UNCERTAINTIES

Ass. Prof. Oleksandr Mokliachuk
NATIONAL POLITECHNICAL
UNSVERSITY KPI

Prof. Lidiya Moklyachuk
Head of the Ecotoxicology Department
INSTITUTE OF AGROECOLOGY
AND ENVIRONMENTAL
MANAGEMENT of NAAS

Joint Committee for Guides in Metrology (JCGM)

Overview of uncertainty analysis

Uncertainty estimates includes methods for:

• Determining uncertainties in individual variables used in the inventory;

• Aggregating the component uncertainties to the total inventory;

• Determining the uncertainty in the trend;

• Identifying significant sources of uncertainty in the inventory to help prioritise data collection and efforts to improve the inventory.

Quantitative uncertainty analysis is performed by estimating the 95 percent confidence interval of the data for individual categories and for the total inventory.

Key concepts and terminology

***Accuracy*:** Agreement between the true value and the average of repeated measured observations or estimates of a variable. An accurate measurement does not have systematic error.

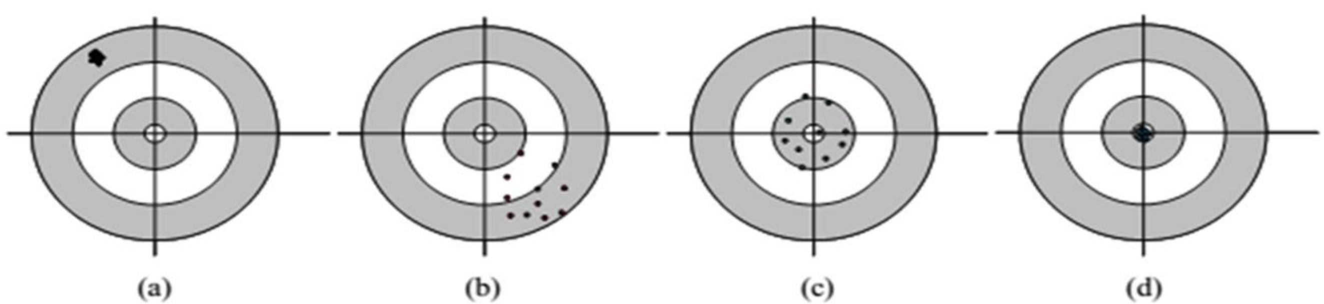
***Bias*:** Lack of accuracy. Bias (systematic error), can occur because of instrument error.

***Confidence Interval*:** The confidence interval is a range that encloses the true value of this unknown quantity, with a specified confidence (probability). Typically, a 95 percent confidence interval is used.

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

Figure 3.2 Illustration of accuracy and precision

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



***Probability density function (PDF)*:** The PDF describes the range of possible values of unknown quantity, such as the annual total ammonia emissions.

***Random errors*:** Random error is a distinct concept compared to systematic error.

***Systematic error*:** Another term for *bias*.

***Uncertainty*:** Lack of knowledge of the true value of a variable

***Variability*:** Changing of a variable over time, space or members of a population.

Basis for uncertainty analysis

Two main statistical concepts:

- the probability density function (PDF) and confidence interval are used in uncertainty analysis.
- in contrast, systematic errors can be much more difficult to quantify.

Causes of uncertainty

We describe eight broad causes of uncertainty:

• *Lack of completeness*: This is a case where measurement or other data are not available.

• *Model*: The use of models to estimate data can introduce uncertainty, including both bias and random error, for a variety of reasons:

(i) Models are not exact

(ii) Interpolation is application of a model within a range of inputs.

(iii) **Extrapolation can lead to uncertainty;**

(iv) Alternative formulations of the model may result in different estimates;

(v) Model inputs are generally approximated based on limited information that create additional uncertainties.

• *Lack of data*: In some situations, there simply may not yet be data available

• *Lack of representativeness of data*: This uncertainty is associated with lack of complete correspondence between conditions associated with the available data and the conditions associated with real world emissions/removals or activity.

