ANNEX 3

GLOSSARY

CO-CHAIRS, EDITORS AND EXPERTS

Co-chairs of the Expert Meeting on Cross-sectoral Methodologies for Uncertainty Estimation and Inventory Quality

Taka Hiraishi (Japan) and Buruhani Nyenzi (Tanzania)

REVIEW EDITOR

Richard Odingo (Kenya)

AUTHORS

Milos Tichy (Czech Republic) and Simon Bentley (Australia)

REVIEWERS

Roberto Acosta (UNFCCC Secretariat), Simon Eggleston (UK), Ian Galbally (Australia), Katarina Mareckova (Slovak Republic), Thomas Martinsen (IPCC/OECD), Jos Olivier (Netherlands), Jim Penman (UK), and Kristin Rypdal (Norway)

Contents

ANNEX 3 GLOSSARY

INDEX TO THE GLOSSARY A3.4
A3.1 INTRODUCTION A3.6
A3.1.1 Selection of entries A3.6
A3.1.2 Formulation of entries A3.6
A3.2 GLOSSARY A3.7
REFERENCES A3.21

Index

A

Accuracy A3.7

Activity data A3.7

Arithmetic mean A3.7

Auto-correlation A3.7

Auto-covariance A3.8

B

Bias A3.8

Bootstrap technique A3.8

C

Central limit theorem A3.8

Coefficient of variation A3.8

Comparability A3.8

Completeness A3.8

Confidence A3.9

Confidence interval A3.9

Consistency A3.9

Correlation A3.9

Correlation coefficient A3.9

Covariance A3.10

Cumulative distribution function *See* Distribution function

D

Decision tree A3.10

Distribution function A3.10

E

Elasticity A3.10

Emission Factor A3.10

Error A3.10

Estimation A3.10

Estimator A3.11

Expectation A3.11

Expected value A3.11

Expert Judgement A3.11

Extreme value A3.11

G

Good Practice A3.11

I

Independence A3.12

K

Key source category A3.12

Kurtosis A3.12

L

Latin hypercube sampling A3.12

Law of large numbers A3.12

Linear model A3.12

Linear regression A3.12

Lognormal distribution A3.13

M

Mean A3.13

Median A3.13

Mode A3.13

Model A3.14

Moments (of random variable) A3.14

Monte Carlo method A3.14

N

Non-linear model A3.14

Normal distribution A3.14

P

Parameters of population A3.14

PDF *See* Probability density function

Percentile A3.15

Population A3.15

Population percentile *See* Percentile

Population standard deviation *See* Standard deviation

Precision A3.15

Probability A3.15

Probability density function – PDF A3.15

Probability distribution A3.16

Propagation of uncertainties A3.16

Q

Quality Assurance (QA) A3.16

Quality Control (QC) A3.16

R

Random error A3.16

Random variable A3.16

Residual A3.17

S

Sample A3.17

Sample mode *See* Mode

Sample percentile *See* Percentile

Sensitivity A3.17

Sensitivity analysis A3.17

Sigma interval A3.17

Simple random sample A3.17

Skewness A3.17

Standard deviation A3.18

Standard error of the mean A3.18

Statistic A3.18

Statistics A3.18

Systematic and random errors A3.18

Systematic error A3.18

T

Time series A3.18

Transparency A3.18

Trend A3.18

Triangular distribution A3.19

U

Unbiased estimator A3.19

Uncertainty A3.19

Uncertainty analysis A3.19

Uniform distribution A3.19

V

Validation A3.20

Variability A3.20

Variance A3.20

Variance of sample mean A3.20

Verification A3.20

ANNEX 3 GLOSSARY

A3.1 INTRODUCTION

The Glossary provides a convenient reference for inventory compilers and policy makers covering general statistical terms and terms that have a particular meaning in the context of emission inventories.

A3.1.1 Selection of entries

The main purposes for the selection of terms and the formulation of entries were to:

- Distinguish between terms that have different meanings when used in the context of greenhouse gas inventory compilation and when used in a technical, statistical, or mathematical sense – e.g. the term 'consistency';
- Provide unified notation for the basic terms (mostly statistical) considered fundamental to the practical reporting of inventories;
- Define other terms that assist in the understanding and development of good practice guidelines for uncertainties in national inventories.

