6

QUANTIFYING UNCERTAINTIES IN PRACTICE

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)

Expert Group on Quantifying Uncertainties in Practice

CO-CHAIRS

Jim Penman (UK) and Semere Habetsion (Eritrea)

AUTHORS OF BACKGROUND PAPERS

Kay Abel (Australia), Simon Eggleston (UK) and Tinus Pullus (Netherlands)

CONTRIBUTORS

Simon Eggleston (UK), Christopher Frey (USA), Kari Gronfors (Finland), Niklas Höhne (UNFCCC Secretariat), Charles Jubb (Australia), Katarina Mareckova (Slovak Republic), Jeroen Meijer (IPCC-NGGIP/TSU), Frank Neitzert (Canada), Todd Ngara (Zimbabwe), Tinus Pulles (Netherlands), Emmanuel Rivière (France), Arthur Rypinski (USA), Martiros Tsarukyan (Armenia), and Peter Zhou (Botswana)

Contents

Figures

Tables

6 QUANTIFYING UNCERTAINTIES IN PRACTICE

6.1 OVERVIEW

This chapter describes *good practice* in estimating and reporting uncertainties associated with both annual estimates of emissions, and emission trends over time. It identifies types of uncertainty from the viewpoint of the inventory practitioner and shows how to obtain expert judgements in a consistent manner. It provides two tiers for combining source category uncertainties into an uncertainty estimate for total national emissions, and presents an example of the application of the Tier 1 method.

The chapter is consistent with source-specific *good practice guidance* described in Chapters 2 to 5, the general principles discussed in Annex 1, Conceptual Basis for Uncertainty Analysis, and the chapters on methodological choice (Chapter 7, Methodological Choice and Recalculation) and QA/QC (Chapter 8, Quality Assurance and Quality Control).

Uncertainty estimates are an essential element of a complete emissions inventory. Uncertainty information is not intended to dispute the validity of the inventory estimates, but to help prioritise efforts to improve the accuracy of inventories in the future and guide decisions on methodological choice, as indicated in Chapter 7, Methodological Choice and Recalculation. Inventory practitioners understand that for most countries and source categories, greenhouse gas emissions estimates are reasonably accurate. However, national inventories prepared pursuant to the *Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories* (*IPCC Guidelines*) will typically contain a wide range of emission estimates, varying from carefully measured and demonstrably complete data on emissions of certain engineered chemicals, to order-of-magnitude estimates of highly variable nitrous oxide (N_2O) fluxes from soils and waterways.

Inventory estimates can be used for a range of purposes. For some purposes, only the national total matters, while for others, the detail by greenhouse gas and source category is important. In order to match the data to the intended purpose, users must be able to understand the actual reliability of both the total estimate and its component parts. For this reason, the methods used to communicate uncertainty must be practical, scientifically defensible, robust enough to be applicable to a range of source categories, methods and national circumstances, and presented in ways comprehensible to non-specialist inventory users.

There are many reasons that actual emissions and sinks may differ from the number calculated in a national inventory. These reasons are discussed at greater length in Annex 1. Some sources of uncertainty (i.e. sampling error or limitations on instrument accuracy) may generate well-defined, easily characterised estimates of the range of potential error. However, other sources of uncertainty may be much more difficult to characterise. This chapter describes how to account for both well-defined statistical uncertainties and less specific information characterising other forms of uncertainty, and how to combine this information into a characterisation of the uncertainty of both the total inventory and its components.

Ideally, emissions estimates and uncertainty ranges would both be derived from source-specific measured data. Since it is not practical to measure every emission source in this way, estimates are often based on the known characteristics of typical sources taken to be representative of the population. This introduces additional uncertainties, because it must be assumed that the population of these sources behave, on average, like the sources that have been measured. Sometimes enough will be known about these typical sources to determine their uncertainty distributions empirically. In practice, however, expert judgement will often be necessary to define the uncertainty ranges.

The pragmatic approach to producing quantitative uncertainty estimates in this situation is to use the best available estimates; a combination of the available measured data and expert judgement. The methods proposed in this chapter can therefore be used with the source category-specific uncertainty ranges discussed in Chapters 2 to 5, and also allow for new empirical data to be incorporated as they become available. This chapter also describes methods for eliciting expert judgement in a way that minimises the risk of bias, and it discusses how to combine uncertainties in emission factors and activity data to estimate source category and total uncertainties in inventories, as well as uncertainties in the trend.