***Statistical random sampling error*:** This uncertainty is associated with data that are a random sample of finite size.

- ***Measurement error*:** Measurement error, which may be random or systematic
- ***Misreporting or misclassification*:** Uncertainty here may be due to incomplete, unclear, or faulty definition.
- ***Missing data*:** Uncertainties may result where no value was available. This cause of uncertainty can lead to both bias and random error.

Reducing uncertainty

Depending on the cause of uncertainty present, uncertainties could be reduced in seven ways:

- ***Improving conceptualisation*:** Improving the structural assumptions.
- ***Improving models***: Improving the model structure and parameterisation
- ***Improving representativeness*:** This is particularly important for categories in the agriculture, forestry and land use parts of an inventory.
- ***Using more precise measurement methods*:** Measurement error can be reduced by using more precise measurement methods.
- ***Collecting more measured data*:** Uncertainty associated with random sampling error can be reduced by increasing the sample size.
- ***Eliminating known risk of bias*:** This is achieved by ensuring instrumentation is properly positioned and calibrated.
- ***Improving state of knowledge***: Improving the understanding of the categories and the processes .

Reducing uncertainty

Depending on the cause of uncertainty present, uncertainties could be reduced in seven ways:

- ***Improving conceptualisation*: Improving the structural assumptions.**
- ***Improving models*:** Improving the model structure and parameterisation
- ***Improving representativeness***: This is particularly important for categories in the agriculture, forestry and land use parts of an inventory.
- ***Using more precise measurement methods*:** Measurement error can be reduced by using more precise measurement methods.
- ***Collecting more measured data*:** Uncertainty associated with random sampling error can be reduced by increasing the sample size.
- ***Eliminating known risk of bias***: This is achieved by ensuring instrumentation is properly positioned and calibrated.
- ***Improving state of knowledge*:** Improving the understanding of the categories and the processes leading to emissions and removals

QUANTIFYING UNCERTAINTIES

Sources of data and information

There are three broad sources of data and information:
information contained in models;
empirical data associated with measurements.
quantified estimates of uncertainties based upon expert judgement.

Uncertainties associated with models

There are two key considerations in model uncertainty:
(1) has the correct, most relevant real-world system been identified,
(2) is the model an accurate representation of the chosen system.
l development relative to the intended system and conceptualisation(s).

UNCERTAINTY OF EMPIRICAL DATA DEPENDENT:

(a) representativeness of the data and potential for bias;
(b) precision and accuracy of the measurements;
(c) sample size and inter-individual variability in measurements;
(d) inter-annual variability data and whether estimates are based upon an average of several years or on the basis of a particular year.

Representative sampling (or sampling design) implies that measurements are made for typical system characteristics. The precision and accuracy of individual measurements will depend upon the equipment and protocols used to make the measurements.

If the data sample size is large enough, standard statistical goodness-of-fit tests can be used, in combination with expert judgement, to help in deciding which PDF to use for describing variability in the data (partitioned if necessary) and how to parameterise it.

UNCERTAINTY ESTIMATES FOR EMISSION FACTORS AND OTHER PARAMETERS OBTAINED FROM PUBLISHED REFERENCES

When site-specific data are unavailable, the associated uncertainties should be estimated from:

- *Original research including country-specific data*:
- *Default Values from Guidelines*:.

UNCERTAINTIES ASSOCIATED WITH ACTIVITY DATA

Activity data are often more closely linked to economic activity.

There are several approaches to identifying the potential uncertainty of activity data for complete samples:

• To check for the size of random errors, look for fluctuations over time, and differential fluctuations in series.

• To check for bias errors, cross-check the data of interest with other, related information.

***Activity data based on random samples*:** The data will be subject to sampling errors, which will be normally distributed and uncorrelated over time.