A3.1.2 Formulation of entries

The Glossary takes a pragmatic approach and provides one or more of the following types of definitions for each entry. First, any definition developed specially for inventory applications is marked 'Inventory definition.' In some cases, examples are used to illustrate specific meanings for inventory preparation. The second type is 'statistical definition,' that are used to explain the statistical or common mathematical definition for a certain term. Again, in some cases, examples are provided to clarify the application of these meanings for inventory use. The final type of definitions are those that come from other sources – including pre-existing SBSTA or IPCC definitions agreed by the Subsidiary Body for Scientific and Technological Advice (SBSTA) of the United Nations Framework Convention on Climate Change (UNFCCC) (indicated by FCCC/SBSTA/1999/6 Add. 1), the *Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories*, and the International Standardization Organization (ISO). The clauses, that fall under the following definitions: arithmetic mean, expectation, population, probability, probability distribution, random variable, statistic and uncertainty include a definition indicated by [7] are taken from the publication *Guide to the Expression of Uncertainty in Measurement* and have been reproduced with permission from the International Organization for Standardization (ISO). This ISO publication can be obtained from any member body or directly from the Central Secretariat, ISO, Case Postale 56, 1211 Geneva 20, Switzerland. Copyright remains with the ISO.

The definitions provided in this Glossary are not rigorous in a full mathematical or statistical sense. Most of the statistical definitions given here lie within the context of 'classical' frequency-based statistical inference, although it is acknowledged that this is not the only theory of statistical inference. As with any reference manual, compromises have been made between understandability, clarity, exactness, and brevity. To this end, mathematical notation has been kept to a minimum.

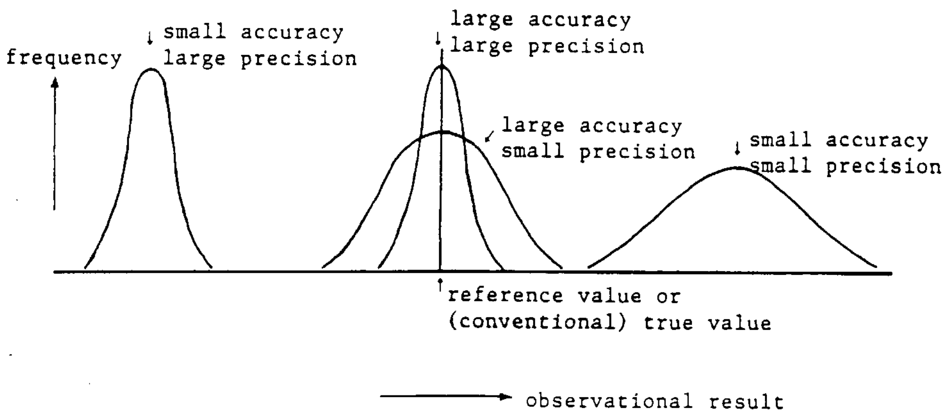
A3.2 GLOSSARY

ACCURACY

Inventory definition: Accuracy is a relative measure of the exactness of an emission or removal estimate. Estimates should be accurate in the sense that they are systematically neither over nor under true emissions or removals, as far as can be judged, and that uncertainties are reduced as far as practicable. Appropriate methodologies conforming to guidance on good practices should be used to promote accuracy in inventories. (FCCC/SBSTA/1999/6 Add. 1)

Statistical definition: Accuracy is a general term which describes the degree to which an estimate of a quantity is unaffected by bias due to systematic error. It should be distinguished from precision as illustrated on Figure A3.1.

Figure A3.1 Accuracy and Precision (from [3]*)



ACTIVITY DATA

Inventory definition: Data on the magnitude of human activity resulting in emissions or removals taking place during a given period of time. In the energy sector for example, the total amounts of fuel burned is annual activity data for fuel combustion sources, and the total number of animals being raised, by species, is annual activity data for methane emissions from enteric fermentation. (*Revised 1996 IPCC Guidelines* [9]*)

ARITHMETIC MEAN

Statistical definition: The sum of the values divided by the number of values. [7]*

AUTO-CORRELATION

Statistical definition: The correlation coefficient calculated for two data items in a time series.

Example: Observed animal numbers in two successive years are usually highly auto-correlated when the lifetime of the animals significantly exceeds two years.

---

* See references (p A3.21).

AUTO-COVARIANCE

Statistical definition: The covariance calculated for two data items in a time series.

BIAS

Inventory definition: A systematic error of the observation method, whose value in most cases is unknown. It can be introduced by using measuring equipment that is improperly calibrated, by selecting items from a wrong population or by favouring certain elements of a population, etc.

Statistical definition: The difference between the expected value of a statistic and the parameter which it estimates. See **Unbiased estimator**.

Example: Estimating the total fugitive emission from gas transport and distribution using only measurements of leakage from high/medium pressure pipelines can lead to bias if the leakage in the lower pressure distribution network (which is significantly more difficult to measure) is neglected.

BOOTSTRAP TECHNIQUE

Statistical definition: Bootstrap technique is a type of computationally intensive statistical methods which typically use repeated resampling from a set of data to assess variability of parameter estimates.

