Uncertainty Levels associated with Input Data used in National Greenhouse Gas Inventories from developing countries

25-29 October 2021

GHG Support Unit, Transparency Division

MONDAY

1. Overview of Uncertainty Analysis in National GHG Inventories

2. Causes of uncertainties associated with input data used in National GHG Inventories

3. How to reduce the uncertainty associated with input data used in National GHG Inventories

TUESDAY

4. Uncertainty associated with the use of national statistics, surveys, censuses and sampling

Practical examples in Energy, IPPU, AFOLU and Waste

WEDNESDAY

5. Uncertainty associated with the Use of Empirical Data

Practical examples in Energy, IPPU, AFOLU and Waste

THURSDAY

6. Uncertainty associated with the use of proxy, splicing techniques and expert judgment to fill data gaps

Practical examples in Energy, IPPU, AFOLU and Waste

FRIDAY

7. Selecting Probability Density Functions (PDF) and addressing correlation

Practical examples in Energy, IPPU, AFOLU and Waste

Uncertainty overview [1]

General approach

Uncertainty

Lack of knowledge of the true value of a variable that can be described as a probability density function (PDF). Uncertainty depends on the analyst's state of knowledge, which in turn depends on the quality and quantity of applicable data as well as knowledge of underlying processes and inference methods.

Uncertainty analysis

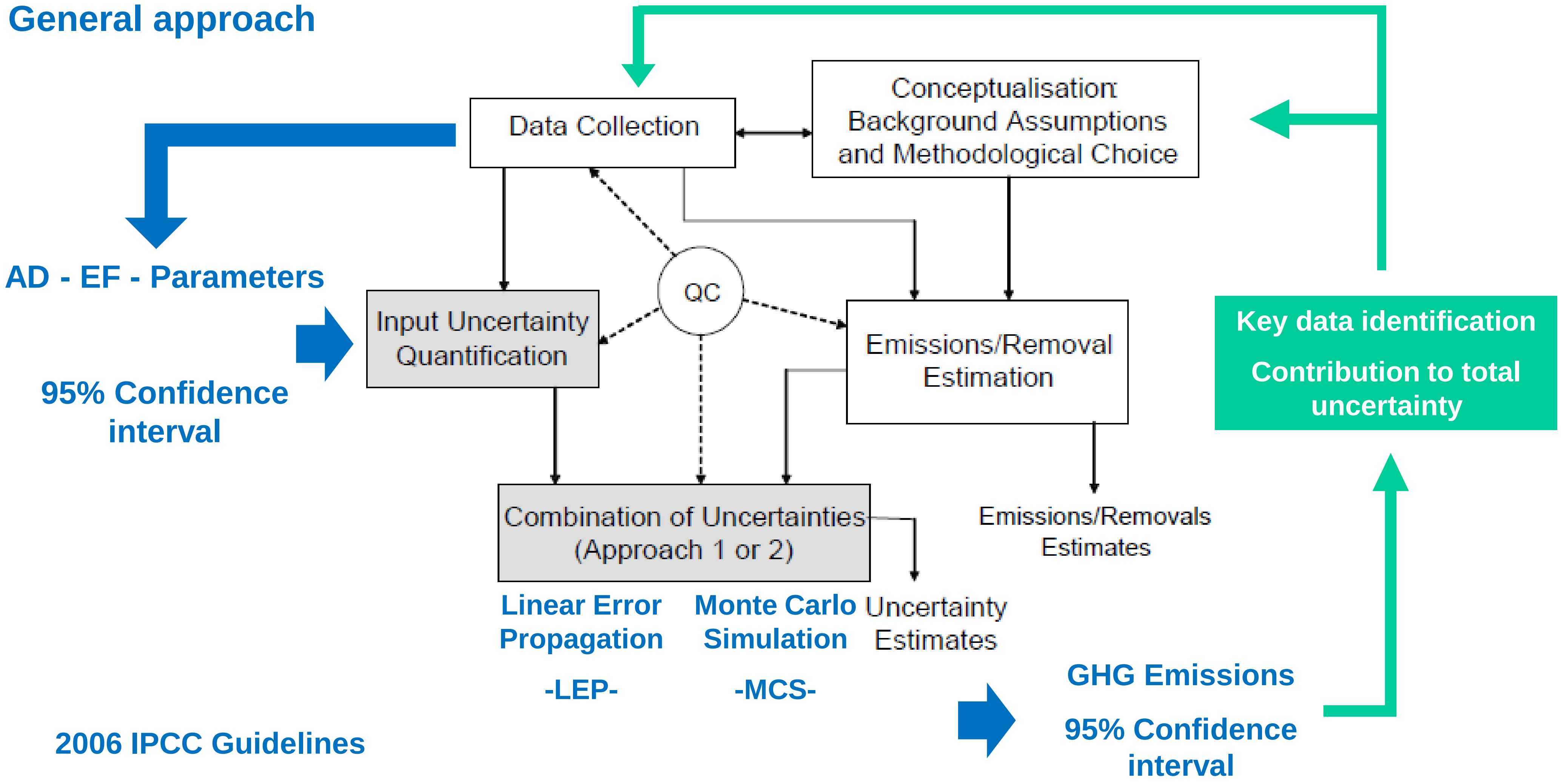
An uncertainty analysis should be seen, first and foremost, as a means to help prioritise national efforts to reduce the uncertainty of inventories in the future, and guide decisions on methodological choice.
Quantitative uncertainty analysis is performed by estimating the 95 percent confidence interval of the emissions and removals estimates for individual categories and for the total inventory

Uncertainty assessment

The term "ASSESSMENT" is intended to convey an exercise that includes the investigation of quantitative and qualitative aspects. In the glossary to the Guidelines, "uncertainty analysis" is defined as only a quantitative exercise.

Uncertainty overview [2]

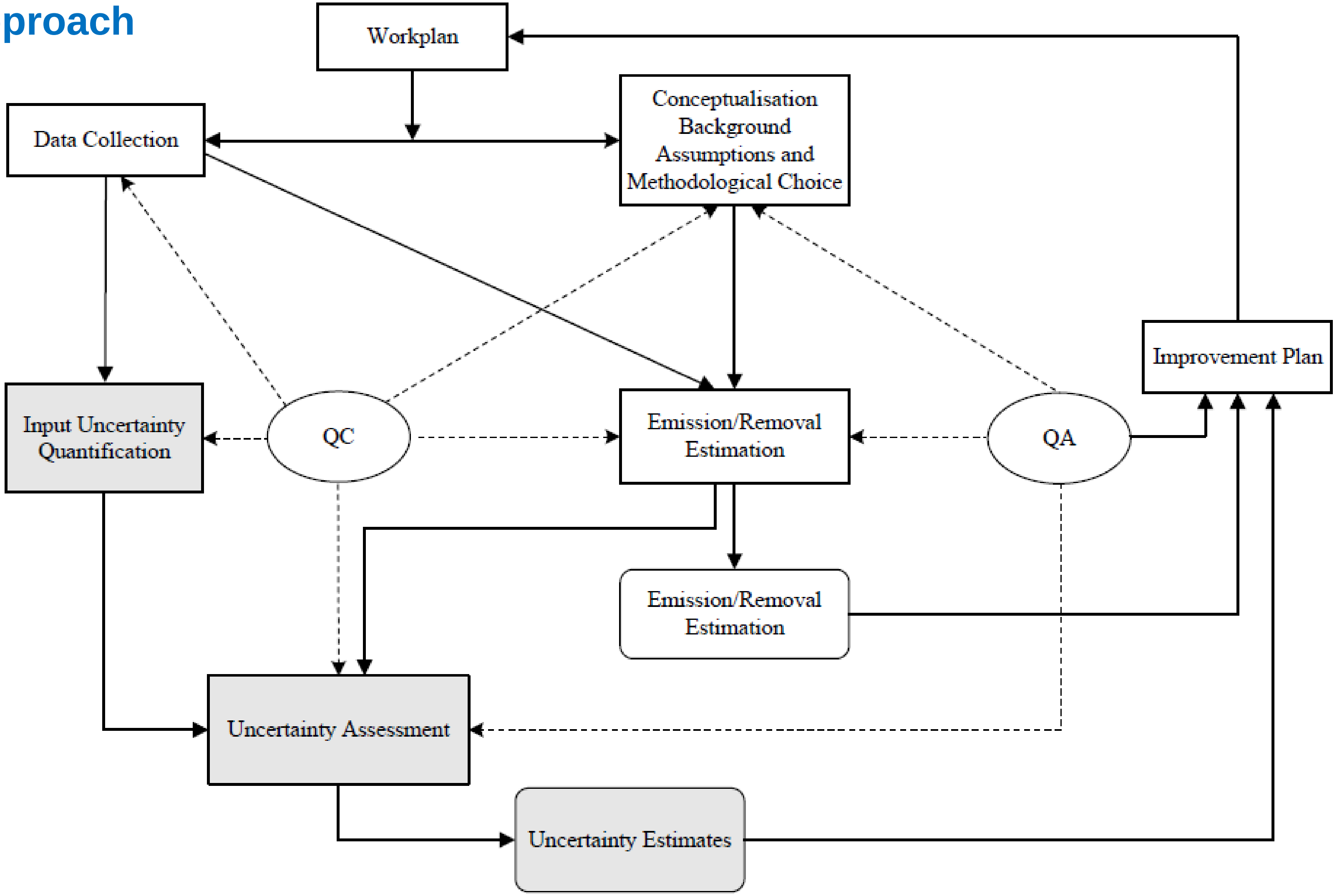
General approach



2006 IPCC Guidelines

Uncertainty overview [3]

General approach



Uncertainty overview [4]

Key concepts

Confidence interval: range that encloses the true, but unknown value, with a determined confidence (probability). Typically, a 95 percent confidence interval is used in greenhouse gas inventories.
Alternative interpretation: Range that may safely be declared to be consistent with observed data or information

Probability Density Function (PDF): describes the range and relative likelihood of possible values.
For emission inventory, it is used to describe uncertainty in the estimate of a quantity that is a fixed constant whose value is not exactly known.

Sensitivity analysis: method to determine which of the input uncertainties to an inventory contributes most substantially to the overall uncertainty.

Uncertainty overview [5]

Key concepts

Accuracy: *Agreement between the true value and the average of repeated measured observations or estimates of* a variable.
An accurate measurement or prediction lacks bias or, equivalently, systematic error.

Bias / Systematic error: *Lack of accuracy. Bias (systematic error), can occur because of failure to capture all relevant processes involved or because the available data are not representative of all real-world situations, or because of instrument error.*

Precision: *Agreement among repeated measurements of the same variable.*
Better precision means less random error. Precision is independent of accuracy.

Random errors: *Random variation above or below a mean value. Random error is inversely proportional to* precision. Usually, the random error is quantified with respect to a mean value, but the mean could be biased or unbiased. Thus, random error is a distinct concept compared to systematic error.

Uncertainty overview [6]

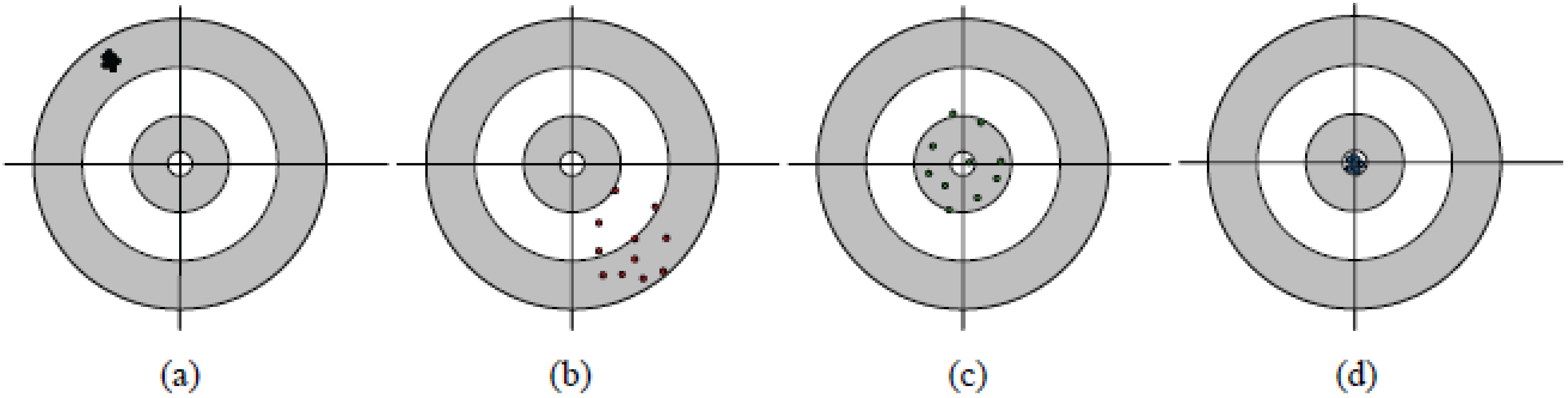
Key concepts

Lack of knowledge of the true value

How far is the true value from the value used?