EXPERT JUDGEMENT AS A SOURCE OF INFORMATION

In many situations, directly relevant empirical data are not available for sources, sinks, or activity inputs to an inventory. In such situations, a practical solution to dealing with the absence of adequate data is to obtain well informed judgements from domain experts regarding best estimates and uncertainties of inputs to the inventory.

UNCERTAINTY IN MODELS

There are at least three major approaches for estimating uncertainty:
(1) comparison of model results with independent data for purposes of verification;
(2) comparison of the predictions of alternative models;
(3) expert judgement regarding the magnitude of model uncertainty.
These approaches can be used in combination.

The remaining uncertainty can then be quantitatively assessed by expert judgement about how uncertainties in the data used to drive the model and the model parameters combine, or more formally by Monte Carlo analysis.

STATISTICAL ANALYSIS OF EMPIRICAL DATA

Statistical analysis of empirical data can be summarised as the following major steps:
Step 1: Compilation and evaluation of a database,
Step 2: Visualisation of data by developing empirical distribution functions
Step 3: Fitting, evaluation, and selection of alternative PDF models for representing variability in activity data.
Step 4: Characterisation of uncertainty in the mean of the distributions for variability.
Step 5: Once uncertainties have been appropriately specified, these can be used as input to a probabilistic analysis for purposes of estimating uncertainty in total process.
Step 6: Sensitivity analysis is recommended to determine which of the input uncertainties to an inventory contributes most substantially to the overall uncertainty.

METHODS FOR ENCODING EXPERT JUDGEMENTS

When empirical data are lacking or are not considered fully representative for all causes of uncertainty, expert judgement may be necessary for estimating uncertainty.

The goal of the process of obtaining expert judgement is to develop a Probability density function (PDF) taking into account relevant information.