The chapter uses two main statistical concepts – the probability density function and confidence limits – that are defined formally in Annex 3, the Glossary, and discussed in more detail in Annex 1, Conceptual Basis for

Uncertainty Analysis. Briefly, the probability density function describes the range and relative likelihood of possible values. Confidence limits give the range within which the underlying value of an uncertain quantity is thought to lie for a specified probability. This range is called the confidence interval. The *IPCC Guidelines* suggest the use of a 95% confidence interval which is the interval that has a 95% probability of containing the unknown true value.

The analysis of uncertainty presented in this chapter does not consider the uncertainties in Global Warming Potentials (GWPs). For reporting, the GWP values adopted at the Third Session of the UNFCCC Conference of the Parties become in effect fixed weighting factors. However, it should be kept in mind that GWP values actually have significant uncertainties associated with them and that an overall assessment of total equivalent emissions should take this fact into account.

6.2 IDENTIFYING UNCERTAINTIES

The estimated uncertainty of emissions from individual sources (e.g. power plants, motor vehicles, dairy cattle) is either a function of instrument characteristics, calibration and sampling frequency of direct measurements, or (more often) a combination of the uncertainties in the emission factors for typical sources and the corresponding activity data. The uncertainties in the emission factors and activity data should be described using probability density functions. Where data are available to do so, the shape of the probability density function should be determined empirically. Otherwise, expert judgement will be necessary, following the rules set out in Section 6.2.5, Expert Judgement, below. Sections 6.2.1 to 6.2.4 (below) give examples of typical situations that can arise under different circumstances of data availability. These subsections are ranked in order of their desirability in producing uncertainty assessments.

Uncertainties are affected by choice of estimation algorithm, and this is reflected in *good practice*, where higher tier methods (provided they are well implemented) should usually be associated with lower uncertainties. In general, uncertainties related to model choice will be reflected in the uncertainty ranges derived for use in the context of the model selected.

6.2.1 Uncertainties associated with continuous monitoring of emissions

Continuous monitoring of emissions, though comparatively rare, is usually consistent with source category-specific *good practice*. In this case, the probability density function, and therefore the uncertainty in emissions including 95% confidence limits, can be determined directly. Representative sampling requires that the equipment used to make the measurements be installed and operated according to the principles and references set out in Chapter 8 on QA/QC issues. Provided this is done, it is unlikely that there will be correlation of errors between years. Therefore, the probability density function of the difference in emissions between two years (the trend uncertainty) will be simply related to the probability density functions of the annual emissions. Assuming both probability density functions are normal, the probability density function of the difference in emissions will also be normal with:

EQUATION 6.1

$$\text{mean} = \mu_1 - \mu_2$$

EQUATION 6.2

$$\text{standard deviation} = (\sigma_1^2 + \sigma_2^2)^{1/2}$$

where μ_1 and μ_2 are the mean values of the emissions in years t_1 and t_2, and σ_1 and σ_2 are the standard deviations of the probability density functions of the emissions in year t_1 and t_2. The 95% confidence limits (this time of the mean or the difference in the means) will be given by plus or minus approximately two standard deviations.[1]

[1] For sample sizes less than about 30, a Student *t* distribution should be used to estimate confidence intervals.

6.2.2 Uncertainties associated with direct determination of emission factors

In some cases, periodic emission measurements may be available at a site. If these measurements can be linked to representative activity data, which of course is crucial, then it is possible to determine a site-specific emission factor, together with an associated probability density function to represent annual emissions.

This can be a complex task. To achieve representativeness it may be necessary to partition (or stratify) the data to reflect typical operating conditions. For example:

- *Start-up and shut down can give different emission rates relative to activity data*. In this case, the data should be partitioned, with separate emission factors and probability density functions derived for steady state, start-up and shut down conditions.
- *Emission factors can depend on load*. In this case, the total emissions estimation and uncertainty analysis may need to be stratified to take account of load, expressed, for example, as percentage of full capacity. This could be done by regression analysis and scatter plots of the emission rate against likely controlling variables (e.g. emissions *versus* load) with load becoming part of the activity data needed.
- *Measurements taken for another purpose may not be representative*. For example, methane measurements made for safety reasons at coal mines and landfills may not reflect total emissions. In such cases, the ratio between the measured data and total emissions should be estimated for the uncertainty analysis.