Accuracy (systematic errors o bias) vs. Precision (random errors)

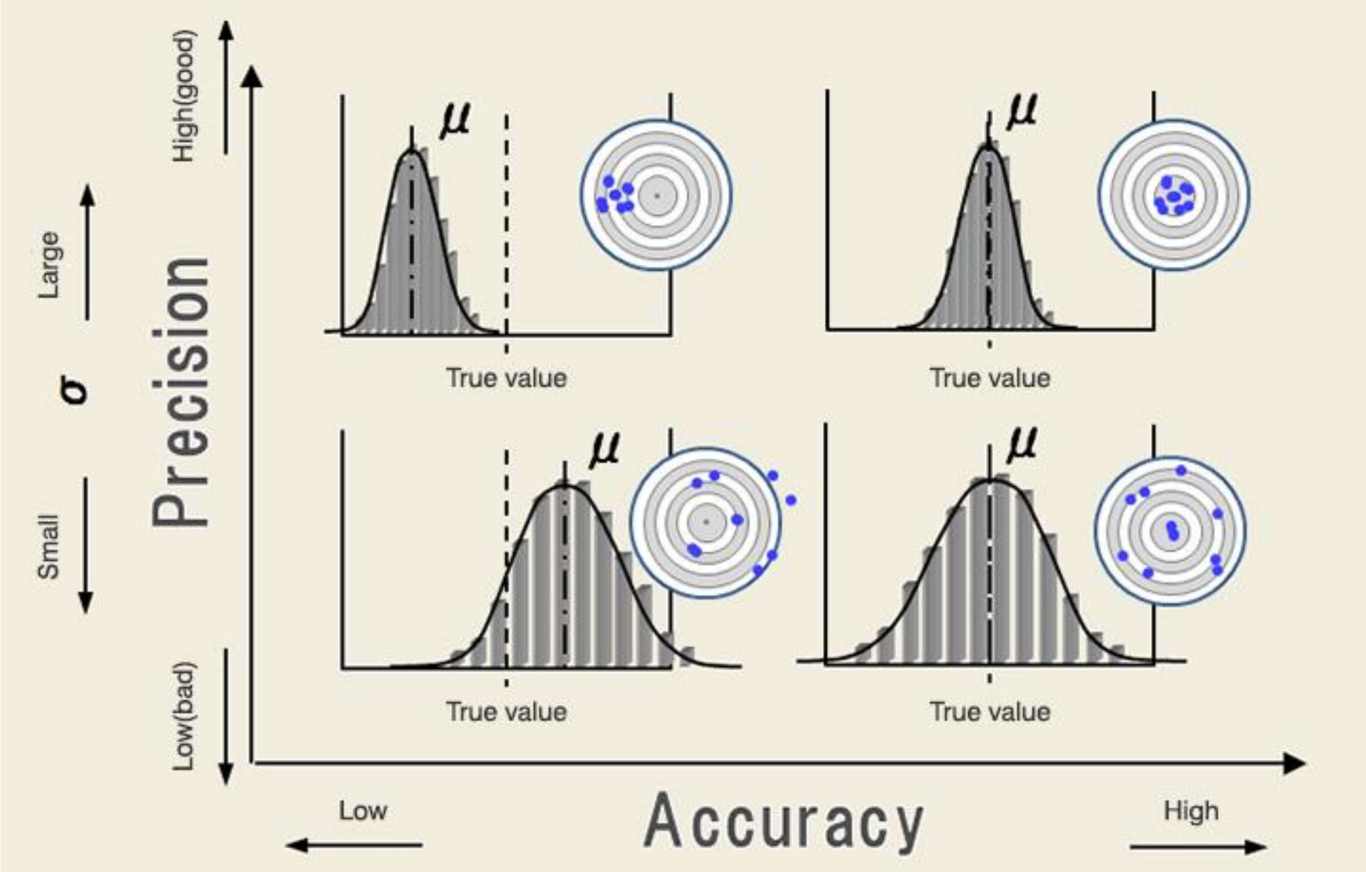
(a) inaccurate but precise; (b) inaccurate and imprecise; (c) accurate but imprecise; and (d) precise and accurate



Uncertainty overview [7]

Key concepts

Accuracy (systematic errors or bias) vs. Precision (random errors)



Uncertainty overview [8]

Key concepts

The **quantitative uncertainty analysis** tends to deal primarily with **random errors** based on the **inherent variability** of a system and the **finite sample size** of available data, **random components of measurement error**, or **inferences** regarding the random component of uncertainty obtained from **expert judgment**

→ **PRECISION**

It is important to recognize that **some uncertainties that are not addressed by statistical means may exist**, including those arising from **omissions or double counting**, or other **conceptual errors**, or from **incomplete understanding** of the processes that may lead to inaccuracies in estimates developed from models.

Bias or systematic errors

→ **ACCURACY**

Uncertainty overview [9]

Terminology

i) **Confidence interval**: range that encloses the true value with a determined confidence (probability)

95 % CI → 2.5 th [……………………..] 97.5 th

Absolute | Relative

Uncertainty

U_i

Symmetrical: 80 (64-96) 80 ± 16 80 ± 20%

64 [--------80--------] 96

$\underbrace{}_{16}\underbrace{}_{16}$

16/80 = 0.2

Not symmetrical: 80 (72-104) 80 -8/+24 80 -10%/+30%

72 [-----80--------------] 104

Large uncertainty:

100 (50-200) 100 -50/+100 100 -50%/+100% 100 ± factor of 2 (half or double)

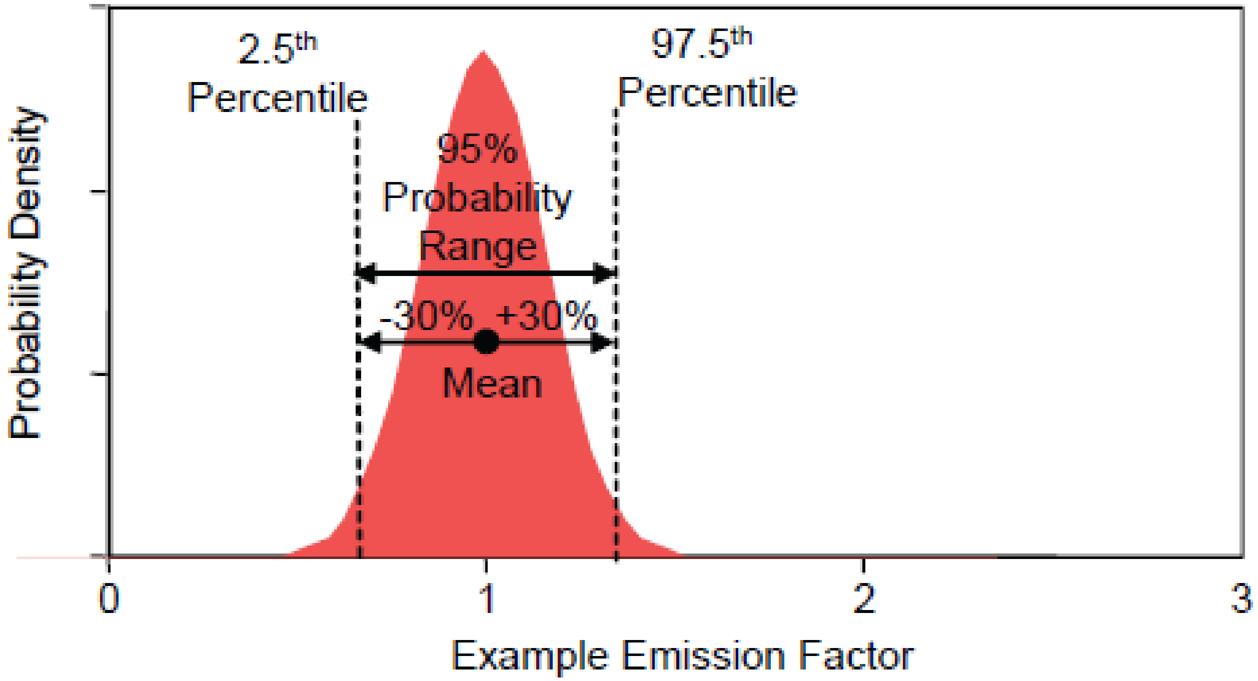
Factor of "n" (mean/n) and (mean x n)

Uncertainty overview [10]

Terminology

ii) Probability Density Function: range and relative likelihood of possible values

Symmetrical (normal distribution)

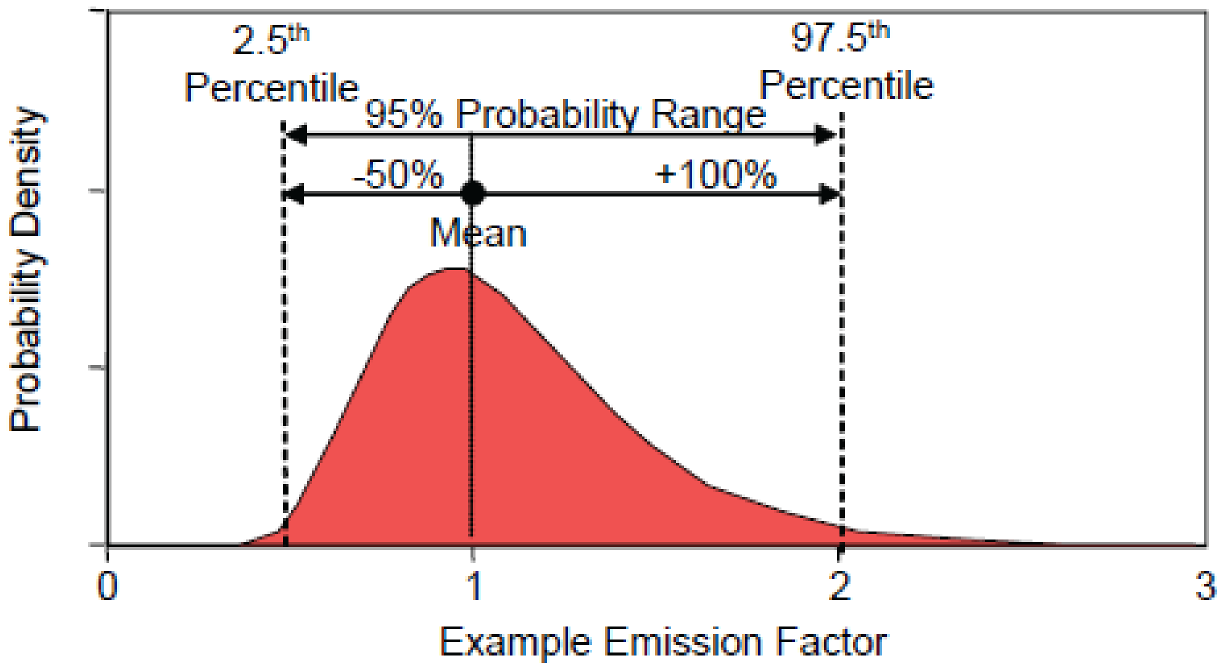


Uncertainty overview [11]

Terminology

ii) Probability Density Function: range and relative likelihood of possible values

Not symmetrical (Skewed)



Causes of uncertainty [1]

Causes of uncertainty: Animal population? Age? Livestock characterization? Diet?

Causes of uncertainty [2]

Causes of uncertainty: Land representation? Stand volume? Carbon stock? Below ground biomass?

Causes of uncertainty [3]

Causes of uncertainty: Pipe length? No. of fittings? Gas composition? Maintenance? Venting?

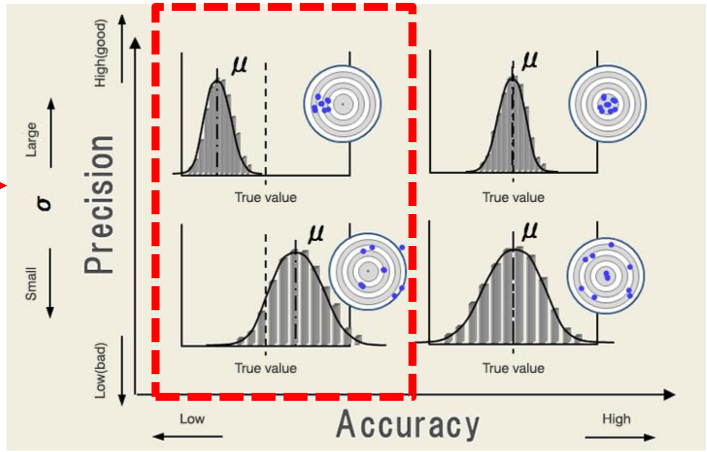
Causes of uncertainty [4]

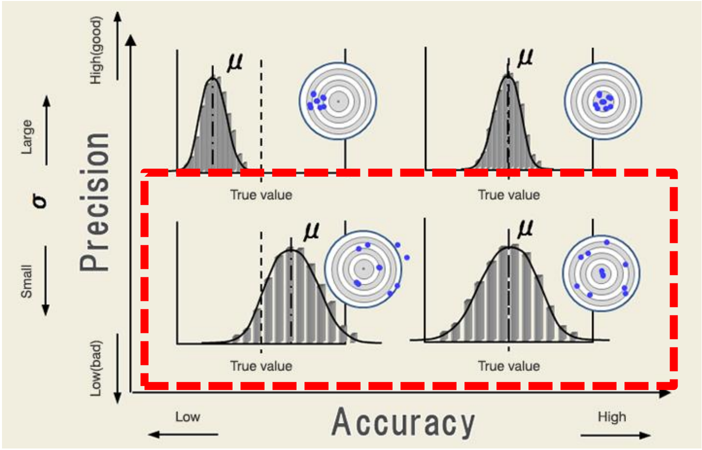
Causes of uncertainty: Waste generation? Composition? Climate? Treatment? Management type?

Causes of uncertainty [5]

Causes of uncertainty

Lack of completeness	-Bias
Model	-Bias and Random errors
Lack of data	-Bias and Random errors
Representativeness of data	-Bias
Random sampling error	-Random errors
Measurement errors	-Bias and Random errors
Misreporting	-Bias
Data gaps	-Bias and Random errors





Causes of uncertainty [6]

Causes of uncertainty & examples

Lack of completeness **-Bias**

e.g. sources/sink categories not included in the inventories

Energy: emissions from coke production

IPPU: fluorinated gases not reported

AFOLU: methane emissions from enteric fermentation for dairy cows.