CENTRAL LIMIT THEOREM

Statistical definition: A general name for the class of mathematical/statistical theorems which, very broadly stated, says that the arithmetic mean of *n* independently distributed and random variables approximates a normal distribution as *n* tends to infinity. This is true for underlying distributions of variables likely to be encountered in practice, and certainly for any distributions likely to be encountered in the context of greenhouse gas inventories. For inventories, the theorem gives a guide to the interpretation of combined variances of total emission (which is the sum of sectoral emissions). Also, under some conditions, the central limit theorem can justify the approximation that the total emissions from a bottom-up inventory has a normal distribution.

COEFFICIENT OF VARIATION

Statistical definition: The coefficient of variation, v_x is the ratio of the population standard deviation, σ_x , and mean, μ_x, where $v_x = \sigma_x/\mu_x$. It also frequently refers to the sample coefficient of variation, which is the ratio of the sample standard deviation and sample mean.[1]

COMPARABILITY

Inventory definition: Comparability means that estimates of emissions and removals reported by Parties in inventories should be comparable among Parties. For this purpose, Parties should use the methodologies and formats agreed by the Conference of the Parties (COP) for estimating and reporting inventories. The allocation of different source/sink categories should follow the split of the *Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories*, at the level of its summary and sectoral tables.

COMPLETENESS

Inventory definition: Completeness means that an inventory covers all sources and sinks as well as all gases included in the *Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories* in addition to other existing relevant source/sink categories which are specific to individual Parties (and therefore may

[1] ‘Coefficient of variation’ is the term, which is frequently replaced by ‘error’ in a statement like ‘error is 5%’.

not be included in the *IPCC Guidelines*). Completeness also means full geographic coverage of sources and sinks of a Party[2].

CONFIDENCE

Inventory definition: The term 'confidence' is used to represent trust in a measurement or estimate. Having confidence in inventory estimates does not make those estimates more accurate or precise; however, it will eventually help to establish a consensus regarding whether the data can be applied to solve a problem [6].* This usage of confidence differs substantially from the statistical usage in the term confidence interval.

CONFIDENCE INTERVAL

Statistical definition: A confidence interval is the range in which it is believed that the true value of a quantity lies. The level of belief is expressed by the probability, whose value is related to the size of the interval. It is one of the ways in which uncertainty can be expressed (see Estimation).

In practice a confidence interval is defined by a probability value, say 95%, and confidence limits on either side of the mean value $\overline{x}$. In this case the confidence limits *L1* and *L2* would be calculated from the probability density function such that there was a 95% chance of the true value of the quantity being estimated by $\overline{x}$ lying between *L1* and *L2*. Commonly *L1* and *L2* are the 2.5 percentile and 97.5 percentile respectively.

Example: 'An emission is between 90 and 100 kt with a probability of 95%.' Such a statement can be provided when the confidence interval is calculated (the numerical values in this example are arbitrarily chosen).

CONSISTENCY

Inventory definition: Consistency means that an inventory should be internally consistent in all its elements over a period of years. An inventory is consistent if the same methodologies are used for the base and all subsequent years and if consistent data sets are used to estimate emissions or removals from sources or sinks. Under certain circumstances referred to in paragraphs 10 and 11 of FCCC/SBSTA/1999/6 Add.1, an inventory using different methodologies for different years can be considered to be consistent if it has been recalculated in a transparent manner taking into account any good practices.

Statistical definition: A statistical estimator for a parameter is said to be consistent, if the estimator tends towards the parameter as the size of the sample used for the estimator increases – i.e. precision is improved by an increasing number of observations.

CORRELATION

Statistical definition: Mutual dependence between two quantities. See Correlation coefficient.

CORRELATION COEFFICIENT

Statistical definition: A number lying between –1 and +1 which measures the mutual dependence between two variables which are observed together. A value of +1 means that the variables have a perfect direct straight line relation; a value of –1 means that there is a perfect inverse straight line relation; and a value of 0 means that there is no straight line relation. It is defined as the covariance of the two variables divided by the product of their standard deviations.

---

[2] According to the instruments of ratification, this is the acceptance, approval or accession of the Convention by a given Party.

* See references (p A3.21).

COVARIANCE

Statistical definition: The covariance between two variables is a measure of the mutual dependence between two variables.

The sample covariance of paired sample of random variables *X* and *Y* is calculated using the following formula: $s_{xy}^2 = \frac{1}{n}\sum_i^n (x_i - \bar{x})(y_i - \bar{y})$ where x_i, y_i, $i = 1,\ldots,n$ are items in the sample and $\bar{x}$ and $\bar{y}$ are sample means.

CUMULATIVE DISTRIBUTION FUNCTION

See Distribution function.

DECISION TREE

Inventory definition: A decision tree is a flow chart describing the specific ordered steps which need to be followed to develop an inventory or an inventory component in accordance with the principles of good practice.

DISTRIBUTION FUNCTION

Statistical definition: A distribution function or cumulative distribution function *F*(*x*) for a random variable *X* specifies the probability Pr(*X* ≤ *x*) that *X* is less than or equal to *x*.

