



# Approaches to Data Collection

Regional African Workshops on REDD+ National Forest Monitoring Systems  
and Greenhouse Gas National Inventory Systems

25-27 February 2014

Livingstone, Zambia

Kiyoto Tanabe

Technical Support Unit, IPCC TFI

ipcc

INTERGOVERNMENTAL PANEL ON climate change

# Outline

- 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
- Internet resources to identify data agencies and other potential data sources

# Introduction

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

# Possible sources of country-specific data

1. National Statistics Agencies
2. Sectoral experts, stakeholder organisations
3. Other national experts
4. IPCC Emission Factor Database (EFDB)
5. Other international experts
6. 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)
7. Reference libraries (National Libraries)
8. Scientific and technical articles in environmental books, journals and reports
9. Universities
10. Web search for organisations & specialists
11. National Inventory Reports from Parties to the United Nations Framework Convention on Climate Change

# Restricted data and confidentiality

- Data providers might restrict access to information because it is confidential, unpublished, or not yet finalized.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

# General guidance for gathering existing data

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 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:



# Refining data requirements (specifications)...

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

# General guidance to gathering existing data

## 3. Choosing between published national and international data

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



# General guidance to gathering existing data

## 4. Surrogate data

- alternative data that have a correlation with the data that they are replacing
- should be physically and statistically related to the emissions:
  - Confirm and document the physical relationship between emissions/removals and the surrogate activity data.
  - Confirm and document a statistically significant correlation between emissions/removals and the surrogate activity data.
  - Using regression analysis, develop a country-specific factor relating emissions/removals to the surrogate data

(Use of surrogate data to address gaps in time series data sets will be explained in detail in the other presentation.)

# Generating new data

## 1. By measurements

- representative sample
- suitable measurement method
- well-designed measurement programme (Table 2.1):
  - 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

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

# Adapting data for inventory use

## Combining data sets numerically:

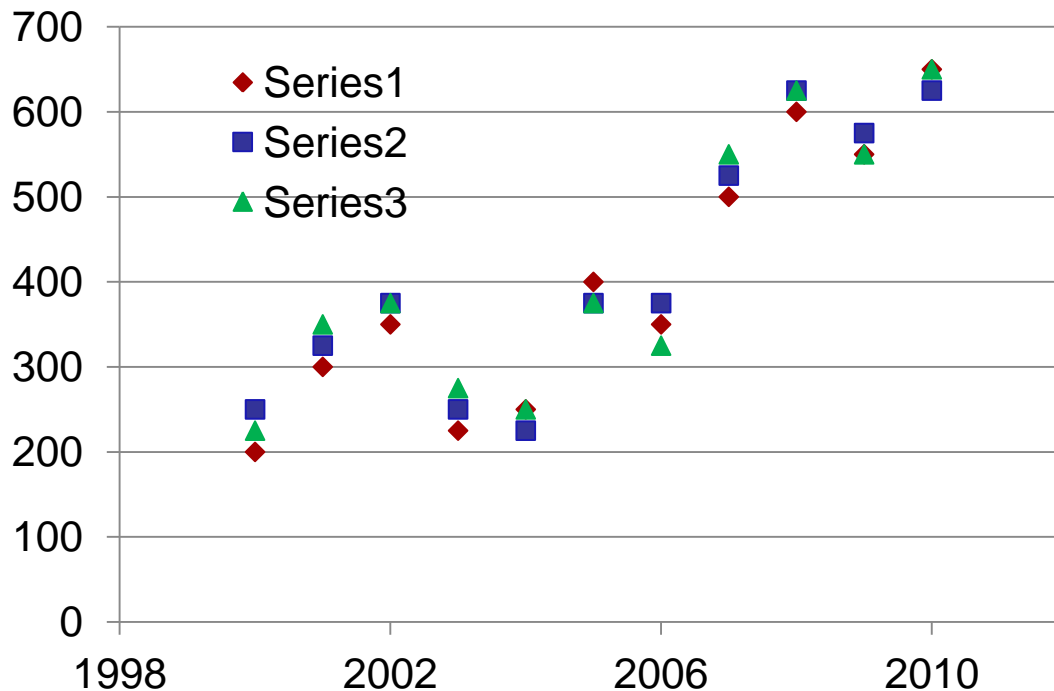
Presented with several potential datasets for the same estimates?