Waste: industrial wastewater for some products not included

Causes of uncertainty [7]

Causes of uncertainty & examples

Lack of data **-Bias and Random errors**

e.g. activity data obtained by interpolation or other methods for missing year

Energy: provisional information in National Energy Balance for recent year

IPPU: cement production missing in national statistics

AFOLU: headcount for some animal category not informed in the statistics

Waste: amount of MSW surrogated from population, extrapolated from census

Causes of uncertainty [8]

Causes of uncertainty & examples

Representativeness of data **-Bias**

e.g. emission factor based on particular conditions

Energy: N_2O from internal combustion engines from laboratory test instead of real driving conditions

IPPU: CO_2 emissions form steel production based on full load capacity plant

AFOLU: biomass growth rate based on sampling that do not cover all relevant regions

Waste: wastewater treatment efficiency based on newly built plants data

Causes of uncertainty [9]

Causes of uncertainty & examples

Random sampling error **-Bias and Random errors**

e.g. activity data or emission factors based on limited sampling

- Energy: limited reporting in census of liquid fuels used in transport
- IPPU: amount of glass recovered (by type) based on surveys
- AFOLU: C stored in forest based on limited sampling capacity (few trees)
- Waste: MSW treatment distribution based on information from few cities in the country

Reducing uncertainty [1]

Causes of uncertainty & strategies

Lack of completeness	Bias	Concept, QA/QC
Model	Bias and Random errors	Concept, QA/QC
Lack of data	Bias and Random errors	Experts, QA/QC
Representativeness of data	Bias	QA/QC, verification
Random sampling error	Random errors	Statistics sizes
Measurement errors	Bias and Random errors	QA/QC, verification
Misreporting	Bias	QA/QC
Data gaps	Bias and Random errors	Statistics, experts

Reducing uncertainty [2]

Improving accounting

Improving conceptualization	Structural assumptions
Improving models	Structure and parameterization
Improving representativeness	Sampling strategies
Using + precise measurement methods	Measurement technologies
Collecting more measured data	Sample size
Eliminating known risk of bias	Following decision trees
Improving state of knowledge	Understanding of the categories

Reducing uncertainty [3]

Improving accounting

Improving conceptualization	**Structural assumptions**
Improving models	**Structure and parameterization**

e.g. better treatment of seasonality effects leading to more accurate annual estimates of emissions or removals in AFOLU.

e.g. moving to higher Tiers in steel production to account for local data at plant level.

Reducing uncertainty [4]

Improving accounting

Improving representativeness **Sampling strategies**

Collecting more measured data **Sample size**

e.g. including emissions data for situations involving start-up or load changes, if frequent, instead of only full load operations.

e.g. perform stratified sampling in forest to account for different characteristics, climate and species

e.g. increasing the sample size for determination of soil organic carbon.

Reducing uncertainty [5]

Improving accounting

Using more precise measurement methods	**Measurement technologies**

e.g. collecting data using standardized measurement methods (i.e. ISO)

e.g. using measured parameters instead of simplified assumptions (density, temperature, mass vs. volume)

Eliminating known risk of bias	**Following decision trees, expert knowledge**
Improving state of knowledge	**Understanding of the categories**

e.g. Verifying the correct positioning and calibration of instruments in gas measurement.

e.g. moving to higher tiers to account for national conditions

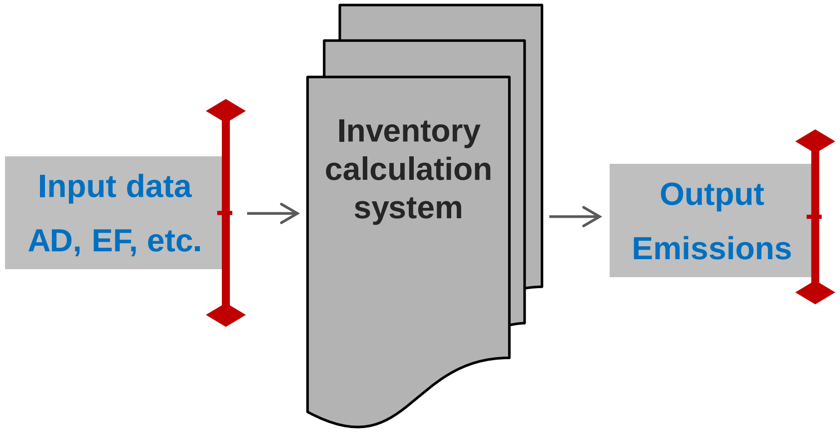
e.g. Involving producers to better understand the details and appropriateness of assumptions.

Reducing uncertainty [6]

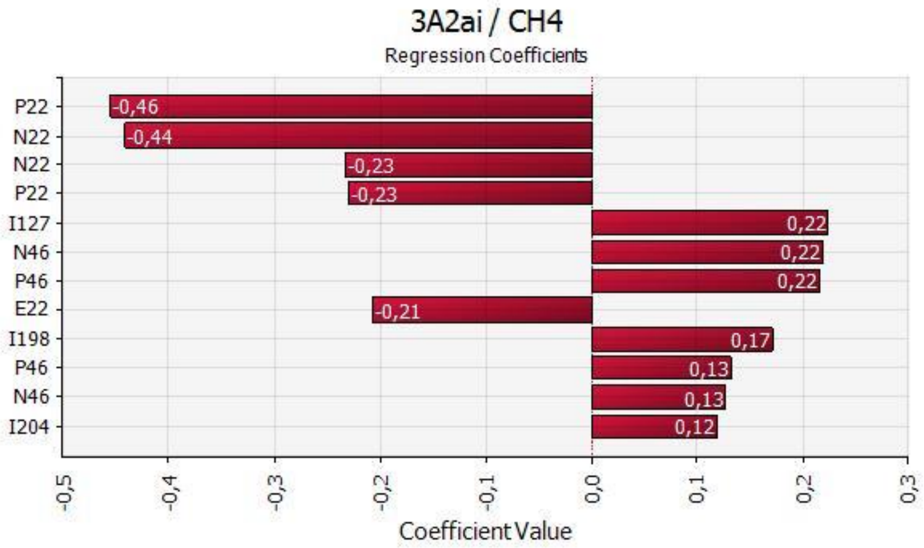
Sensitivity analysis

Purpose: Identify categories, and key variables used that contribute the most to overall uncertainty of the inventory. → **ALLOCATE RESOURCES**

How: Introduce a perturbation to one variable, of the magnitude of its uncertainty, and assess the variation in the result, one at a time (ceteris paribus).



e.g. Sensitivity in CH_4 emission from manure



Uncertainty overview

Conclusions

- It is a means to help prioritise national efforts to reduce the uncertainty of inventories in the future
- It guides decisions on methodological choice
- It helps understand the quality of the information use
- It is a requirement of GHG Inventories

Assessment of uncertainty in the input parameters should be part of the data collection !

End of day 1

Thank you!

Diego M. Ezcurra

MONDAY

1. Overview of Uncertainty Analysis in National GHG Inventories

2. Causes of uncertainties associated with input data used in National GHG Inventories

3. How to reduce the uncertainty associated with input data used in National GHG Inventories

TUESDAY

4. Uncertainty associated with the use of national statistics, surveys, censuses and sampling

Practical examples in Energy, IPPU, AFOLU and Waste

20%

Uncertainty associated with activity data [1]

National statistics: census, survey

- Activity data are closely linked to economic activity
- well established price incentives and fiscal requirements for accurate accounting

National statistics

Censuses	Surveys
Counting every instance of a particular activity. Typically includes both systematic and random errors.	Counting a portion of a particular activity. Typically includes both systematic and random errors.
Random errors are typically normally distributed and uncorrelated	**Random errors are typically normally distributed and uncorrelated**

V1_Ch2. - Annex 2A.2
2006 IPCC Guidelines
General guidance on performing surveys

Contact the statistical agencies directly

Uncertainty associated with activity data [2]

National statistics: census

National statistics

Censuses

Surveys

- Check for the size of random errors, look for fluctuations over time, and differential fluctuations in series that ought to be highly correlated with the data of interest
- To check for bias errors, cross-check the data of interest with other, related information. (look up and down the supply chain for fuels, or highly correlated activities with the data of interest, for instance reported fuel input vs. electricity output).
- Interpretation of statistical differences, within, for instance, national energy data are an example of cross checking (i.e. reference approach)..

Often ‘precise but inaccurate’

Uncertainty associated with activity data [3]

National statistics: census

National statistics

Censuses

Surveys

Hands-on exercises

Uncertainty associated with activity data [4]

National statistics: census

Periodic publications

ENERGY: liquid fuels commercialization. Fuel consumption statistics are published by the Ministry of Energy every year for gasoil and gasoline. The statistics contains the most updated information for the current and previous years.

Evaluate consistency and identify fluctuations over time in series to derive the uncertainty of the data

Year	Gasoil	Gasoline
2012	**659.034**	**479.291**
2013	**662.157**	**504.563**
2014	**666.065**	**533.358**
2015	**770.377**	**599.242**
2016	**866.303**	**554.953**
2017		
2018		

Source: commercialization tables, 12 July 2017

Year	Gasoil	Gasoline
2012	**659.034**	**479.291**
2013	**662.157**	**504.563**
2014	**697.641**	**533.358**
2015	**770.377**	**599.242**
2016	**866.303**	**554.953**
2017	**847.566**	**498.429**
2018		

Source: commercialization tables, 18 June 2018

Year	Gasoil	Gasoline
2012	**659.034**	**479.291**
2013	**662.157**	**504.563**
2014	**697.641**	**533.358**
2015	**770.377**	**599.242**
2016	**866.303**	**554.953**
2017	**847.566**	**480.723**
2018	**884.250**	**426.849**

Source: commercialization tables, 23 June 2019

Uncertainty associated with activity data [4]

National statistics: census

4,5% | U ≈ 5% | 3,7%

Periodic publications

ENERGY: liquid fuels commercialization. Fuel consumption statistics are published by the Ministry of Energy every year for gasoil and gasoline. The statistics contains the most updated information for the current and previous years.

Evaluate consistency and identify fluctuations over time in series to derive the uncertainty of the data

Year	Gasoil	Gasoline
2012	**659.034**	**479.291**
2013	**662.157**	**504.563**
2014	**666.065**	**533.358**
2015	**770.377**	**599.242**
2016	**866.303**	**554.953**
2017		
2018		

Source: commercialization tables, 12 July 2017

Year	Gasoil	Gasoline
2012	**659.034**	**479.291**
2013	**662.157**	**504.563**
2014	**697.641**	**533.358**
2015	**770.377**	**599.242**
2016	**866.303**	**554.953**
2017	**847.566**	**498.429**
2018		

Source: commercialization tables, 18 June 2018

Year	Gasoil	Gasoline
2012	**659.034**	**479.291**
2013	**662.157**	**504.563**
2014	**697.641**	**533.358**
2015	**770.377**	**599.242**
2016	**866.303**	**554.953**
2017	**847.566**	**480.723**
2018	**884.250**	**426.849**

Source: commercialization tables, 23 June 2019

Uncertainty associated with activity data [5]

National statistics: census

Highly correlated activities
ENERGY: fuel consumption vs electricity generation. Fuel consumption and electricity generation are reported every year by the electricity grid administrator.
Evaluate the reported data, calculate the efficiency to identify any inconsistency and assess uncertainty.

	Total
Gas Turbine Generation (MWh)	**59.757.516**
Natural Gas Consumption (10^3 m^3)	**14.975.637**

42%

NCV: 48 TJ/Gg

Density: 0,714 kg/m^3

efficiency ?

Uncertainty associated with activity data [6]

National statistics: census

Highly correlated activities
ENERGY: fuel consumption vs electricity generation. Fuel consumption and electricity generation are reported every year by the electricity grid administrator.
Evaluate the reported data, calculate the efficiency to identify any inconsistency and assess uncertainty

	January	February	March	April	May	June	July	August	September	October	November	December	Total
GT (MWh)	5.155.358	4.554.661	5.010.912	4.657.051	4.714.604	5.288.447	5.432.010	5.592.661	5.316.355	4.290.471	4.512.174	5.232.812	59.757.516
NG (10^3 m^3)	1.711.896	1.375.323	1.480.325	1.331.261	1.345.921	1.572.943	157.041	1.674.304	1.632.617	1.376.352	1.323.954	1.622.381	14.975.637
	0,33	0,30	0,30	0,29	0,29	0,30	0,03	0,30	0,31	0,32	0,29	0,31	0,25

NCV: 48 TJ/Gg

Density: 0,714 kg/m^3

Monthly efficiency ?