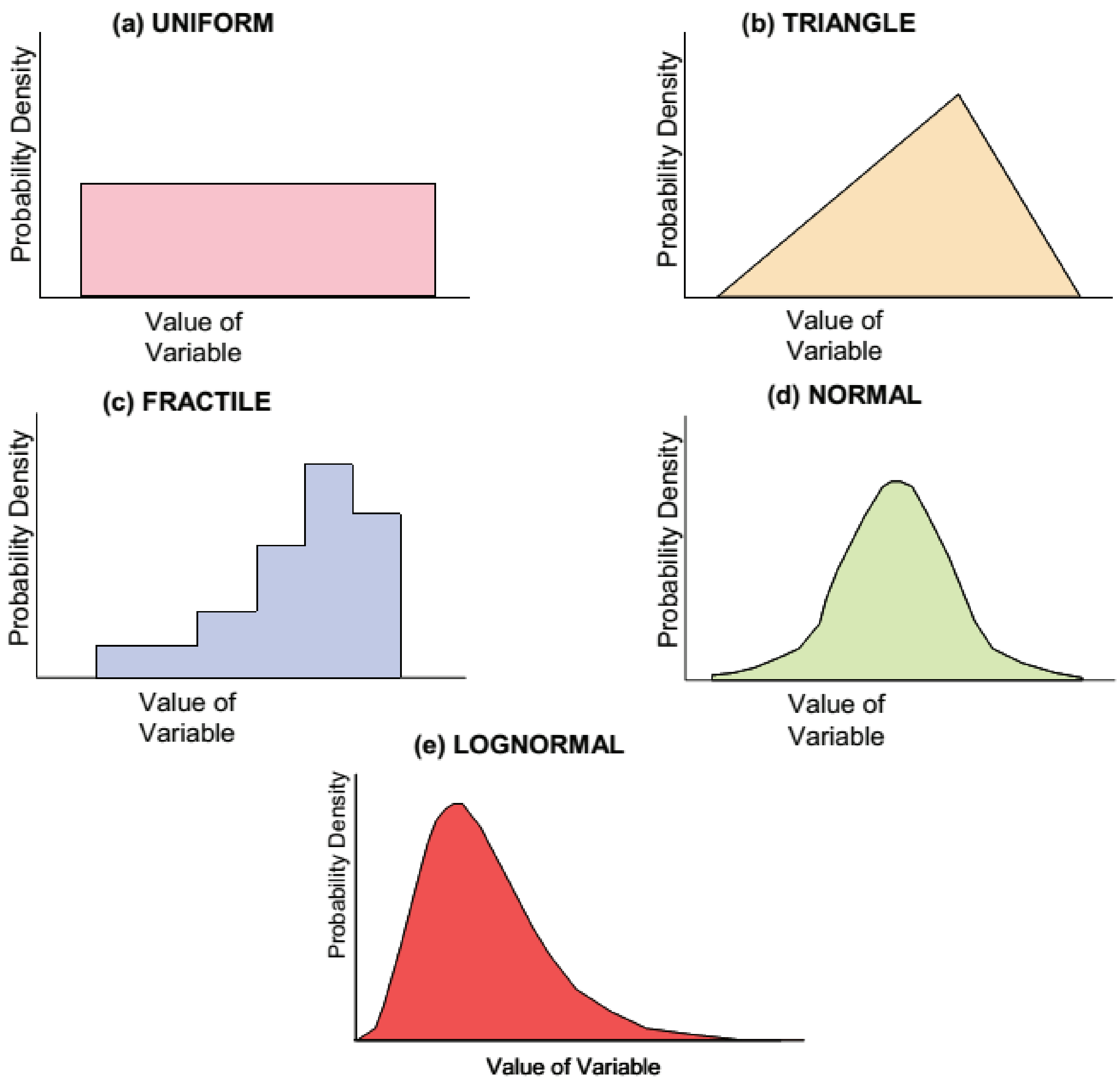
TYPES OF PROBABILITY DENSITY FUNCTIONS

The choice of a particular type of PDF depends, at least in part, on the domain of the function, the range of the function, the shape, and processes that generated the data.

Examples of such functions and the situations they represent are:

- *The normal distribution* is most appropriate when the range of uncertainty is small, and symmetric relative to the mean. The normal distribution arises in situations where many individual inputs contribute to an overall uncertainty, and in which none of the individual uncertainties dominates the total uncertainty.
- *The lognormal distribution* may be appropriate when uncertainties are large for a non-negative variable and known to be positively skewed. The emission factor for nitrous oxide from fertiliser applied to soil provides a typical inventory example.
- *Uniform distribution* describes an equal likelihood of obtaining any value within a range. The uniform distribution is a special case of the Beta distribution.
- *The triangular distribution* is appropriate where upper and lower limits and a preferred value are provided by experts but there is no other information about the PDF.
- *Fractile distribution* is a type of empirical distribution in which judgements are made regarding the relative likelihood of different ranges of values for a variable, such as illustrated in Figure 3.5.

Figure 3.5 **Examples of some commonly used probability density function models (e.g., based on Frey and Rubin, 1991)**



Methods to combine uncertainties

Two approaches for the estimation of combined:

Approach 1 uses simple error propagation equations.

Approach 2 uses Monte Carlo or similar techniques.

APPROACH 1: PROPAGATION OF ERROR

Approach 1 is based upon error propagation and is used to estimate uncertainty in individual categories, in the buget as a whole, and in trends between a year of interest and a base year.

KEY ASSUMPTIONS OF APPROACH 1

In Approach 1 uncertainty can be propagated from uncertainties in the activity data, emission factor and other estimation parameters through the error propagation equation.

KEY REQUIREMENTS OF APPROACH 1

In order to quantify uncertainty using Approach 1, estimates of the mean and the standard deviation for each input are required, as well as the equation through which all inputs are combined to estimate an output. The simplest equations include statistically independent (uncorrelated) inputs.

PROCEDURE OF APPROACH 1

The Approach 1 analysis estimates uncertainties by using the error propagation equation in two steps.

First, the Equation 3.1 approximation is used to combine emission factor, activity data and other estimation parameter.

