Annex 1: Glossary of Terms

This Glossary contains terms identified at the Paris Workshop, and some added subsequently. It is provisional, and comments are invited both on the definitions provided, and on additional terms. Please send comments to Ian Galbally (name and contact details in Annex 6)

Absolute Error [1]

The absolute deviation of a variable from its 'true' value. See also *Deviation from the Mean*.

Accuracy [2]

The tendency of values of an *estimator* to come close to the quantity they are intended to estimate. See also *Precision*. So far as inventories are concerned accuracy means that estimates are neither over estimates nor underestimates of true avulse, so far as can be judged, and that the uncertainties are reduced so far as practicable

Arithmetic Mean[8]

The arithmetic mean of n quantities $x_1, x_2, \ldots, x_n$ is the sum of the quantities divided by their number n.

Average

A generic term for a measure of central tendency for a set of sample observations.

Bias [2]

(1) In problems of estimation, an estimator is said to be *biased* if its expected value does *not* equal the parameter it is intended to estimate. (2) In sampling, a *bias* is a systematic error introduced by selecting items from a wrong population, favouring some of the elements of a population, or poorly phrasing questions.

Comparability

In the IPCC inventory context inventories are said to be *comparable* with the Guidelines if they are produced using the Guidelines, or by using methods that are mathematically equivalent to those in the Guidelines, or which can be shown to give more accurate estimates than the methods in the Guidelines for source categories contained in the Guidelines.

Compliance [3]

Compliance is the act of conforming or yielding to a specified norm or protocol. In the inventory development process, compliance indicates conformity to development protocols or international agreements. In this sense the compliance issue can be thought of as verification of adherence to these established and agreed norms.

Confidence [3]

The term confidence is used to represent trust in a measurement or estimate. Having confidence in inventory estimates does not make those estimates accurate or precise, but will help to develop a consensus that the data can be applied to problem solving.

Confidence Interval [2]

An interval for which one can assert with a given probability, called the *degree of confidence* or the *confidence coefficient*, that it will contain the true value of the parameter it is intended to estimate. The endpoints of a confidence interval are referred to as the (*upper and lower*) *confidence limits*; they

are generally values of random variables calculated on the basis of sample data. In the IPCC Guidelines specified confidence intervals correspond to a two standard deviations or 95%.

Consistency [2]

An *estimator* (depending on the sample size of *n*) is said to be *consistent* if the probability that it will assume a value arbitrarily close to the parameter it is intended to estimate *approaches one* when *n* becomes infinite. In the IPCC inventory context, consistency can mean that the methods used are the same throughout the time series being reported.

Correlation [2]

In general, the term denotes the relationship (association or dependence) between two or more qualitative *or* quantitative variables. See also *Correlation Coefficient*.

Correlation Coefficient [2]

A measure of the linear relationship between two quantitative variables. It is denoted by the letter *r* and its values range from -1 to +1, where 0 indicates the absence of linear relationship, while -1 and +1 indicate, respectively, a perfect *negative* (inverse) and a perfect *positive* (direct) relationship.

Deviation from the Mean [2]

For a given set of data, the amount by which an individual observation differs from the mean; thus, the deviation of the *ith* observation x_i from the mean $\bar{x}$ is given by $x_i - \bar{x}$. An important property of deviations from the mean is that, for any set of data, their sum is always equal to zero. The absolute value of $x_i - \bar{x}$, namely $| x_i - \bar{x} |$, is referred to as an *absolute deviation* from the mean.

Distribution function [2]

A function whose values F(t) are the probabilities that a random variable assumes a value less than or equal to *t*.

Error of estimation [1]

In regression analysis where the regression equation is used to estimate the 'dependent' from given values of the 'independent' variates, the difference between the estimated and the observed value of the dependent variate.

Estimate [2]

A number or an interval, based on sample data produced by an estimator; also the result of a model calculation or other estimation procedure.

Estimator [1]

A rule or method of estimating a constant of a parent population.

Expected Value[8]

The expected value of a random variable is a theoretically defined quantity intended to be an analogue of the (observable) arithmetic mean. It can be regarded as corresponding to a limit of the arithmetic mean as the sample size increases indefinitely.

Gaussian Distribution

See *Normal Distribution*

Mean

See *Arithmetic Mean*

Median[9]

The median value of a sample is the value which 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.

Normal Distribution

The normal distribution is the probability distribution function whose probability density function is:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right] \quad \text{for } -\infty < x < \infty$$

Precision [2]

The precision of an *estimator* is its tendency to have its values cluster closely about the expected value of its sampling distribution; thus, it is related inversely to the *variance* of this sampling distribution - the smaller the variance, the greater the precision.

Probability density function [2]

A function with non-negative values, whose integral from a to b ($a \leq b$) gives the probability that a corresponding random variable assumes a value on the interval from a to b.

Probability distribution [2]

The probability distribution (or simply the distribution) of a random variable is its *probability structure* as described, for example, by a *probability density function* or a probability distribution function.

Quality Assurance [3]

In the context of emissions inventory development, quality assurance is used to represent the sum of activities that are implemented to ensure the collection and presentation of high quality data.