- combination can be achieved by pooling the raw data and re-estimating the mean and 95% confidence limit
- 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

## Combining data sets numerically:

year	A	B	C
2000	200	250	225
2001	300	325	350
2002	350	375	375
2003	225	250	275
2004	250	225	250
2005	400	375	375
2006	350	375	325
2007	500	525	550
2008	600	625	625
2009	550	575	550
2010	650	625	650
<b>avg</b>	<b>398</b>	<b>411</b>	<b>414</b>
<b>95% CI</b>	<b>112</b>	<b>108</b>	<b>109</b>

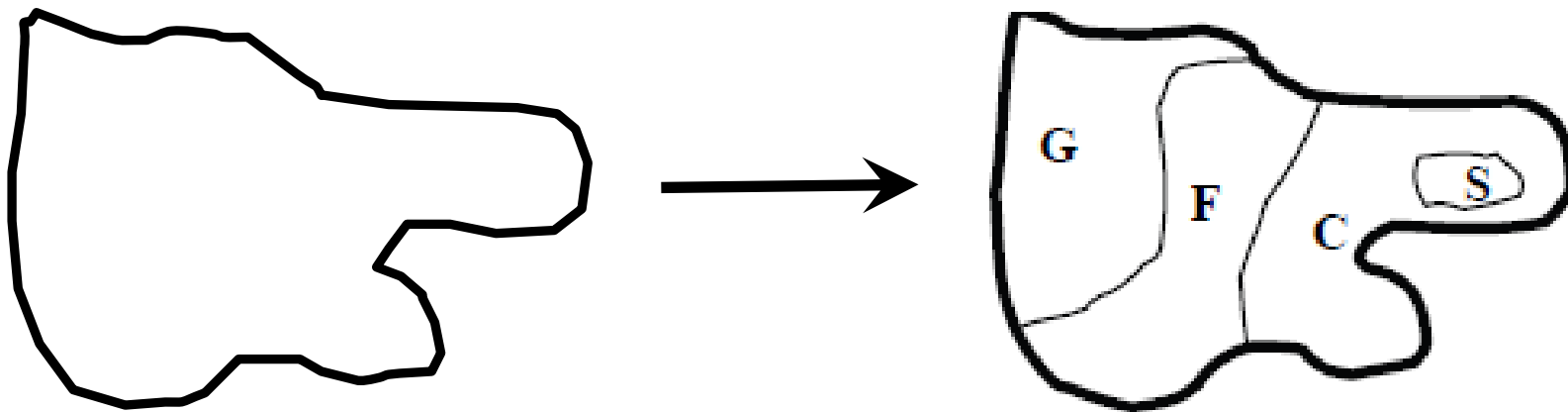


**pooled  
average: 408**  
**pooled 95% CI: 56**

# 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 be careful: 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 CH<sub>4</sub> 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?

- 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, TSC, 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**  
**Potential sources of literature data**

<b>Literature Type</b>	<b>Where to find it</b>	<b>Comments</b>
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/CORINAIR Emission Inventory Guidebook	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 Volume 1, Section 2.2 to assess quality of measurement data)

- emissions can be determined directly or calculated
- require a well-designed measurement programme
- require suitable and well-documented measurement methods



London School of Economics Library  
and Political Science, 1964

# How are EFs derived?

## 2. Data obtained by measurements (applies concepts presented in Volume 1, Section 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:

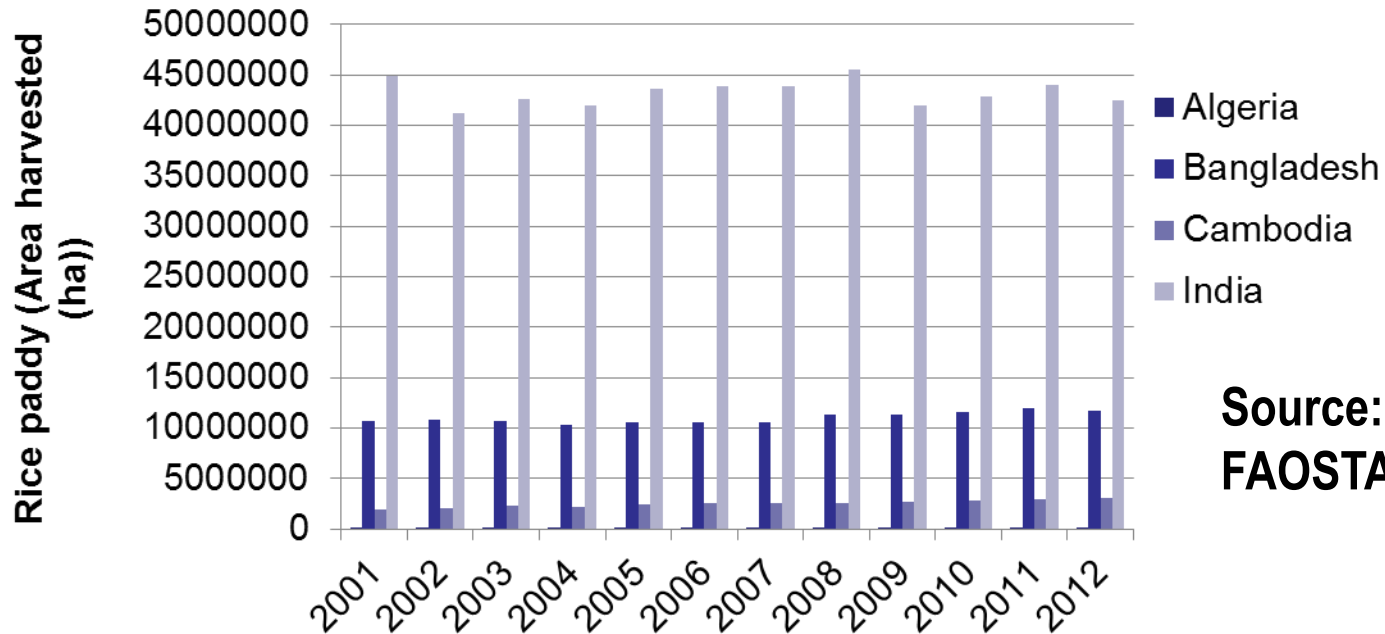
1. information on specialized data sources
2. surveys and censuses
3. measurement related data, where appropriate