ELASTICITY

Statistical definition: Elasticity (or normalised sensitivity) is a measure of how responsive one quantity is to a change in another related quantity. The elasticity of a quantity *Y* that is affected by changes in another quantity *X* is defined as the percentage change in *Y* divided by the percentage change in *X* which caused the change in *Y*.

EMISSION FACTOR

Inventory definition: A coefficient that relates the activity data to the amount of chemical compound which is the source of later emissioms. Emission factors are often based on a sample of measurement data, averaged to develop a representative rate of emission for a given activity level under a given set of operating conditions (*Revised 1996 IPCC Guidelines* [9]*).

ERROR

Statistical definition: In statistical usage, the term ‘error’ is a general term referring to the difference between an observed (measured) value of a quantity and its ‘true’ (but usually unknown) value and does not carry the pejorative sense of a mistake or blunder.

ESTIMATION

Statistical definition: Estimation is the assessment of the value of a quantity or its uncertainty through the assignment of numerical observation values in an estimation formula, or estimator. The results of an estimation can be expressed as follows:

---

* See references (p A3.21).

- a point estimation which provide a number which can be used as an approximation to a parameter (such as the sample standard deviation which estimates the population standard deviation), or
- an interval estimate specifying a confidence level.

Example: A statement like ‘The total emission is estimated to be 100 kt and its coefficient of variation is 5%’ is based upon point estimates of the sample mean and standard deviation, whereas a statement such as ‘The total emission lies between 90 and 110 kt with probability 95%’ expresses the results of estimation as a confidence interval.

ESTIMATOR

Statistical definition: An estimator is a formula specifying how to calculate a sample estimate value of a population parameter from the sampled data. For example, emission factors are often estimated as the sample means of sets of measurements. There can be more than one estimator for a population parameter, and each estimator in general has its own sampling properties with consistency and unbiasedness being among the most important.

Examples of point estimators include the arithmetic mean $\bar{x}$, which is a commonly used estimator for the expected value (mean), and the sample variance s^2, which is a commonly used estimator for the variance.

EXPECTATION

Statistical definition: 1. For a discrete random variable X taking the values x_i with the probabilities p_i, the expectation is $\mu = E(X) = \Sigma\, p_i\, x_i$; and 2. For a continuous random variable X having the probability density function $f(x)$, the expectation, if it exists, is $\mu = E(X) = \int x\, f(x)\, dx$, the integral being extended over the interval(s) of variation of X. [7]*

EXPECTED VALUE

Statistical definition: See Mean.

EXPERT JUDGEMENT

Inventory definition: A carefully considered, well-documented qualitative or quantitative judgement made in the absence of unequivocal observational evidence by a person or persons who have a demonstrable expertise in the given field.

EXTREME VALUE

Statistical definition: The extreme values of a sample are the maximum and minimum values of the sample. The statistical theory of extreme values is concerned with estimating the distributions of these extreme values for large values of n.

GOOD PRACTICE

Inventory definition: Good Practice is a set of procedures intended to ensure that greenhouse gas inventories are accurate in the sense that they are systematically neither over nor underestimates so far as can be judged, and that uncertainties are reduced so far as possible.

Good Practice covers choice of estimation methods appropriate to national circumstances, quality assurance and quality control at the national level, quantification of uncertainties and data archiving and reporting to promote transparency.

---

* See references (p A3.21).

INDEPENDENCE

Statistical definition: Two random variables are independent if there is a complete absence of association between how their sample values vary. The most commonly used measure of the lack of independence between two random variables is the correlation coefficient.

KEY SOURCE CATEGORY

Inventory definition: A key source category is one that is prioritised within the national inventory system because its estimate has a significant influence on a country's total inventory of direct greenhouse gases in terms of the absolute level of emissions, the trend in emissions, or both. (See Chapter 7, Methodological Choice and Recalculation.)

KURTOSIS

Statistical definition: Kurtosis is a measure of the flatness of a PDF. It is a simple function of two moments.

Kurtosis is given by: $\gamma = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^2}$ where μ_2 and μ_4 are the second and fourth population central moments. For the normal distribution, kurtosis equals 3. The sample kurtosis has a corresponding definition, with sample moments replacing population moments; it is very sensitive to 'outlier' points.

LATIN HYPERCUBE SAMPLING

Statistical definition: Latin hypercube sampling is a technique of selecting values for inputs to computer realisation runs of a model by stratifying the range of each of the model inputs, and ensuring that input values across the whole of each model input range are selected.

LAW OF LARGE NUMBERS

Statistical definition: A mathematical theorem which formalises the generally known wisdom that an average becomes a better approximation to the mean as the number of observations is increased.