EQUATION 3.1

COMBINING UNCERTAINTIES – APPROACH 1 – MULTIPLICATION

$$U_{total} = \sqrt{U_1^2 + U_2^2 + \ldots + U_n^2}$$

Where:

U_{total} = the percentage uncertainty in the product of the quantities (half the 95 percent confidence interval divided by the total and expressed as a percentage);

U_i = the percentage uncertainties associated with each of the quantities.

- Where uncertain quantities are to be combined by addition or subtraction, the standard deviation of the sum will be the square root of the sum of the squares of the standard deviations of the quantities that are added with the standard deviations all expressed in absolute terms (this rule is exact for uncorrelated variables).

  Using this interpretation, a simple equation (Equation 3.2) can be derived for the uncertainty of the sum, expressed in percentage terms:

Equation 3.2 can be used to derive the uncertainty of the sum, expressed in percentage terms:

EQUATION 3.2

COMBINING UNCERTAINTIES – APPROACH 1 – ADDITION AND SUBTRACTION

$$U_{total} = \frac{\sqrt{(U_1 \bullet x_1)^2 + (U_2 \bullet x_2)^2 + \ldots + (U_n \bullet x_n)^2}}{|x_1 + x_2 + \ldots + x_n|}$$

Where:

U_{total} = the percentage uncertainty in the sum of the quantities (half the 95 percent confidence interval divided by the total (i.e., mean) and expressed as a percentage). This term 'uncertainty' is thus based upon the 95 percent confidence interval;

x_i and U_i = the uncertain quantities and the percentage uncertainties associated with them, respectively.

APPROACH 2: MONTE CARLO SIMULATION

In Monte Carlo simulation (моделирование), pseudo-random samples of model inputs are generated according to the **Probability density function** PDFs specified for each input. Because Monte Carlo simulation is a numerical method, the precision of the results typically improves as the number of iterations is increased.

KEY ASSUMPTIONS OF APPROACH 2

Numerical (численные) statistical techniques, particularly the Monte Carlo technique, are more appropriate than Approach 1 for estimating uncertainty when:

- uncertainties are large;
- their distribution are non-Gaussian;
- algorithms are complex functions;
- correlations occur between some of the activity data sets, emission factors, or both;
- uncertainties are different for different years of the inventory.

KEY REQUIREMENTS OF APPROACH 2

Monte Carlo simulation requires the analyst to specify PDFs that reasonably represent each model input for which the uncertainty is quantified. A key consideration is to develop the distributions for the input variables to the emission/removal calculation model so that they are based upon consistent underlying assumptions regarding averaging time, location, and other conditioning factors relevant to the particular assessment.

PROCEDURES OF APPROACH 2

The principle of Monte Carlo analysis is to select random values of emission factor, activity data and other estimation parameters from within their individual probability density functions, and to calculate the corresponding emission values. This procedure is repeated many times, using a computer, and the results of each calculation run build up the overall emission probability density function. The results will only be valid to the extent that the input data, including any expert judgements, are sound.

Thank you very much for your attention!

ОСНОВНЫЕ ТРЕБОВАНИЯ ПОДХОДА 2

Монте-Карло требует от аналитика указать PDF-файлы, которые разумно представляют каждая модель вход для которых неопределенность количественно.Ключевым фактором является разработка распределения для входных переменных в модели расчета выбросов / абсорбции, так что они основаны на согласованных базовых предположениях о времени усреднения, местоположения и других факторов кондиционирования, относящихся к конкретной оценке.

ПРОЦЕДУРЫ ПОДХОДА 2

Принцип анализа методом Монте-Карло является выбор случайных значений коэффициента выбросов, данных о деятельности и других параметров оценки в пределах их от отдельных функций плотности вероятности, и вычислить соответствующие значения выбросов. Эта процедура повторяется много раз, используя компьютер, а результаты каждого расчета перспективе создать общую функцию плотности вероятности выбросов. Результаты будут действительны только в той степени, что входные данные, включая любые экспертные оценки звука.