	January	February	March	April	May	June	July	August	September	October	November	December	Total
Generation (TJ)	18.559	16.397	18.039	16.765	16.973	19.038	19.555	20.134	19.139	15.446	16.244	18.838	**215.127**
Consumption (TJ)	58.670	47.135	50.734	45.625	46.127	53.908	5.382	57.382	55.953	47.170	45.375	55.602	**513.245**
Efficiency	32%	35%	36%	37%	37%	35%	363%	35%	34%	33%	36%	34%	**42%**

Uncertainty associated with activity data [7]

National statistics: census

Bias ≈ 20%

Highly correlated activities
ENERGY: fuel consumption vs electricity generation. Fuel consumption and electricity generation are reported every year by the electricity grid administrator.
Evaluate the reported data, calculate the efficiency to identify any inconsistency and assess uncertainty

	January	February	March	April	May	June	July	August	September	October	November	December	Total
GT (MWh)	5.155.358	4.554.661	5.010.912	4.657.051	4.714.604	5.288.447	5.432.010	5.592.661	5.316.355	4.290.471	4.512.174	5.232.812	59.757.516
NG (10^3 m^3)	1.711.896	1.375.323	1.480.325	1.331.261	1.345.921	1.572.943	157.041	1.674.304	1.632.617	1.376.352	1.323.954	1.622.381	14.975.637
							1.570.410						18.017.687

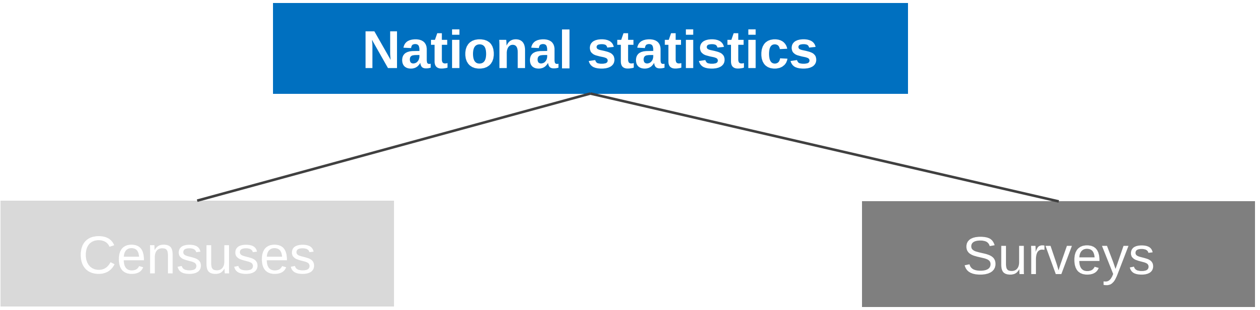
NCV: 48 TJ/Gg

Density: 0,714 kg/m^3

	January	February	March	April	May	June	July	August	September	October	November	December	Total
GT (MWh)	5.155.358	4.554.661	5.010.912	4.657.051	4.714.604	5.288.447	5.432.010	5.592.661	5.316.355	4.290.471	4.512.174	5.232.812	**59.757.516**
NG (10^3 m^3)	1.711.896	1.375.323	1.480.325	1.331.261	1.345.921	1.572.943	1.570.410	1.674.304	1.632.617	1.376.352	1.323.954	1.622.381	**18.017.687**
	32%	35%	36%	37%	37%	35%	36%	35%	34%	33%	36%	34%	35%

Uncertainty associated with activity data [8]

National statistics: survey



- Sample size and inter-individual variability
- Typical cases: consumer surveys, home expenses survey, land use surveys or forest cover surveys
- The agency conducting the sample will normally be able to advise on sampling error.
- If there is no information available, it may be possible to identify, or infer, the sample and population sizes and calculate sampling error directly.

precision depends on sample size, accuracy depends on sampling design

Uncertainty associated with activity data [9]

National statistics: survey

Heterogeneity (standard deviation) vs. uncertainty in sample mean (standard error)

$$Uncertainty = \pm\left(\frac{1.96 \cdot \sigma}{\mu}\right) \cdot 100\%$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$$

$$Uncertainty = \pm\left(\frac{1.96 \cdot SE}{\mu}\right) \cdot 100\%$$

$$SE = \frac{\sigma}{\sqrt{n}*}$$

* n instead of n+1 for large samples

Heterogeneity (standard deviation)	Uncertainty in sample mean (standard error)
Variability of the sample	Variability of the mean of the sample
Applicable for individual value	Applicable for country average
Standard deviation tends to remain constant	Standard error falls as sample size grows

Uncertainty associated with activity data [10]

National statistics: survey

National statistics

Censuses

Survey

Hands-on exercises

Uncertainty associated with activity data [11]

National statistics: survey

AFOLU: Carbon stock in forest from surveys

a) Given a sample with 80 individual values, calculate the mean, standard deviation and standard error.
b) If the sampling is repeated ten times, calculate the mean for each sample and the standard deviation of the sampling distribution of the mean and compare with a).

Excel

	Samp. 1	Samp. 2	Samp. 3	Samp. 4	Samp. 5	Samp. 6	Samp. 7	Samp. 8	Samp. 9	Samp. 10
Mean	102	99	101	100	97	100	100	101	102	96
STD	18	21	18	22	20	19	19	20	20	21
SE	2,0	2,4	2,0	2,5	2,3	2,2	2,1	2,2	2,2	2,4

Mean	STD
101	2,1

Uncertainty associated with activity data [12]

National statistics: survey

AFOLU: Carbon stock in forest from surveys

Calculate the uncertainty of the carbon stock obtained from the sampling to be used in the GHG inventory. The emissions from land use change will be calculated for the entire country in the year in which the survey was carried out.

Sample size: 30

Average C stock: 93.7 tC/ha

Standard deviation: 10.2 tC/ha

$$Uncertainty = \pm\left(\frac{1.96 \cdot SE}{\mu}\right) \cdot 100\%$$

$$SE = \frac{\sigma}{\sqrt{n}} *$$

SE: 1.9 tC/ha

U: 4 %

σ or SE ?

What if the carbon stock is applied to account for emissions in one deforested area?

Uncertainty associated with activity data [13]

National statistics: survey

WASTE: Municipal solid waste amount and composition

1) The amount per capita is obtained from a sample that covers vehicles collecting in a wide range of areas: urban and rural, wealthy and poor, with and without gardens, etc. and covering several periods throughout the year.

Uncertainty for the entire MSW category? µ or SE ?

Uncertainty for the emissions from managed landfills? µ or SE ?

2) The composition of the MSW was done through a survey at different landfills. A national waste composition was calculated based on the results and is used to calculate the methane emissions in each individual landfill.

Uncertainty for the composition based on amount from each landfill? µ or SE ?

End of day 2

Thank you!

Diego M. Ezcurra

MONDAY

1. Overview of Uncertainty Analysis in National GHG Inventories

2. Causes of uncertainties associated with input data used in National GHG Inventories

3. How to reduce the uncertainty associated with input data used in National GHG Inventories

TUESDAY

4. Uncertainty associated with the use of national statistics, surveys, censuses and sampling

Practical examples in Energy, IPPU, AFOLU and Waste

WEDNESDAY

5. Uncertainty associated with the Use of Empirical Data

Practical examples in Energy, IPPU, AFOLU and Waste

40%

Uncertainty associated with empirical data [1]

Techniques for quantifying uncertainties

Measurement | Publications | Statistics

Statistical analysis

V1_Ch.3 – 3.2.2.2

2006 IPCC Guidelines

- Approach to quantify uncertainty in inventories, mainly associated to emission factors and other estimation parameters

- 6 steps approach to apply systematically

Uncertainty associated with empirical data [2]

Techniques for quantifying uncertainties

Statistical analysis

1. Compilation and evaluation of a database
2. Visualisation of data by developing empirical distribution functions
3. Fitting, evaluation, and selection of alternative PDF
4. Characterisation of uncertainty in the mean of the distributions for variability
5. Input to a probabilistic analysis to estimate uncertainty in total emissions
6. Sensitivity analysis

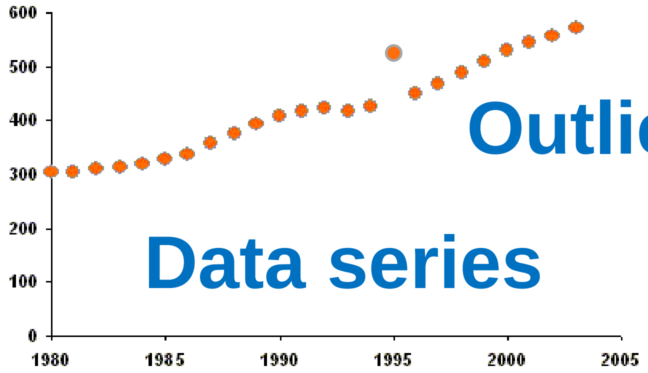
Uncertainty associated with empirical data [3]

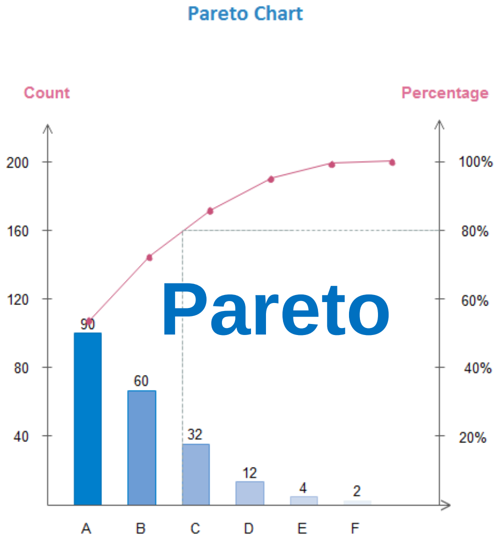
Techniques for quantifying uncertainties

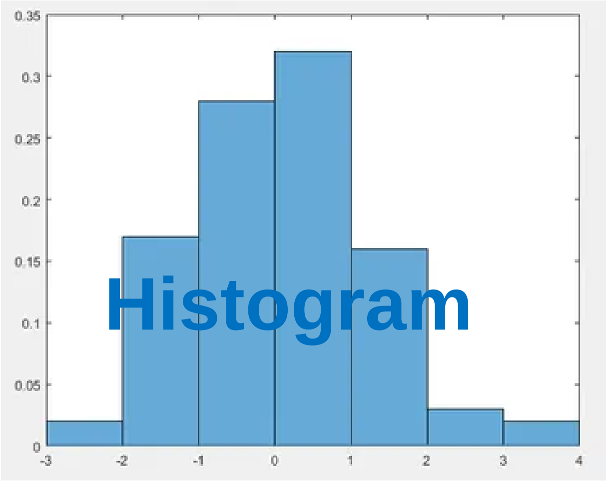
Statistical analysis

1. Compilation and evaluation of a database

2. Visualisation of data by developing empirical distribution functions







Consistency

Uncertainty associated with empirical data [4]

Techniques for quantifying uncertainties

Statistical analysis

3. Fitting, evaluation, and selection of alternative PDF

Possible PDF

Physical process

Expert judgement

Variability

Goodness of fit test

Uncertainty associated with empirical data [5]

Techniques for quantifying uncertainties

Statistical analysis

4. Characterisation of uncertainty in the mean of the distributions for variability
5. Input to a probabilistic analysis to estimate uncertainty in total emissions
6. Sensitivity analysis

σ or SE ? **95% CI** $\rightarrow$ U_i

U_i $\rightarrow$ U_{Total}

U_{Total} $\rightarrow$ X_i

Uncertainty associated with empirical data [6]

Techniques for quantifying uncertainties

Measurement

Publications

Statistics

Statistical analysis

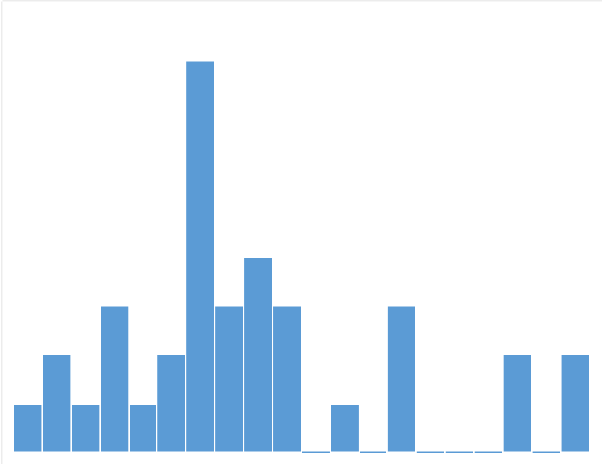
Hands-on exercises

Uncertainty associated with empirical data [7]

Statistical analysis

IPPU: Emission factor for aluminum production

A monitoring system was set to detect the anode effect. Average monthly EF were calculated for CF_4 and recorded for three years.
Perform statistical analysis to identify PDF candidates, calculate the mean, standard deviation, standard error and uncertainty.



μ: 64.3 g CF_4/tAl

σ: 13.8 g CF_4/tAl

SE: 2.3 g CF_4/tAl

U: 5 %

Uncertainty associated with empirical data [8]

Statistical analysis

Attention!

Measurements taken for another purpose may not be representative.

For example, methane measurements made for safety reasons at coal mines and landfills may not necessarily reflect total emissions because they may have been made only when methane emissions were suspected of being high, as a compliance check.

In such cases, the ratio between the measured data and total emissions should be estimated for the uncertainty analysis.

End of day 3

Thank you!

Diego M. Ezcurra

MONDAY

1. Overview of Uncertainty Analysis in National GHG Inventories

2. Causes of uncertainties associated with input data used in National GHG Inventories

3. How to reduce the uncertainty associated with input data used in National GHG Inventories

TUESDAY

4. Uncertainty associated with the use of national statistics, surveys, censuses and sampling

Practical examples in Energy, IPPU, AFOLU and Waste

WEDNESDAY

5. Uncertainty associated with the Use of Empirical Data

Practical examples in Energy, IPPU, AFOLU and Waste

THURSDAY

6. Uncertainty associated with the use of proxy, splicing techniques and expert judgment to fill data gaps

Practical examples in Energy, IPPU, AFOLU and Waste

60%

Proxy: Splicing techniques and Expert Judgement [1]

Splicing techniques for Data gaps

Data gaps

Periodic data

V1_Ch5 . 5.3 Resolving Data Gaps

2006 IPCC Guidelines

Changes and gaps in data availability

Data are available less frequently than annual or may not cover the entire country.
Estimates need to be updated each time new data become available.
New estimates should be extrapolated and then recalculated when new data become available

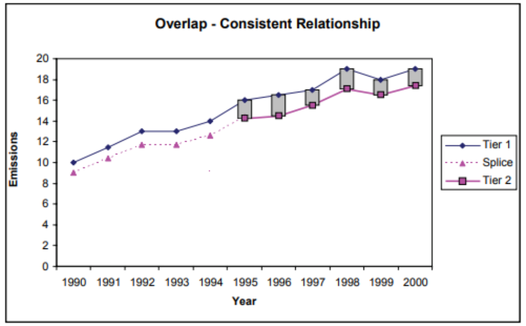
Changes in data collection systems or methodologies (improve or decrease or gaps)
Higher tier methods can be applied for recent years, but not for earlier years
Discontinuation of data sets or different definitions, classifications and levels of aggregation