LINEAR MODEL

Statistical definition: A variable *y* is said to be linearly related to (or a linear function of) variables x_1, x_2, … if *y* can be expressed by the formula $y = b_0 + b_1x_1 + b_2x_2 + \ldots$ where the *b* terms are constant numbers.

Whether a function is considered to be linear or not can vary depending upon the context in which it is applied.

Example: An emission *E* is usually expressed as the product of an emission factor *F* and an activity level *A*. In the case where *F* is a fixed constant and *E* only varies when *A* does, *E* is linearly related to *A*. However, when both *F* and *A* are considered to be variables (such as when applying the error propagation equation to estimate the variance of *E* as a function of the variances and covariance of *A* and *F*) *E* is not a linear function of *F* and *A*.

LINEAR REGRESSION

Statistical definition: Linear regression provides a way of fitting a straight line to a set of observed data points, taking into account the effects of observational variability.

Example: If emissions observations are plotted against corresponding activity levels, the slope of the line fitted by a linear regression provides an estimate of the appropriate emission factor. The technique can also be used for estimating a straight-line trend for a quantity which varies over time.

LOGNORMAL DISTRIBUTION

Statistical definition: The lognormal distribution is an asymmetric distribution, which starts from zero, rises to a maximum and then tails off more slowly to infinity. It is related to the normal distribution: X has a lognormal distribution if ln (X) has a normal distribution.

The PDF of the lognormal distribution is given by:

$$f(x) = \frac{1}{\sigma_l x \sqrt{2\pi}} e^{-\frac{(\ln x - \mu_l)^2}{2\sigma_l^2}} \quad , \text{ for } 0 \le x \le \infty$$

The parameters required to specify the function are: μ_l the mean of the natural log transform of the data; and $\sigma_l^{\ 2}$ the variance of the natural log transform of the data. The data and information that the inventory compiler can use to determine the input parameters are: mean = μ; variance = σ^2; and the relationships:

$$\mu_l = \ln \frac{\mu^2}{\sqrt{(\sigma^2 + \mu^2)}}$$

and

$$\sigma_l = \sqrt{\ln\left(\frac{\sigma^2}{\mu^2} + 1\right)}$$

MEAN

Statistical definition: The mean, population mean, expectation or expected value is, broadly speaking, a measure of a central value around which values sampled from a probability distribution tend to lie. The sample mean or arithmetic average is an estimator for the mean. It is an unbiased and consistent estimator of the population mean (expected value) and itself is a random variable with its own variance value. The **sample** mean is the sum of values divided by the number of values:

$$\bar{x} = \frac{1}{n}\sum_{i}^{n} x_i \quad (x_i \text{ , } i = 1,..,n \text{ are items of a sample})$$

MEDIAN

Statistical definition: The median or population median is a value which divides the integral of a PDF into two halves. For symmetric PDFs, it equals the mean. The median is the 50th population percentile.

The sample median is an estimator of the population median. It is the value that divides an ordered sample into two equal halves. If there are $2n + 1$ observations, the median is taken as the $(n + 1)^{th}$ member of the ordered sample. If there are $2n$, it is taken as being halfway between the nth and $(n + 1)^{th}$.

MODE

Statistical definition: Distributions can have one or more modes. In practice, we usually encounter distributions with only one mode. In this case, **the mode or population mode of a PDF is the measure of a central value around which values sampled from a probability distribution tend to lie and is broadly speaking, the value which has the highest probability of occurrence**.

The sample mode is an estimator for the population mode calculated by subdividing the sample range into equal subclasses, counting how many observations fall into each class and selecting the centre point of the class (or classes) with the greatest number of observations.

MODEL

Statistical definition: A model is a quantitatively-based abstraction of a real-world situation which may simplify or neglect certain features to better focus on its more important elements.

Example: the relationship that emissions equal an emission factor times an activity level is a simple model. The term ‘model’ is also often used in the sense of a computer software realisation of a model abstraction which calculates a set of output values for a given set of input values – e.g. numerical global climate models.

MOMENTS (OF RANDOM VARIABLE)

Statistical definition: A population moment of a variable *X* about a given constant α is defined as the expected value of the random variable ($X – \alpha)^k$, i.e. E($X – \alpha)^k$. For the case where α equals the population mean, μ, the moment E($X – \mu)^k$ is termed the k^{th} central moment of *X*. They are important because statistical calculations are usually based upon the moments of the PDF rather than the PDF itself. The most commonly encountered moments are the mean and the variance.

The sample mean is the first moment around zero and the variance is the second central moment. Skewness and kurtosis are two frequently used functions of central moments which characterise the shape of the PDF.

The sample moments are estimators of population moments. The sample moment of k^{th} order is the arithmetic mean of the k^{th} power of the difference between the observed values and their average.