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 probability density function to use for describing variability in the data (partitioned if necessary) and how to parameterise it. However, in many cases, the number of measurements from which to make an inference regarding uncertainty will be small. Typically, as long as there are three or more data points, and as long as the data are a random representative sample of the quantity of interest, it is possible to apply statistical techniques to estimate the values of the parameters of many two-parameter distributions (e.g. normal, lognormal) that can be used to describe variability in the data set (Cullen and Frey, 1999, pp. 116-117). With small sample sizes, there will be large uncertainties regarding the parameter estimates that should be reflected in the quantification of uncertainty for use in the emissions inventory. Furthermore, it is typically not possible to rely on statistical methods to differentiate goodness-of-fit of alternative parametric distributions when sample sizes are very small (Cullen and Frey, 1999, pp. 158-159). Therefore, considerable judgement is required in selecting an appropriate parametric distribution to fit to a very small data set. In situations where the coefficient of variation is less than approximately 0.3, a normal distribution may be a reasonable assumption (Robinson, 1989). When the coefficient of variation is large and the quantity is non-negative, then a positively skewed distribution such as a lognormal one may be appropriate. Guidance on the selection of distributions is provided in Annex 1, Conceptual Basis for Uncertainty Analysis, and the use of expert judgements in this context is outlined in Section 6.2.5, Expert Judgement, below.

6.2.3 Uncertainties associated with emission factors from published references

When site-specific data are unavailable, *good practice* will usually be to develop emission estimates using emission factors drawn from references consistent with the *IPCC Guidelines* and the source category-specific *good practice guidance* described in Chapters 2 to 5. These factors will have been measured under particular circumstances that are judged to be typical. There will be uncertainties associated with the original measurements, as well as with the use of the factors in circumstances other than those associated with the original measurements. It is a key function of *good practice guidance* for each source category to guide the choice of emission factors to minimise this second source of uncertainty to the extent possible. The source category-specific guidance also indicates, wherever possible, the uncertainty ranges likely to be associated with using these factors.

Where such emission factors are used, the associated uncertainties should be estimated from:

- *Original research including country-specific data.* For measurement-based emission factors, the data from the original measurement programme may enable an assessment of the uncertainty and possibly the probability density function. Well-designed measurement programmes will provide sample data that cover the range of types of plants and their maintenance, size and age, so that the factors and their uncertainties can

be used directly. In other cases, expert judgement will be needed to extrapolate from the measurements to the full population of plants in that particular source/sink category.

- *Good practice guidance.* For most emission factors, source category-specific *good practice guidance* provides default uncertainty estimates that should be used in the absence of other information. Unless clear evidence to the contrary is available, the probability density functions are assumed to be normal. However, the inventory agency should evaluate the representativeness of the default for its own situation. If the default is judged to be unrepresentative and the source category is important to the inventory, improved assumptions based upon expert judgement should be developed.

An emission factor that over or underestimates emissions in the base year will probably do so in subsequent years. Therefore, uncertainties due to emission factors will tend to be correlated over time.

6.2.4 Uncertainties associated with activity data

Activity data are often more closely linked to economic activity than are emission factors. There are often well established price incentives and fiscal requirements for accurate accounting of economic activity. Activity data therefore tend to have lower uncertainties and a lower correlation between years. Activity data are often collected and published regularly by national statistical agencies. It is possible that these agencies have already assessed the uncertainties associated with their data as part of their data collection procedures. These uncertainties can be used to construct probability density functions. This information will not necessarily have been published, so it is *good practice* to contact the statistical agencies directly. Since economic activity data are not usually collected for the purpose of estimating greenhouse gas emissions, the applicability of the data should be assessed before using them.