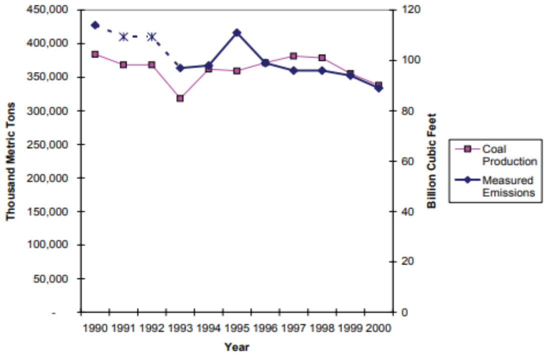
Splicing techniques

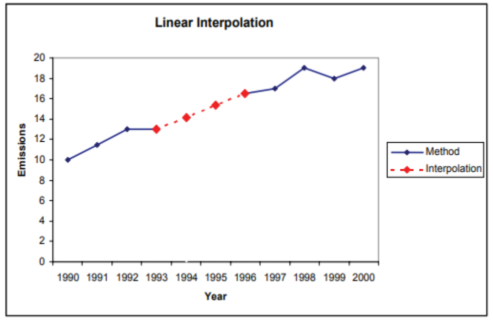
Proxy: Splicing techniques and Expert Judgement [2]

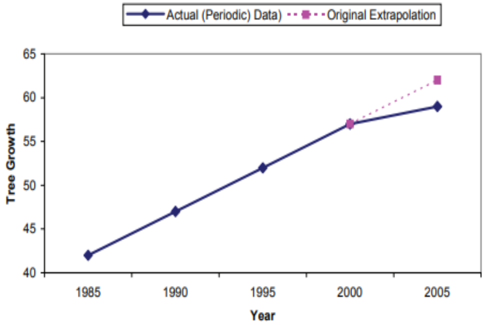
Splicing techniques for Data gaps

Overlap	Surrogate	Interpolation	Extrapolation
New method implemented not available before (or after)	No data for the category but other related available Ratios or regression	Data gap for some years or outliers identified	Data unavailable for last or first year





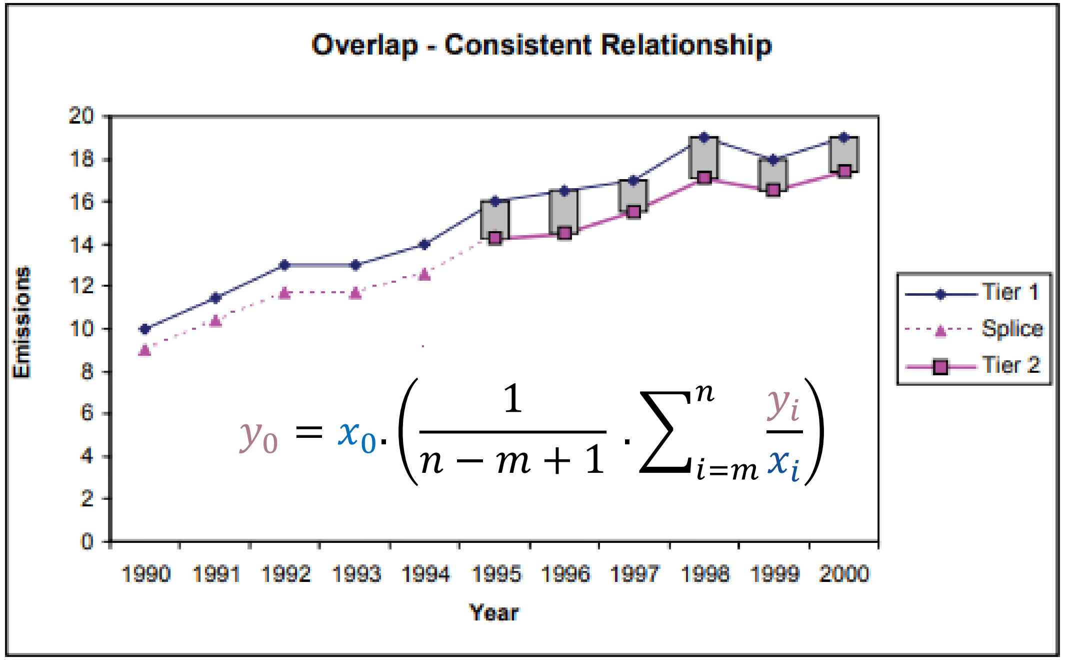




Proxy: Splicing techniques and Expert Judgement [3]

Splicing techniques for Data gaps

Overlap	Surrogate	Interpolation	Extrapolation



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

New method implemented not available before (or after)

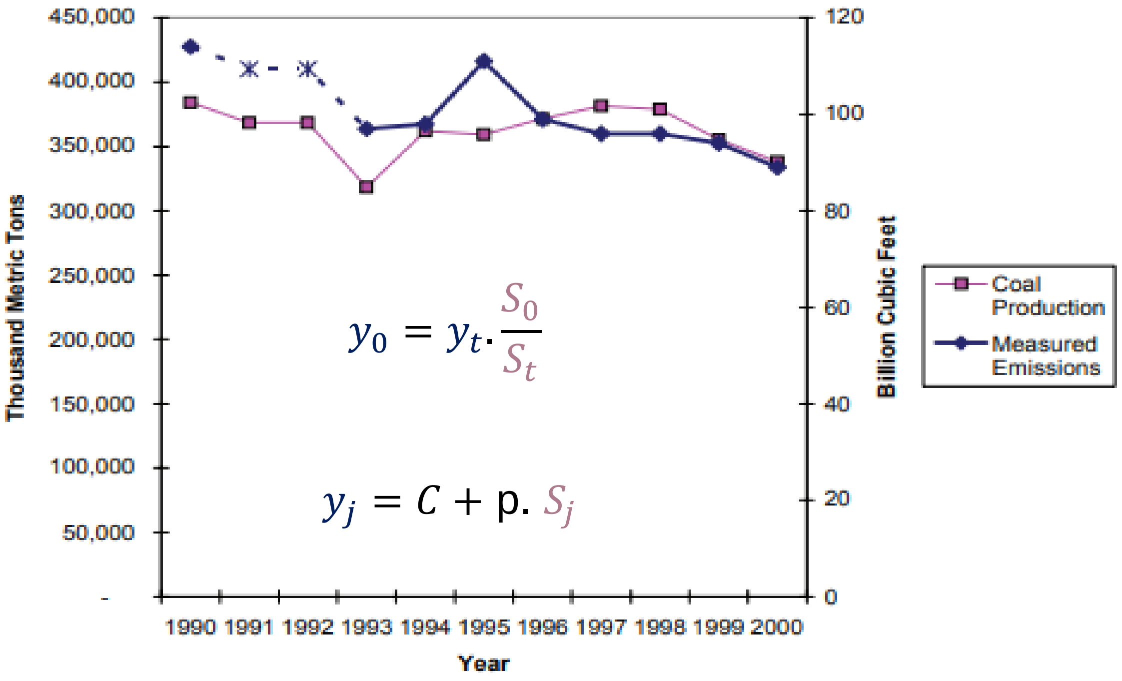
Requires consistent relationship between the two methods

Variability in ratios can be used to assess consistency and uncertainty

Proxy: Splicing techniques and Expert Judgement [4]

Splicing techniques for Data gaps

Overlap	**Surrogate**	Interpolation	Extrapolation



$$y_0 = y_t . \frac{S_0}{S_t}$$

$$y_j = C + \mathrm{p}. S_j$$

No data for the category but other related available

Ratios or regression analysis (simple, multiple, linear, non-linear)

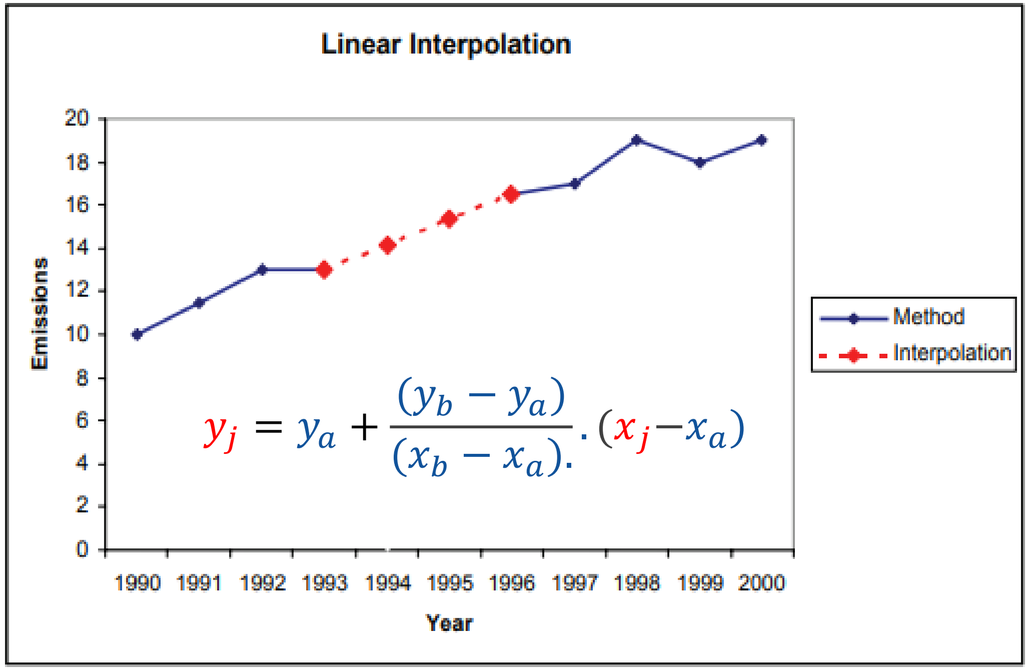
Requires identifying the dependent and independent variable/s as well a the dependency (correlation)

Typical error for the regression can be used to assess uncertainty (other criteria also possible)

Proxy: Splicing techniques and Expert Judgement [5]

Splicing techniques for Data gaps

Overlap | Surrogate | **Interpolation** | Extrapolation



Data gap for some years or outliers identified.

Linear models are commonly used but others may apply (quadratic)

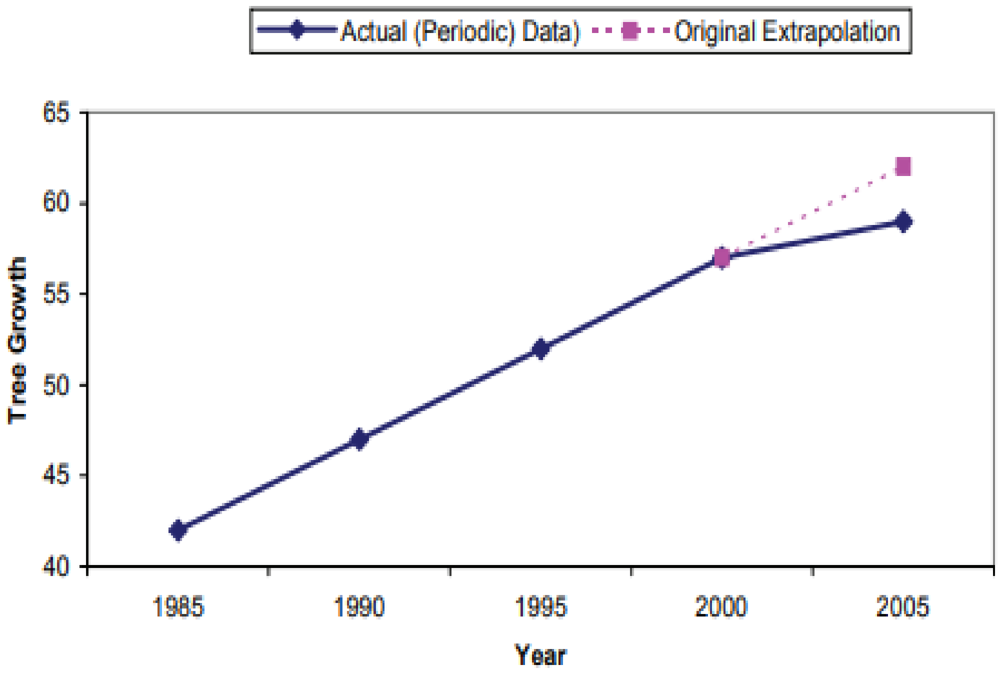
It can be applied intermittently as necessary. Requires smooth or stable trend (low variability). If not, surrogate is a better practice.

Variability in available data can be used to assess uncertainty

Proxy: Splicing techniques and Expert Judgement [6]

Splicing techniques for Data gaps

Overlap | Surrogate | Interpolation | **Extrapolation**



Data unavailable for last or first year/s

Continuation of the trend, if stable, or surrogate, if higher variability, beyond data´s period.

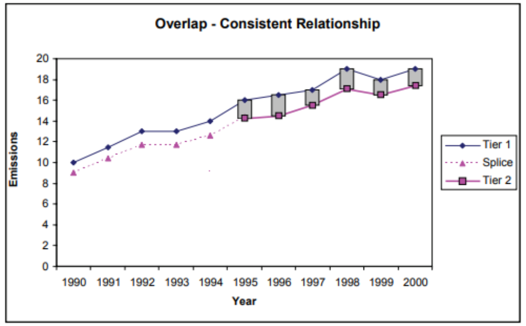
Not recommended for estimations over long periods of time

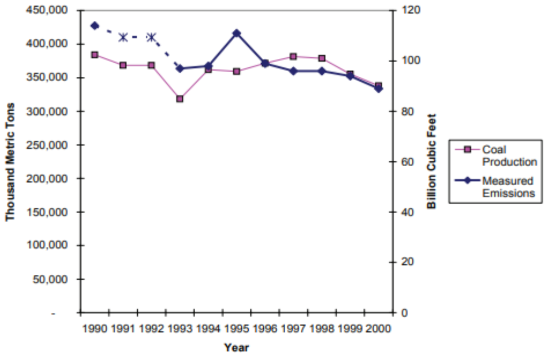
Uncertainty depending on the extrapolation method (trend or surrogate). Could be evaluated ex post if data becomes available.