MONTE CARLO METHOD

Inventory definition: The principle of Monte Carlo analysis is to perform the inventory calculation many times by electronic computer, each time with the uncertain emission factors or model parameters and activity data chosen randomly (by the computer) within the distribution on uncertainties specified initially by the user. Uncertainties in emission factors and/or activity data are often large and may not have normal distributions. In this case the conventional statistical rules for combining uncertainties become very approximate. Monte Carlo analysis can deal with this situation by generating an uncertainty distribution for the inventory estimate that is consistent with the input uncertainty distributions on the emission factors, model parameters and activity data.

NON-LINEAR MODEL

Statistical definition: A model is non-linear if the relationship between its inputs and its outputs is non-linear (see Linear model).

NORMAL DISTRIBUTION

Statistical definition: The normal (or Gaussian) distribution has the PDF given in the following equation and is defined by two parameters (the mean μ and the standard deviation σ):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{, for } -\infty \leq x \leq \infty$$

PARAMETERS OF POPULATION

Statistical definition: Parameters of the probability distribution that characterise the population. The most commonly used population parameters are the moments – e.g. the mean and the standard deviation for the normal distribution. A quantity used in describing the probability distribution of a random variable. [7]*

* See references (p A3.21).

PDF

See Probability density function.

PERCENTILE

Statistical definition: The k^{th} percentile or population percentile is a value which separates the lowest k^{th} part of the integral of the PDF – i.e. an integral of a PDF tail from the k^{th} percentile towards lower probability densities.

The k^{th} population percentile ($0 \le k \le 100$) of a population with a distribution function $F(x)$ equals to z where z satisfies $F(z) = k/100$

Sample k^{th} percentile is an approximation for the population percentile which is derived from a sample. It is the value below which *k* percent of the observations lie.

POPULATION

Statistical definition: The population is the totality of items under consideration. In the case of a random variable, the probability distribution is considered to define the population of that variable [7]*.

Example: all conceivable experiments or events of a given type.

PRECISION

Inventory definition: Precision is the inverse of uncertainty in the sense that the more precise something is, the less uncertain it is.

PROBABILITY

Statistical definition: A probability is a real number in the scale 0 to 1 attached to a random event. ([7]*, C.2.1) There are different ways in which probability can be interpreted. One interpretation considers a probability as having the nature of a relative frequency (i.e. the proportion of all outcomes corresponding to an event), whilst another interpretation regards a probability as being a measure of degree of belief. The probability that a random event *E* occurs is often denoted as Pr(*E*). Probabilities may also be expressed in percentage terms. Probability theory is a branch of mathematics developed from axiomatic foundations, whose results underlay statistical inference.

PROBABILITY DENSITY FUNCTION – PDF

Statistical definition: A probability density function (PDF) is a mathematical function which characterises the probability behaviour of population. It is a function $f(x)$ which specifies the relative likelihood of a continuous random variable X taking a value near x, and is defined as the probability that X takes a value between x and $x+dx$, divided by dx where dx is an infinitesimally small number. Most PDFs require one or more parameters to specify them fully.

The probability that a continuous random variable X lies in between the values a and b is given by the interval of the PDF, $f(x)$, over the range between a and b.

$$\Pr(a \le x < b) = \int_b^a f(x)dx$$

The PDF is the derivative (when it exists) of the distribution function:

* See references (p A3.21).

$$f(x) = \frac{dF(x)}{dx}$$

In practical situations, the PDF used is chosen from a relatively small number of standard PDFs and the main statistical task is to estimate its parameters. Thus, for inventory applications, a knowledge of which PDF has been used is a necessary item in the documentation of an uncertainty assessment.

PROBABILITY DISTRIBUTION

Statistical definition: A function giving the probability that a random variable takes any given value or belongs to a given set of values. The probability on the whole set of values of the random variable equals 1. [7]*

PROPAGATION OF UNCERTAINTIES

Statistical definition: The rules for propagation of uncertainties specify how to algebraically combine the quantitative measures of uncertainty associated with the input values to the mathematical formulae used in inventory compilation, so as to obtain corresponding measures of uncertainty for the output values. See Chapter 6, Quantifying Uncertainties in Practice, and Annex 1, Conceptual Basis for Uncertainty Analysis.

QUALITY ASSURANCE (QA)

Inventory definition: Quality Assurance (QA) activities include a planned system of review procedures conducted by personnel not directly involved in the inventory compilation/development process to verify that data quality objectives were met, ensure that the inventory represents the best possible estimate of emissions and sinks given the current state of scientific knowledge and data available, and support the effectiveness of the quality control (QC) programme.

QUALITY CONTROL (QC)

Inventory definition: Quality Control (QC) is a system of routine technical activities, to measure and control the quality of the inventory as it is being developed. The QC system is designed to:

- **(i) Provide routine and consistent checks to ensure data integrity, correctness, and completeness;**
- **(ii) Identify and address errors and omissions;**
- **(iii) Document and archive inventory material and record all QC activities.**

QC activities include general methods such as accuracy checks on data acquisition and calculations and the use of approved standardised procedures for emission calculations, measurements, estimating uncertainties, archiving information and reporting. Higher tier QC activities include technical reviews of source categories, activity and emission factor data, and methods.