More generally, quality assurance describes the activities that are completed after the development of a product, usually by an independent party to verify that data quality objectives were met and that the product conforms to specifications. In experimental programs, audits with standard instruments and standard measures are used to establish the reliability of the experimental procedures. In the calculated emission inventories development, however, few such standards have been developed so far.

Quality Control [3]

Quality control activities are those procedures and tests that can be performed during the planning and development of an inventory to ensure that the data quality objectives are being met. These activities may include criteria tests for data on operations, completeness criteria, or averaging techniques for use in developing default parameters. Quality control activities are generally applied by the developers.

Random error [4]

Is the result of a measurement minus its expected value. Random error is equal to absolute error minus systematic bias. Because only a finite number of measurements can be made, it is possible to determine only an estimate of random error.

Reliability [2]

When speaking of the *reliability* of an estimator, one is referring to the possible size of chance errors. This term is often used interchangeably with *consistency*. In the context of emissions inventories,

reliability and confidence are closely linked. If the approaches and data sources used in an inventory development project are considered reliable, then users will have an acceptable degree of confidence in the emissions data developed from those techniques.

Sample

A *sample* among other meanings, can be a set of data.

Sampling Distribution [2]

The distribution of a statistic; for example, the distribution of the arithmetic mean for random samples from normal populations, or the distribution of the coefficient of correlation under the assumptions of normal correlation analysis.

Sensitivity Analysis [5, 6]

(1) Determining the contribution of individual analysis inputs to the variability in model parameters. (2) The systematic investigation of the reaction of the simulation and response to either extreme values of the model's quantitative factors (parameter and input variables) or to drastic changes in the model's quantitative factors (modules). So the focus is not on marginal changes in inputs.

Standard Deviation [2]

The standard deviation of a sample of size *n* (usually called the "sample standard deviation") is given by the square root of the sum of the squared deviations from the arithmetic mean divided by *n* -1.

Systematic error

See *Bias*

Transparency [3]

Transparency is used to represent the condition of being clear and free from pretence. In the context of compiling emission inventories under the UNFCCC this means (1) the construction of the emission estimates is clearly explained; (2) the documentation of the inventory is sufficient for another party to reconstruct it; (3) the documentation sufficiently clarifies the major causes of emission trends in the inventory. Transparency will be greatly increased if the data collected and reported by different agencies will be similar and, therefore, easily understood by other parties and comparable to the data presented by the other parties.

Uncertainty [3]

Uncertainty is a statistical term that is used to represent the degree of accuracy and precision of data. It often expresses the range of possible values of a parameter or a measurement around a mean or preferred value.

Uncertainty analysis [6]

In uncertainty analysis values of the model inputs are sampled from pre-defined distributions to quantify the consequences of the uncertainties in the model inputs, for the model outputs. So in uncertainty analysis the input variables range between extreme values investigated in *sensitivity analysis*.

Validation [3]

Validation is the establishment of sound approach and foundation. The legal use of validation is to give an official confirmation or approval of an act or product. 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.

Variability [7]

This refers to observed differences attributable to true heterogeneity or diversity in a population or parameter. Sources of variability are the result of random processes. Variability is usually not reducible by further measurement or study, but can be characterised.

Variance [8]

The *variance* of a sample is (1) the square of the standard deviation (2) the second central moment of a population.

Verification [3]

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 and/or concentration gradients of these gases.

Definitions are extracted from:

1. M.G. Kendall and W.R. Buckland, *A Dictionary of Statistical Terms*; International Statistics Institute; UK 1975.
2. Freund, J.E., and F.J. Williams, *Dictionary/Outline of Basic Statistics*, 195 pp., Dover Publications, Inc., New York, NY, 1966.
3. The European Environment Agency, *Atmospheric Emission Inventory Guidebook (CD-ROM)*, section on Procedures for Verification of Emission Inventories; Copenhagen, 1997.
4. International Organisation of Standardisation, *International Vocabulary of Basic and General Terms in Meteorology* (VIM),1993.
5. J.C. Helton, *Uncertainty and Sensitivity Analysis in Performance Assessment for the Waste Isolation Pilot Plant*, In: Second International Symposium on Sensitivity Analysis of Model Output (SAMO, 98), edited by K. Chan, S. Tarantola, and F. Campolongo, Office of Official Publications of the European Communities, Luxembourg, 1998.
6. Kleijnen, J.P.C., *Statistical Theory on Design of Experiments, Applied to Simulation.* In Second International Symposium on Sensitivity Analysis of Model Output (SAMO '98), edited by K. Chan, S. Tarantola, and F. Campolongo, Office of Official Publications of the European Communities, Luxembourg, 1998.
7. United States Environmental Protection Agency, *Guiding principles for Monte Carlo Analysis*, 1997.
8. Kotz, S. and Johnson, N.L., *Encyclopedia of Statistical Sciences*, 9 volumes, John Wiley & Sons, New York, 1988.
9. Kendall M.G. and Stuart A., *The Advanced Theory of Statistics*, Charles Griffin and Company Ltd, London, 1969.