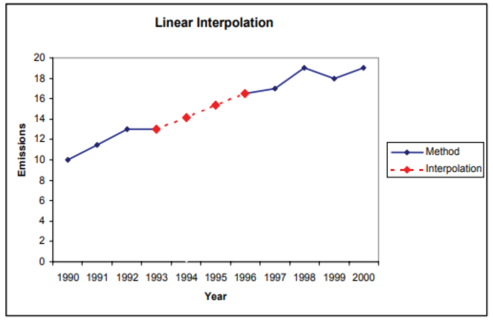
Proxy: Splicing techniques and Expert Judgement [7]

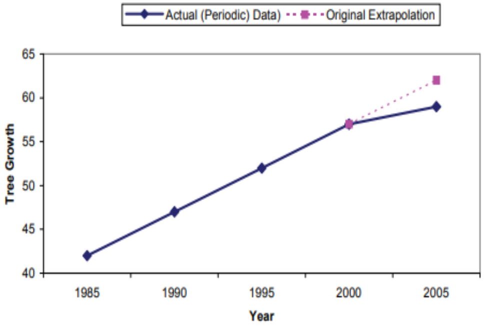
Splicing techniques for Data gaps

Overlap	Surrogate	Interpolation	Extrapolation









Hands-on exercises

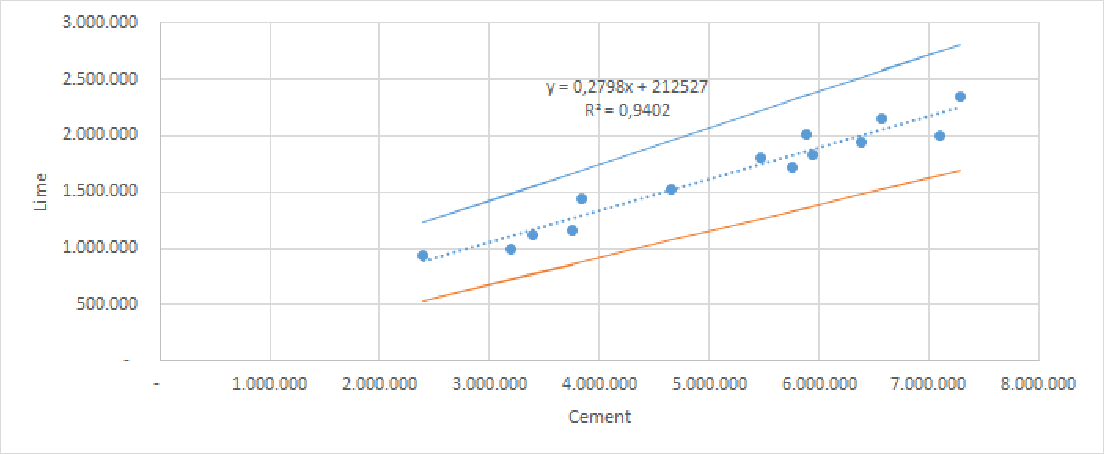
Proxy: Splicing techniques and Expert Judgement [8]

Splicing techniques for Data gaps

Overlap | **Surrogate** | Interpolation | Extrapolation

IPPU: Lime production statistics were not available for the last two years.
Investigate the relationship with the production of cement, surrogate lime production and determine the uncertainty of the estimation used.

Year	Lime	Cement
1	1.167.701	3.751.759
2	1.119.942	3.398.621
3	943.108	2.396.907
4	988.285	3.197.714
5	1.444.813	3.833.116
6	1.527.389	4.654.673
7	1.808.855	5.472.815
8	2.016.509	5.885.219
9	1.834.096	5.947.131
10	1.717.285	5.752.006
11	1.949.142	6.388.311
12	2.000.257	7.104.927
13	2.154.948	6.567.993
14	2.348.841	7.288.507
15		6.992.203
16		7.172.822



U: 15 %

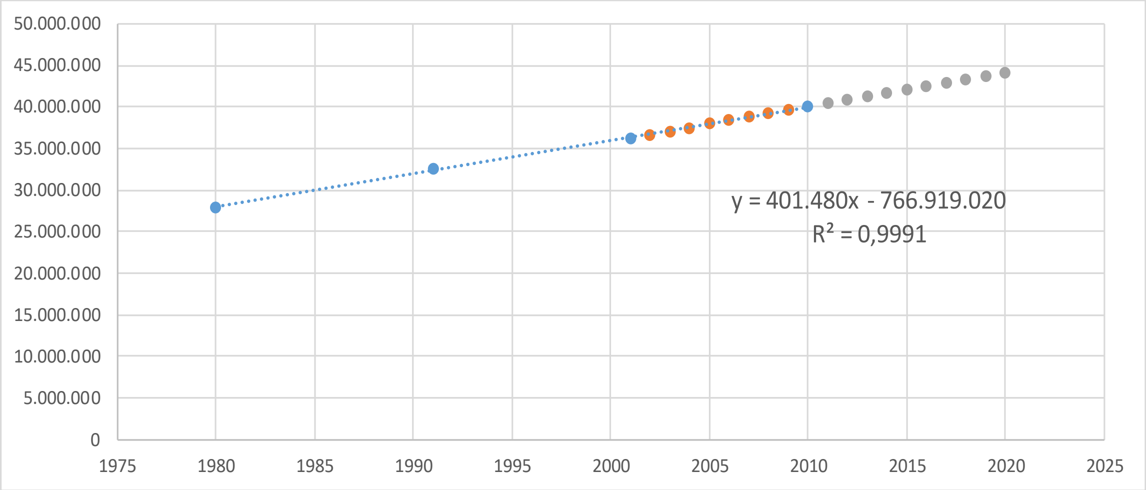
Proxy: Splicing techniques and Expert Judgement [9]

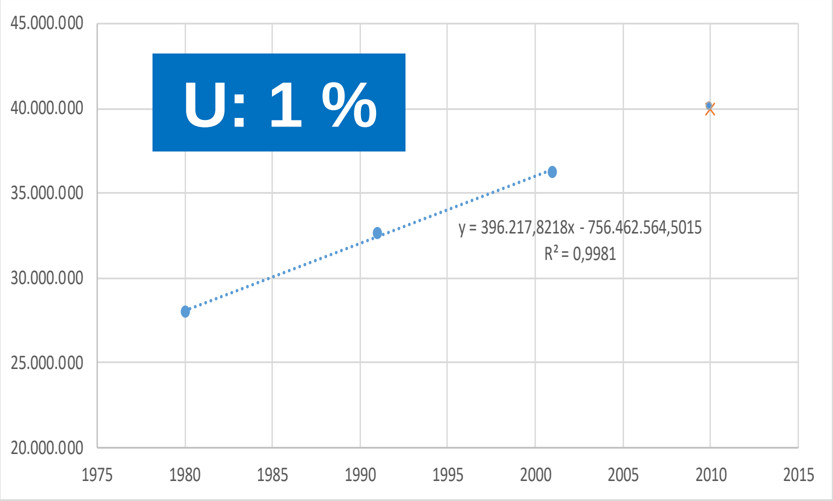
Splicing techniques for Data gaps

Overlap | Surrogate | Interpolation | Extrapolation

Waste: municipal solid waste generation per capita is used to calculate the amount of waste to be treated in a landfill that serves a specific region. Obtain the yearly population for the region using data from the census and asses the uncertainty. Extrapolate the data from the census to calculate the population in 2020 and asses the uncertainty.

Year	Population
1980	27.949.480
1991	32.615.528
2001	36.260.130
2010	40.117.096

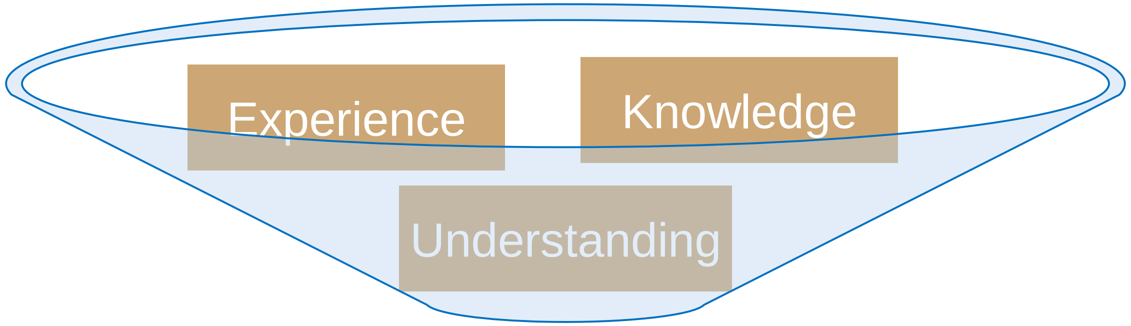




ENCODING EXPERT JUDGEMENT

Proxy: Splicing techniques and Expert Judgement [11]

ENCODING EXPERT JUDGEMENT



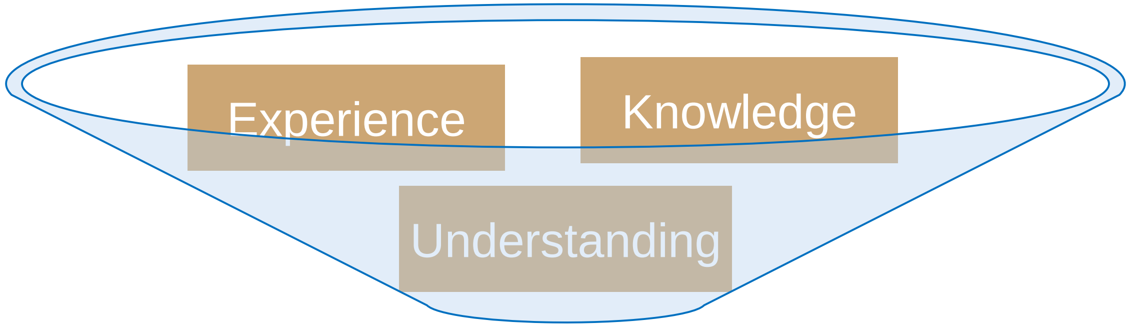
Key goal is to characterise the state of knowledge regarding possible values of a particular variable and to develop a PDF

Full range of values

- When no relevant empirical data is available
- Well informed judgements from domain experts
- Formal expert elicitation protocols

Proxy: Splicing techniques and Expert Judgement [12]

ENCODING EXPERT JUDGEMENT



Elicitation protocols

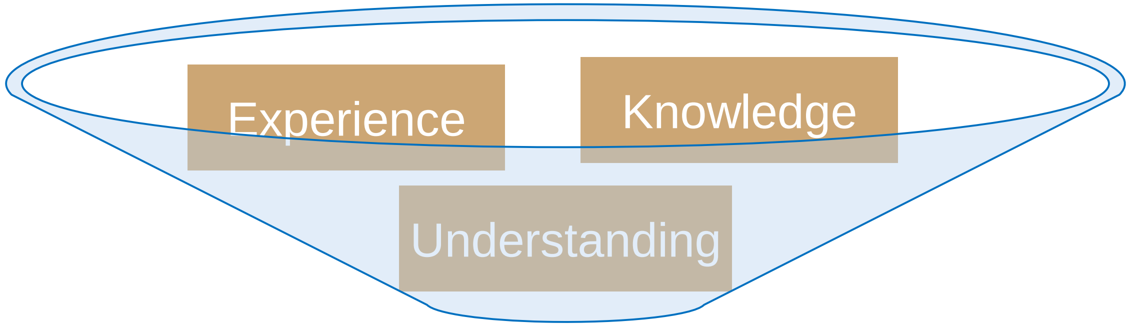
V1_Ch. 2 Annex 2A.1

2006 IPCC Guidelines

- **Motivation:** explaining the problem and purpose
- **Structuring:** defining the specifics of the protocol and problem
- **Conditioning:** previous experiences, models, theory and results
- **Encoding/Elicitation:** obtaining the data and information
- **Verification:** confirmation of outputs, bias testing and feeling of outliers

Proxy: Splicing techniques and Expert Judgement [13]

Elicitation protocols- EXPERT JUDGEMENT



Elicitation protocols

V1_Ch. 2
Annex 2A.1

Key concern with expert elicitation is to overcome the typical heuristic biases of *availability*, *representativeness*, and *anchoring and adjustment*

⊗ ***Overconfident estimate (narrow interval)***

- Fixed Value → Probability?
- Fixed Probability → Value?
- Interval Methods → Median and quartiles?
- Graphing → Drawing PDF

Proxy: Splicing techniques and Expert Judgement [14]

Elicitation protocols- EXPERT JUDGEMENT

Fixed Value: Estimate the probability of being higher (or lower) than an arbitrary value and repeat, typically three or five times.

For example, what is the probability that an emission factor would be less than 100?

Fixed Probability: Estimate the value associated with a specified probability of being higher (or lower). For example, what is the emission factor such that there is only a 2.5 percent probability (or 1 in 40 chance) that the emission factor could be lower (or higher) than that value

Interval Methods: It focuses on the median and the quartiles. For example, the expert would be asked to choose a value of the emission factor such that it is equally likely that the true emission factor would be higher or lower than that value. This yields the median.

Then the expert would divide the lower range into two bins such that he or she felt it to be equally likely (25 percent probability) that the emission factor could be in either bin, and repeat for the other end.