RANDOM ERROR

See Systematic and random errors.

RANDOM VARIABLE

Statistical definition: A variable that may take any of the values of a specified set of values and with which

* See references (p A3.21).

is associated a probability distribution. A random variable which may take only isolated values is said to be 'discrete.' A random variable that may take any value within a finite or infinite interval is said to be 'continuous'. [7]*

RESIDUAL

Statistical meaning: For an observed value whose behaviour is modelled by a statistical model, the residual is the difference between the observed value and the value predicted by the model, e.g. by linear regression. The residual is thus the component of an observation that cannot be explained by the model.

SAMPLE

Statistical meaning: A sample is a finite set of observations drawn from a population.

SENSITIVITY

Statistical definition: A sensitivity is a measurement of how responsive one quantity is to a change in another related quantity. The sensitivity of a quantity *Y* that is affected by changes in another quantity *X*, is defined as the change in *Y* divided by the change in *X* that caused the changes in *Y*.

SENSITIVITY ANALYSIS

Statistical definition: Sensitivity analysis is a study of a model algorithm to determine how sensitive (or stable) it is to variations of its input data or underlying assumptions. It is performed by varying input values or model equations and observing how the model output varies correspondingly. The aim of such a sensitivity analysis can include:

- observing the range of output values corresponding to input variables lying within 'reasonable' ranges; and
- calculating finite difference approximations for elasticities and sensitivities as required by some methodologies for studying error propagation within a system.

SIGMA INTERVAL

Statistical definition: A *c*-sigma interval is a symmetric confidence interval centred on the mean and extending *c* times the standard deviation on either side.

SIMPLE RANDOM SAMPLE

Statistical definition: A sample of *n* items chosen from a population such that every possible sample has the same probability of being chosen.

SKEWNESS

Statistical definition: Skewness is a measure of asymmetry of a PDF. It is a simple function of two moments of the PDF, given by: $\gamma = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3}$ where μ_2, μ_3, and σ, are central moments. Symmetric distributions have $\gamma = 0$. The same name is frequently used for sample skewness, in which case both population moments are replaced by sample moments.

---

* See references (p A3.21).

STANDARD DEVIATION

Statistical definition: The population standard deviation is the positive square root of the variance. It is estimated by the **sample standard deviation** that is the positive square root of the sample variance.

STANDARD ERROR OF THE MEAN

Statistical definition: A term often used to signify the sample standard deviation of the mean.

STATISTIC

Statistical definition: A statistic is a function of the sample random variables. [7]*

STATISTICS

Statistical definition: Statistics can refer either in a general sense to the compilation of data, frequently about human activities, or in a more specific sense to the branch of science concerned with the systematic numerical treatment of data derived from aggregates of items.

SYSTEMATIC AND RANDOM ERRORS

Statistical definition: Systematic error is the difference between the true, but usually unknown, value of a quantity being measured, and the mean observed value as would be estimated by the sample mean of an infinite set of observations. The random error of an individual measurement is the difference between an individual measurement and the above limiting value of the sample mean.

SYSTEMATIC ERROR

Statistical definition: See Systematic and random errors.

TIME SERIES

Statistical definition: A time series is series of values which are affected by random processes and which are observed at successive (but usually equidistant) time points.

TRANSPARENCY

Inventory definition: Transparency means that the assumptions and methodologies used for an inventory should be clearly explained to facilitate replication and assessment of the inventory by users of the reported information. The transparency of inventories is fundamental to the success of the process for the communication and consideration of information.

TREND

Inventory definition: The trend of a quantity measures its relative trend over a time period, with a positive trend value indicating growth in the quantity, and a negative value indicating a decrease. It is defined as the ratio of the change in the quantity over the time period, divided by the initial value of the quantity, and is usually expressed either as a percentage or a fraction.

---

* See references (p A3.21).

TRIANGULAR DISTRIBUTION

Statistical definition: An asymmetric triangular distribution function has a PDF

$f(x) = 2(x - a) / \{(b - a)(m - a)\}$ when $a \le x \le m$ and $a < m \le b$

$= 2(b - x) / \{(b - a)(b - m)\}$ when $m \le x \le b$ and $a \le m < b$

= 0 elsewhere,

where the parameters which specify the distribution are the minimum value *a*, the maximum value *b*, and the most likely position (i.e. mode) *m*, subject to $a \le m \le b$.

