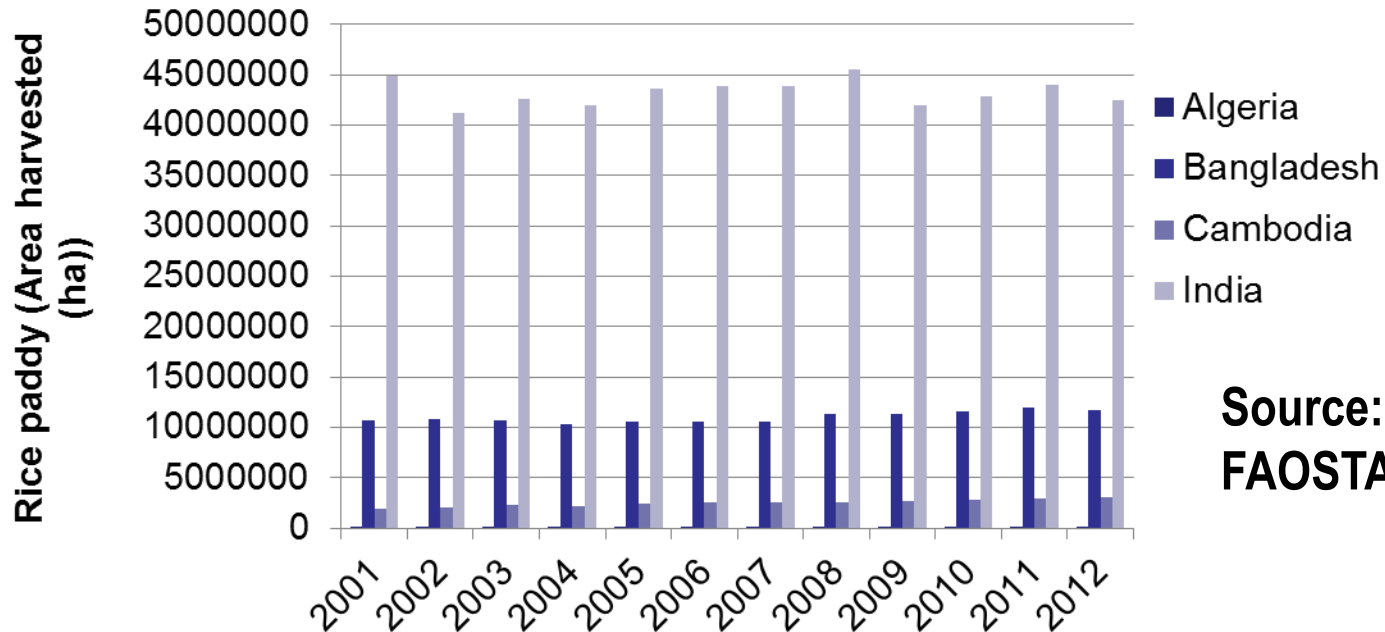
The production or review of activity data include use of:

1. information on specialized data sources
2. surveys and censuses
3. measurement related data, where appropriate

How are Activity Data produced?

1. Information on specialized data sources

- Use data from such bodies as National Statistical Agencies, and national regulatory authorities
- National and international statistics
- Other sources of specialised literature (UN statistics, US Geological Survey, etc.)



Source:
FAOSTAT 2013

How are Activity Data produced?

2. Survey and Census Information

- Provide the best agricultural, production and energy statistics for GHG inventories
- Generally these data are compiled by national statistical agencies or relevant ministries for national policy purposes or to comply with international demand for data

Survey data:

- derived from sampling and do not include real data for the whole population
- representativeness and methods for developing the population estimate require careful review

Census data:

- based on a complete count of the whole population
- limited in detail and can be costly and time consuming