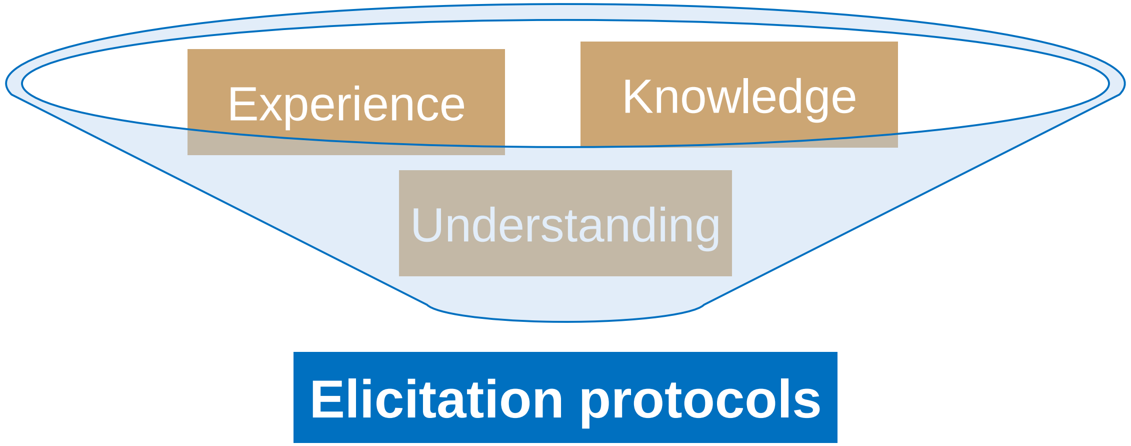
Finally, either fixed probability or fixed value methods could be used to get judgements for extreme values.

Graphing: The expert draws his/her own distributions.

This should be used cautiously since some experts are overconfident about their knowledge of PDFs

Proxy: Splicing techniques and Expert Judgement [15]

Elicitation protocols- EXPERT JUDGEMENT



Hands-on exercises

Proxy: Splicing techniques and Expert Judgement [16]

Elicitation protocols- EXPERT JUDGEMENT

- **Motivation:** to understand the methane content in landfill gas, identify an annual average and assess uncertainty
- **Structuring:** variability, climate & operation conditions, values and probabilities
- **Conditioning:** landfill characteristics, FOD model, cases
- **Elicitation:**

- **Verification:**

End of day 4

Thank you!

Diego M. Ezcurra

MONDAY

1. Overview of Uncertainty Analysis in National GHG Inventories

2. Causes of uncertainties associated with input data used in National GHG Inventories

3. How to reduce the uncertainty associated with input data used in National GHG Inventories

TUESDAY

4. Uncertainty associated with the use of national statistics, surveys/censuses and sampling

Practical examples in Energy, IPPU, AFOLU and Waste

WEDNESDAY

5. Uncertainty associated with the Use of Empirical Data

Practical examples in Energy, IPPU, AFOLU and Waste

THURSDAY

6. Uncertainty associated with the use of proxy, splicing techniques and expert judgment to fill data gaps

Practical examples in Energy, IPPU, AFOLU and Waste

FRIDAY

7. Selecting Probability Density Functions (PDF) and addressing correlation

Practical examples in Energy, IPPU, AFOLU and Waste

80%

Selecting PDF and addressing correlation [1]

GOOD PRACTICE GUIDANCE FOR SELECTING PROBABILITY DENSITY FUNCTIONS

V1_Ch3. 3.2.2.4

2006 IPCC Guidelines

Recommendations for different cases and commonly applied criteria to follow

Selecting PDF and addressing correlation [2]

GOOD PRACTICE GUIDANCE FOR SELECTING PROBABILITY DENSITY FUNCTIONS V1_Ch3. 3.2.2.4

Domain (+, -, ∞)

Range (narrow or broad)

Shape (symmetry)

Underlying process (+, x)

Others

Selecting PDF and addressing correlation [3]

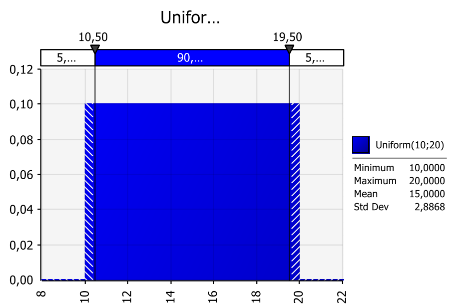
Selecting PDF

Most commonly used PDF

Selecting PDF and addressing correlation [4]

Selecting PDF

Uniform



All values with same probability

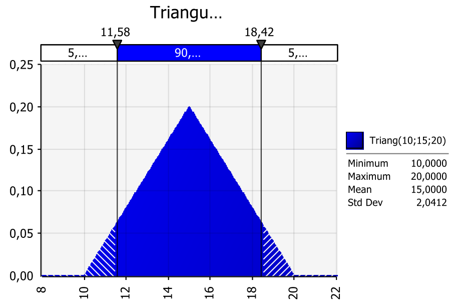
Parameters: Uniform **(min, max)**

Application to inventories: large uncertainty and lack of information

Selecting PDF and addressing correlation [5]

Selecting PDF

Triangular



Intuitive and flexible.

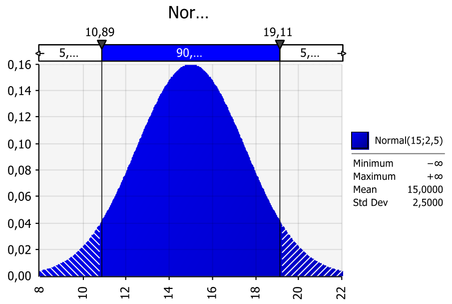
Parameters: Triang **(*min, mean, max*)**

Application to inventories: expert judgment, knowledge from experience. Limited information.

Selecting PDF and addressing correlation [6]

Selecting PDF

Normal



Distribution around a most likely central value.

Parameters: Normal (***mean [μ], std. Dev. [σ]***)

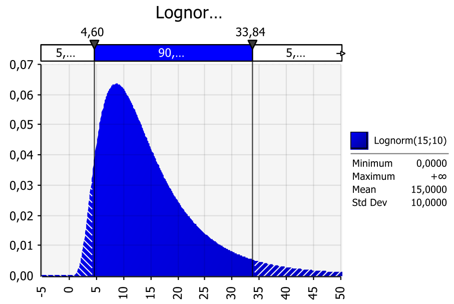
Application to inventories: most of situations (central limit). Additive processes.

The interval +/- 2*σ (1.96) accounts for approx. 95% of the values.
σ may be estimated as: (max - mean) / 2

Selecting PDF and addressing correlation [7]

Selecting PDF

Lognormal



The natural logarithm of the variable adjust to a Normal distribution

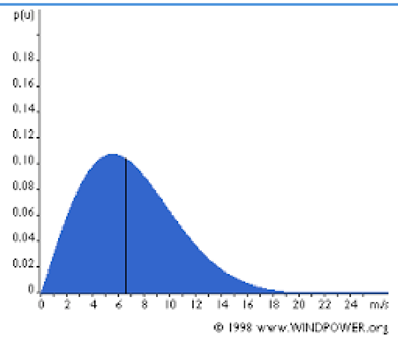
Parameters: Lognormal (***mean [μ], std. Dev. [σ]***)

Application to inventories: Generally good representation for skewed non negative values (emission factors for N_2O). Multiplicative processes.

Selecting PDF and addressing correlation [8]

Selecting PDF

Weibull



Distribution characterized by a shape parameter and a scale parameter. Shape can change drastically.

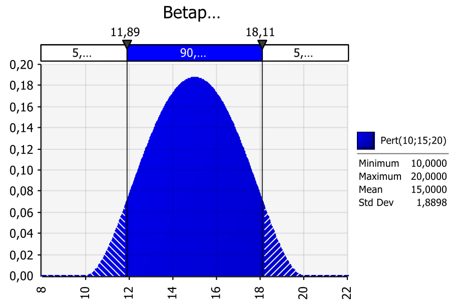
Parameters: Weibull (***λ [scale], k [shape],***)

Application to inventories: positively skewed non-negative values (similar to lognormal).

Selecting PDF and addressing correlation [9]

Selecting PDF

Beta pert



Version of the Beta using PERT (Program Evaluation and Review Technique).

Parameters: BetaPert **(*min, mean, max*)**

Application to inventories: similar to triangular but with lower standard deviation.

Selecting PDF and addressing correlation [10]

Selecting PDF - Good practice guidance

- The minimum number of probability functions are used
- These probability functions are well known and well based (theoretical or empirical)
- Where empirical data are available, the first choice should be to assume a normal distribution
- If the variable must be non-negative and a normal distribution is assumed, the standard deviation should not exceed 30%
- Truncation of the negative tail of the normal distribution should be avoided (use instead lognormal, Weibull, or Gamma).

Selecting PDF and addressing correlation [11]

Selecting PDF - Good practice guidance

- Where expert judgment is used, the distribution function adopted might be normal or lognormal, supplemented by uniform or triangular distributions
- If only the interval is known (upper and a lower value), assume that the probability density function is uniform and that the range corresponds to the 95% confidence interval
- If the distribution observed based on data does not seem correct, the data may be the problem (not representative, not random, small sample size, different timing, etc.)

Selecting PDF and addressing correlation [12]

Selecting PDF - Good practice guidance

- When selecting the PDF from Goodness-of-Fit test, several functions will fit the data satisfactorily within a given probability limit
- Different functions can have radically different distributions at the extremes (few or no data to constrain them), and the choice of one function over another can systematically change the outcome of an uncertainty analysis.

"it must be knowledge of the underlying physical processes that governs the choice of a probability function"

Selecting PDF and addressing correlation [13]

DEPENDENCE AND CORRELATION AMONG INPUTS

Dependence / Correlation

Relationship between two variables that make them not independent

One variable is determined (partially or totally) by another one

Selecting PDF and addressing correlation [14]