UNBIASED ESTIMATOR

Statistical definition: An unbiased estimator is a statistic whose expected value equals the value of the parameter being estimated. Note that this term has a specific statistical meaning and that an estimate of a quantity calculated from an unbiased estimator may lack bias in the statistical sense, but may be biased in the more general sense of the word if the sample has been affected by unknown systematic error. Thus, in statistical usage, a biased estimator can be understood as a deficiency in the statistical evaluation of the collected data, and not in the data themselves or in the method of their measurement or collection. For example, the arithmetic mean (average) $\overline{x}$ is an unbiased estimator of the expected value (mean).

UNCERTAINTY

Statistical definition: An uncertainty is a parameter, associated with the result of measurement that characterises the dispersion of the values that could be reasonably attributed to the measured quantity. [7]* (e.g. the sample variance or coefficient of variation)

Inventory definition: A general and imprecise term which refers to the lack of certainty (in inventory components) resulting from any causal factor such as unidentified sources and sinks, lack of transparency etc.

UNCERTAINTY ANALYSIS

Statistical definition: An uncertainty analysis of a model aims to provide quantitative measures of the uncertainty of output values caused by uncertainties in the model itself and in its input values, and to examine the relative importance of these factors.

UNIFORM DISTRIBUTION

Statistical definition: A random variable with a uniform or rectangular distribution is confined to lie within a range over which all values are equally probable. If the upper and lower limits of the range are *a* and *b* respectively, the PDF is a flat function from *a* to *b* (the two parameters defining the PDF).

The PDF of a uniform distribution is given by:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{elsewhere} \end{cases}$$

where

$$\mu = \frac{a+b}{2}$$

is the mean and

* See references (p A3.21).

$$\sigma^2 = \frac{(b-a)^2}{12}$$

is the variance.

VALIDATION

Inventory definition: Validation is the establishment of sound approach and foundation. In the context of emission inventories, validation involves checking to ensure that the inventory has been compiled correctly in line with reporting instructions and guidelines. It checks the internal consistency of the inventory. The legal use of validation is to give an official confirmation or approval of an act or product. [6]*

VARIABILITY

Statistical definition: This refers to observed differences attributable to true heterogeneity or diversity in a population. Variability derives from processes which are either inherently random or whose nature and effects are influential but unknown. Variability is not usually reducible by further measurement or study, but can be characterised by quantities such as the sample variance. [6]*

VARIANCE

Statistical definition: The variance or population variance is a parameter of a PDF, which expresses the variability of the population. It is the second central moment of a random variable. The **sample variance** is defined as a measure of dispersion, which is the sum of the squared deviations of observations from their average, divided by one less than the number of observations. [7]* $s^2 = \frac{1}{n-1}\sum_{i}^{n}(x_i - \bar{x})^2$

VARIANCE OF SAMPLE MEAN

Statistical definition: The mean of a sample taken from a population is itself a random variable with its own characteristic behaviour and its own variance. For such sample means, the appropriate estimate of the variance is not the sample variance, which estimates the variability associated with a single simple value, but a lower value, equal to the sample variance divided by the sample size.

VERIFICATION

Inventory definition: Verification refers to the collection of activities and procedures that can be followed during the planning and development, or after completion of an inventory that can help to establish its reliability for the intended applications of that inventory. Typically, methods external to the inventory are used to check the truth of the inventory, including comparisons with estimates made by other bodies or with emission and uptake measurements determined from atmospheric concentrations or concentration gradients of these gases. [6]*

---

* See references (p A3.21).

REFERENCES

[1] Brandt S. (1970). *Computational Methods in Statistics and Data Analysis.* North-Holland, Amsterdam, Netherlands.

[2] Wonnacott R. and Wonnacott T. (1990). *Introductory Statistics for Business and Economics*. John Wiley, New York, USA.

[3] Zijp W.L. (1987). *Treatment of measurement of Uncertainties.* ECN-194, Petten, Netherlands.

[4] Martin B.R. (1971). *Statistics for Physicists*. Academic Press, London, UK.

[5] Kendall M. G, Stuart A. (1966). *The Advanced Theory of Statistics*. Charles Grifiin & Co.Ltd, London, UK.

[6] Intergovernmental Panel on Climate Change (IPCC) (1998). *Meeting Report: Managing Uncertainty in National Greenhouse Gas Inventories*, Report of the meeting held in Paris 13-15, October 1998, IPCC/OECD/IEA Programme on National Greenhouse Gas Inventories.

[7] International Organization for Standardization (ISO) (1993). *Guide to the expression of uncertainty in measurement.* ISBN 92-67-10188-9, ISO, Geneva, Switzerland.

[8] Kotz, S. and Johnson, N.L. (1988). *Encyclopaedia of Statistical Sciences.* 9 volumes, John Wiley & Sons, New York, USA.

[9] Intergovernmental Panel on Climate Change (IPCC) (1997). *Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories: Volume 1 Reporting Instructions*. J.T. Houghton et al., IPCC/OECD/IEA, Paris, France.