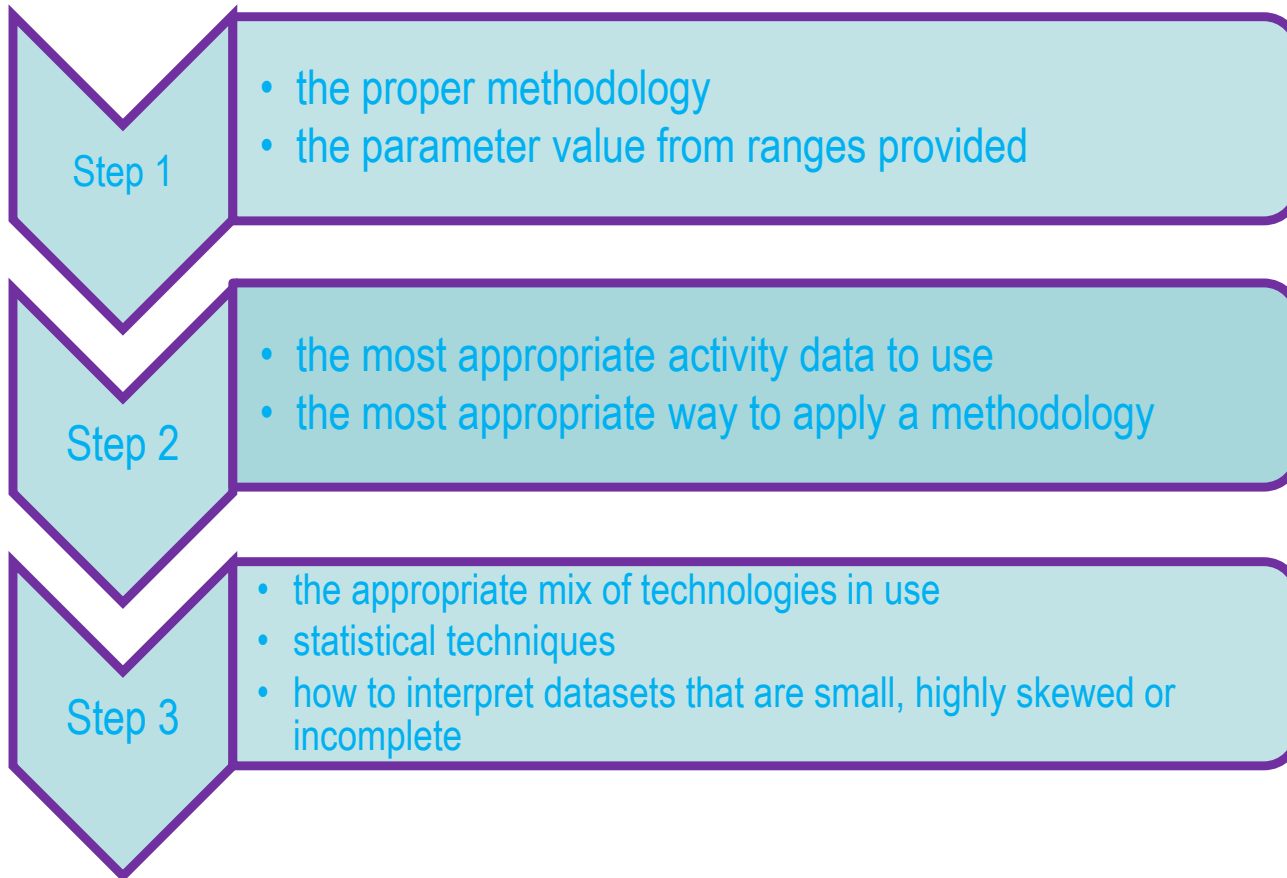
How are Activity Data produced?

2. Survey and Census Information

- **Energy surveys:** The best way to avoid double-counting is to compile energy balances according to the basic principles, concepts and methods developed at the international level.
- **Industry surveys:** Standardised lists of industrial commodities have been established at international level, and countries are encouraged to adopt these lists for collecting harmonized statistics.
- **Agriculture surveys & censuses:** The FAO promotes and provides guidance on national censuses of agriculture.
- **Forest surveys:** The FAO also provides support national forest assessments including on the sampling design, intensity, plot configuration and variables to collect.
- **Waste surveys:** Volume and composition of industrial waste that they produce each year is generally estimated. However, the volume of production is directly proportional. Municipal waste can be estimated by weighing vehicles before and after collection.

Why is Eliciting expert judgment important?-

Aim to be as representative as possible in order to reduce possible bias and increase accuracy.
Use of expert judgment may be needed when determining:



Eliciting expert judgment?

In some cases, an appropriate protocol is especially important. For example, to fill gaps in the available data, to select data from a range of possible values or make judgments about uncertainty ranges.

Motivating: establish a rapport; describe the context

Structuring: define the quantities for which judgements are to be sought (e.g. the year and country, the source/sink category, the structure of the inventory model, etc).

Conditioning: work with the expert to identify and record all relevant data, models, and theory relating to the formulation of the judgements.

Encoding: encoding is the process of converting an expert's judgement regarding uncertainty into a quantitative PDF (see Vol 1. Chapter 3).

Verification: analyze the expert's response and provide the expert with feedback. Is what has been encoded really what the expert meant? Are there inconsistencies in the expert's judgement?

Summary of presentation – Key messages

- IPCC Guidance provides advice on approaches to data collection (Chapter 2, Vol 1, 2006 GL)
- It is good practice to have a formalized data collection process adapted to national circumstances
- Need to follow data collection procedures for collection of new and existing data
- Good practice for Inventory compilers to follow methodological approaches for data collection as listed in the IPCC guidelines
- Advice on emission factors focuses on understanding and generating measured data as well as finding and using default factors
- Advice on activity data focuses on generating and using new census & survey data as well as use of existing national and International data sets
- Guidance on adapting data for inventory use,
- The need for maintaining QA/QC records during every stage of data collection
- Protocol for expert elicitation



Any questions?