DEPENDENCE AND CORRELATION AMONG INPUTS

Dependence / Correlation

Negative

Reduces uncertainty range when included

If not included: overestimation of uncertainty

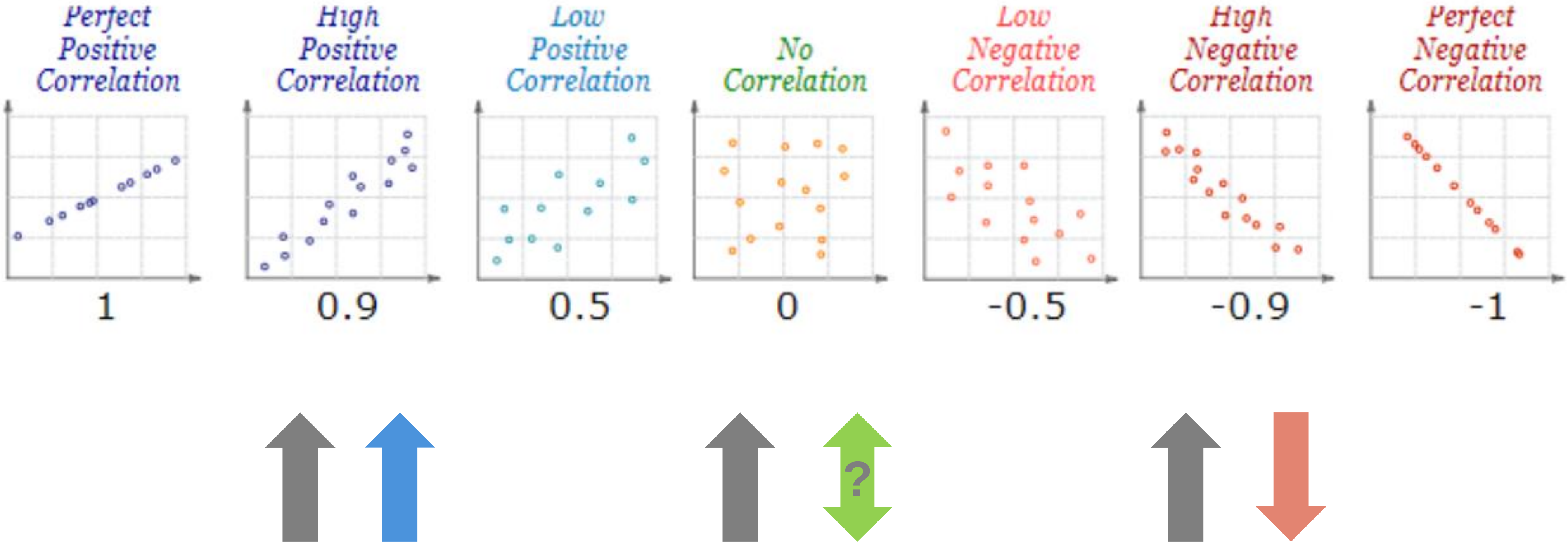
Positive

Increases uncertainty range when included

If not included: underestimation of uncertainty

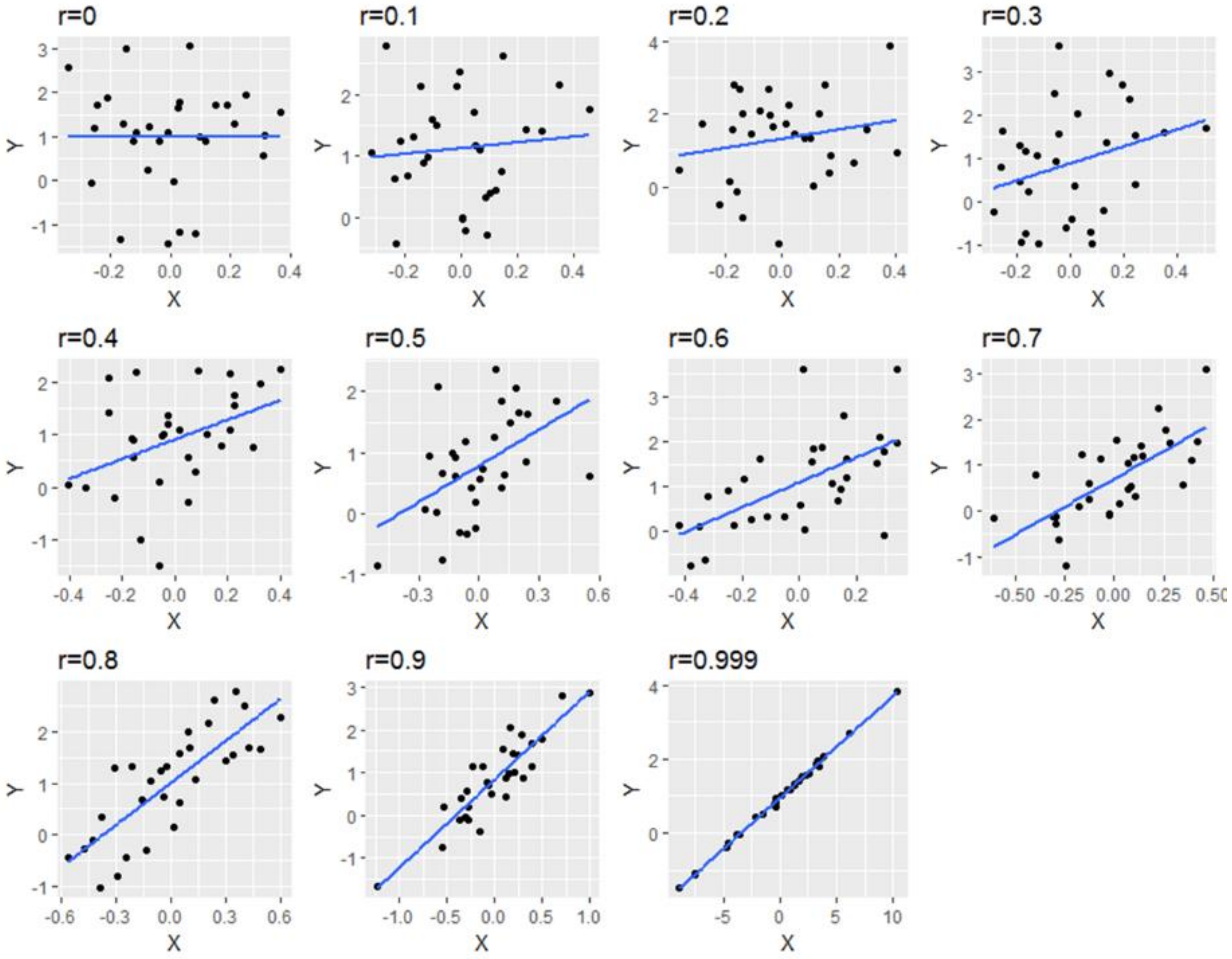
Selecting PDF and addressing correlation [15]

DEPENDENCE AND CORRELATION AMONG INPUTS



Selecting PDF and addressing correlation [16]

DEPENDENCE AND CORRELATION AMONG INPUTS



Correlation degree

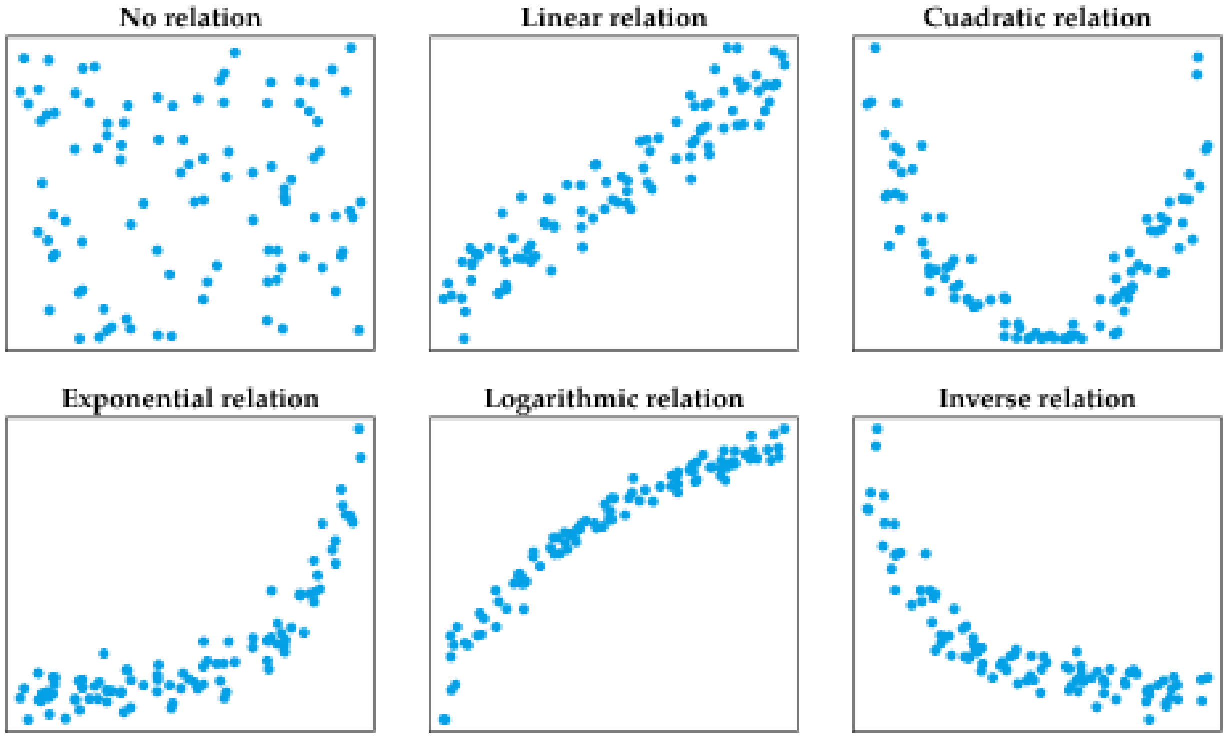
< 0.2	**Very low**
0.2 – 0.4	**Low**
0.4 – 0.6	**Moderate**
0.6 – 0.8	**High**
> 0.8	**Very high**

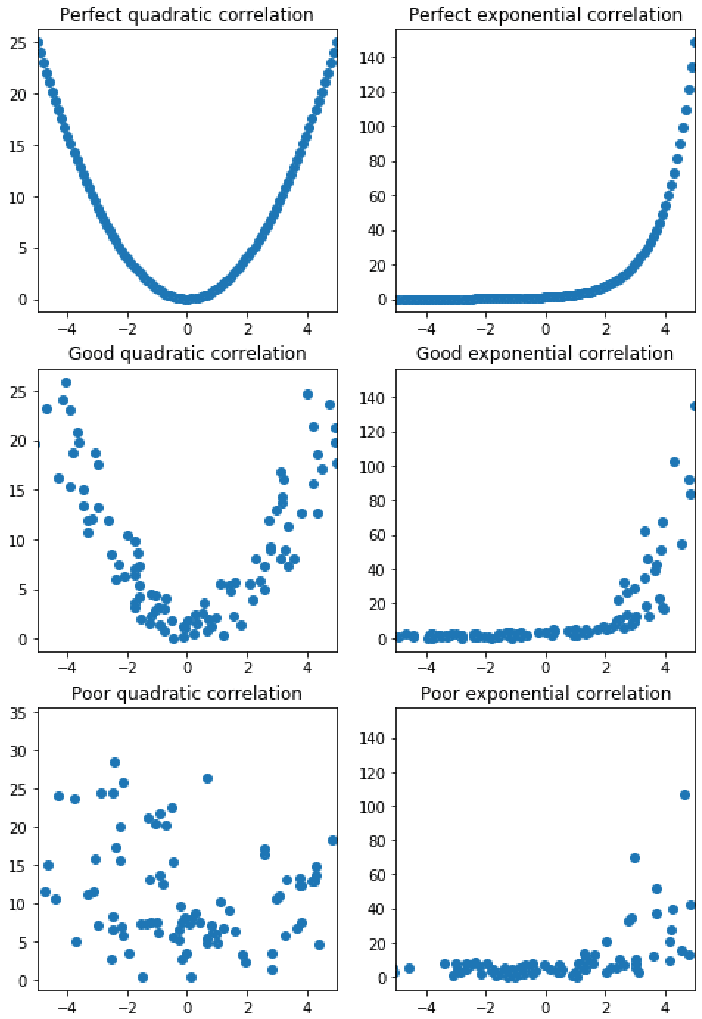
* Indicative ranges

Selecting PDF and addressing correlation [17]

DEPENDENCE AND CORRELATION AMONG INPUTS

Attention! Non-linear correlation also exists





Selecting PDF and addressing correlation [18]

DEPENDENCE AND CORRELATION AMONG INPUTS

Are FC (GJ) and EF independent?

$$E = \overbrace{FC \times NCV}^{FC\ (GJ)} \times EF$$

FC: Fuel consumption (tonnes)
NCV: Net calorific value (GJ/t)
EF: Emission factor (tCO_2/GJ)

$$\boldsymbol{EF = \frac{\%C}{NCV} x \frac{44}{12}}$$

Are NCV and EF independent?

Selecting PDF and addressing correlation [19]

DEPENDENCE AND CORRELATION AMONG INPUTS

Dependencies / Correlations

Are always important to uncertainty assessment?

Degree: strong or weak correlation (i.e.   0.8 or 0.2)

Sensitivity: impact to the overall uncertainty

Selecting PDF and addressing correlation [20]

DEPENDENCE AND CORRELATION AMONG INPUTS

Exists between 2 variables to which uncertainty is NOT sensitive to and dependency is strong	Exists between 2 variables to which uncertainty is sensitive to and dependency is strong
Exists between 2 variables to which uncertainty is NOT sensitive to and dependency is weak	Exists between 2 variables to which uncertainty is sensitive to and dependency is weak

Selecting PDF and addressing correlation [21]

DEPENDENCE AND CORRELATION AMONG INPUTS

Dependence / Correlation

Strategies

- Define the model so that the inputs are as statistically independent as possible
- Stratify or aggregate the category to minimise the dependency effect
- Model dependency explicitly
- Use sensitivity cases (independent, fully positive and fully negative correlated)

Selecting PDF and addressing correlation [22]

DEPENDENCE AND CORRELATION AMONG INPUTS

Dependence / Correlation

Hands-on exercises

Selecting PDF and addressing correlation [23]

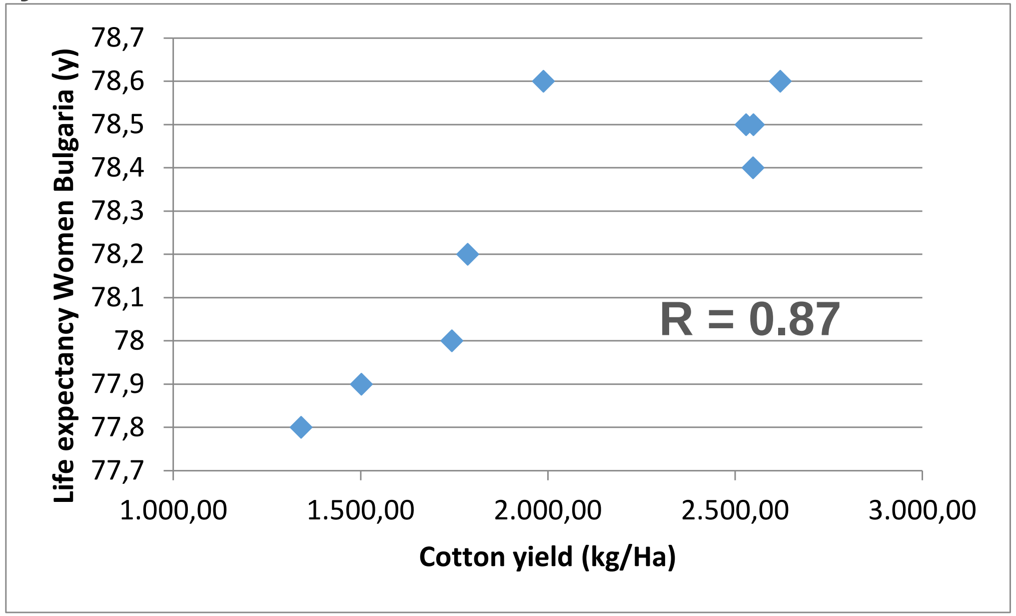
DEPENDENCE AND CORRELATION AMONG INPUTS

Dependence / Correlation

Attention! Correlation does not imply dependency

AFOLU: assess the relationship between different corps cultivated area in order to understand if correlation needs to be included to calculate the uncertainty of the emissions.

	Soybean 1°	*Soybean 2°*	*Soybean*	*Maize*	*Wheat*	*Sunflower*
Soybean 1°	1					
Soybean 2°	-0,10	1				
Soybean	0,96	0,18	1			
Maize	0,26	0,85	0,49	1		
Wheat	-0,86	0,47	-0,72	0,07	1	
Sunflower	-0,56	-0,15	-0,60	-0,49	0,47	1



Selecting PDF and addressing correlation [24]

DEPENDENCE AND CORRELATION AMONG INPUTS

Dependence / Correlation

Energy: Relationship between fuel consumption in different sectors

Total gasoil consumption is obtained at national level from fuel suppliers. The sectoral distribution is estimated for transport, industrial and commercial sectors through sampling. The residential fuel use is obtained as the difference between total fuel consumption and usage in the other sectors.

Assess uncertainty in residential sector without and with correlation

U: 6 % **U: 40%-70%**

MONDAY

1. Overview of Uncertainty Analysis in National GHG Inventories
2. Causes of uncertainties associated with input data used in National GHG Inventories
3. How to reduce the uncertainty associated with input data used in National GHG Inventories

TUESDAY

4. Uncertainty associated with the use of national statistics, surveys/censuses and sampling

Practical examples in Energy, IPPU, AFOLU and Waste

WEDNESDAY

5. Uncertainty associated with the Use of Empirical Data

Practical examples in Energy, IPPU, AFOLU and Waste

THURSDAY

6. Uncertainty associated with the use of proxy, splicing techniques and expert judgment to fill data gaps

Practical examples in Energy, IPPU, AFOLU and Waste

FRIDAY

7. Selecting Probability Density Functions (PDF) and addressing correlation

Practical examples in Energy, IPPU, AFOLU and Waste

100%

End of webinar!

Thank you

Diego M. Ezcurra