# 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

### *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.

# Eliciting expert judgment?

The aim of expert judgment is to be as representative as possible in order to reduce possible bias and increase accuracy.

Use of expert judgment may be needed when determining:

- the proper methodology
- the parameter value from ranges provided
- the most appropriate activity data to use
- the most appropriate way to apply a methodology
- the appropriate mix of technologies in use
- statistical techniques
- how to interpret datasets that are small, highly skewed or incomplete

# 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?

# Internet Resources

## Information

- UNFCCC
  - Policy and introductory material
  - <http://unfccc.int/>
- IPCC
  - Scientific and socio-economic material, assessment reports, methods...
  - <http://www.ipcc.ch/>
  - <http://www.ipcc-nggip.iges.or.jp/>

## Emission Inventories

- UNECE – TFEIP Guidebook
  - <http://reports.eea.eu.int/EMEPCORINAIR4/en>
- USEPA
  - <http://www.epa.gov/>

## Models

- Road Transport
  - COPERT <http://vergina.eng.auth.gr/mech/lat/copert/copert.htm>
  - TREMOVE MOBILE 6 <http://www.epa.gov/otaq/m6.htm>
- Information about Aviation Models
  - SAGE  
[http://www.faa.gov/about/office\\_org/headquarters\\_offices/aep/models/sage/](http://www.faa.gov/about/office_org/headquarters_offices/aep/models/sage/);
  - AERO2K  
<http://www.cate.mmu.ac.uk/aero2k.asp>
- Petroleum Industry
  - API <http://ghg.api.org>
  - API Compendium is available at:
    - <http://api-ec.api.org/policy/index.cfm>



## National Programs

- United States Environmental Protection Agency (2004), *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2003*, National Service Center for Environmental Publications (NSCEP)  
<http://www.epa.gov/globalwarming/publications/emissions>
- North American Carbon Program Plan (NACP). (2002). 'A report of the Committee of U.S. Carbon Cycle Science Steering Group.' Chapter 2, Major Elements of the North American Carbon Program Plan.  
<http://www.esig.ucar.edu/nacp/chap2.html>.
- USEPA NONROAD, 1999  
<http://www.dieselnet.com/standards/us/offroad.html>
- USEPA, 1992 and 1998 <http://www.epa.gov/otaq/locomotv.htm>

## International Organisations

- Energy statistics
  - <http://www.iea.org/Textbase/stats/index.asp>
- Agriculture and Forestry
  - <http://www.fao.org/>
- ISO – background for QA/QC; standards, etc.
  - <http://www.iso.org/>
- Convention on Long-Range Transboundary Air Pollution
  - <http://www.unece.org/env/lrtap/welcome.html>
- ICAO; aviation
  - <http://www.icao.int/>

## Global Emission Inventories

- Emission Database for Global Atmos. Research (EDGAR)
  - <http://www.rivm.nl/edgar/introduction/>
- Global Emission Inventory Activity (GEIA)
  - <http://www.geiacenter.org/>

# Task Force on National Greenhouse Gas Inventories



**Any questions?**

Please note:

All images in this presentation file © IPCC unless otherwise noted.