Examples of generic and specific questions that may arise on coverage, representativeness and repeatability from year to year are:

- *Interpretation of statistical differences.* Statistical differences in energy balance usually represent a difference between amount of reported primary fuels and amount of fuels identified in the categories 'final consumption' and 'in transformation'. They can give an indication of the size of the uncertainties of the data, especially where long time series are considered.
- *Interpretation of energy balances.* Production, use and import/export data should be consistent. If not, this may give an indication of the uncertainties.
- *Cross-checks.* It may be possible to compare two types of activity data that apply to the same source to provide an indication of uncertainty ranges. For example, the sum of vehicle fuel use should be commensurate with the sum by vehicle type of the product of vehicle-km times fuel consumption efficiency.
- *Vehicle numbers and types.* Some countries maintain detailed vehicle registration databases with data on vehicles by type, age, fuel type and emission control technology, all of which can be important for a detailed bottom-up inventory of methane (CH_4) and nitrous oxide (N_2O) emissions from such vehicles. Others do not have such detailed information and this will tend to increase the uncertainty.
- *Smuggling of fuel across borders.* This can be significant and may introduce bias into the activity data. Apparent consumption and the sum of the sectoral fuel use may be compared as a cross-check.
- *Biomass fuels.* Where formal markets for these fuels do not exist, consumption estimates may be much less accurate than for fuels in general.
- *Livestock population data.* Accuracy will depend on the extent and reliability of national census and survey methods and there may be different accounting conventions for animals that do not live for a whole year.

Inventory agencies may also undertake dedicated research to collect additional activity data, consistent with *good practice* in the prioritisation of efforts toward *key source categories* (i.e. those source categories with 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, as described in Chapter 7, Methodological Choice and Recalculation).

Probability density functions associated with activity data can be difficult to assess. The procedures outlined in this chapter should be applied to the available information, in accordance with the advice on interpreting expert judgements in the following section.

6.2.5 Expert judgement

Estimates of uncertainty in emission factors or direct emission measurements will need to be based on expert judgement when empirical data are lacking. Estimates of uncertainty in activity data will often be based on expert judgement, informed wherever possible by cross-checks like those described in the previous section.

Experts are people who have special skills or knowledge in a particular field. A judgement is the forming of an estimate or conclusion from information presented to or available to the expert. It is important to select appropriate experts with respect to the emission inventory inputs for which uncertainty estimates are needed.

The goal of expert judgement here is to develop a probability density function, taking into account relevant information such as:

- Is the emission source similar to other sources? How is the uncertainty likely to compare?
- How well is the emission process understood? Have all possible emission sources been identified?
- Are there physical limits on how much the emission factor can vary? Unless the process is reversible it cannot emit less than zero, and this may constrain a very wide uncertainty range. Mass balance considerations or other process data may place an upper limit on emissions.
- Are the emissions consistent with atmospheric concentrations? Emissions are reflected in atmospheric concentrations at site-specific and larger scales and again this may limit the possible emission rates.

A degree of expert judgement is required even when applying classical statistical techniques to data sets, since one must judge whether the data are a representative random sample and, if so, what methods to use to analyse the data. This may require both technical and statistical judgement. Interpretation is especially needed for data sets that are small, highly skewed or censored.[2] The formal methods for obtaining data from experts are known as *expert elicitation*.

POSSIBLE BIASES IN EXPERT ELICITATION

Wherever possible, expert judgement regarding uncertainty should be elicited using an appropriate protocol. Once experts are identified, elicitation protocols should be designed to overcome the biases that can be introduced by the rules of thumb (sometimes called *heuristics*) that experts use when formulating judgements about uncertainty.

The most common unconscious biases introduced by rules of thumb are:

- *Availability bias.* This is basing judgements on outcomes that are more easily remembered.
- *Representativeness bias.* This is basing judgements on limited data and experience without fully considering other relevant evidence.
- *Anchoring and adjustment bias.* This is fixating on a particular value in a range and making insufficient adjustments away from it in constructing an uncertainty estimate.

To counteract the first two potential sources of biases, elicitation protocols should include a review of relevant evidence. In order to counteract the third potential source of bias, it is important to ask the expert to make judgements regarding extreme values first, before asking for judgements regarding central values of a distribution. When an expert gives too narrow a range of values, that is said to be 'overconfidence'. Experts often systematically underestimate uncertainties according to Morgan and Henrion (1990). It is desirable to avoid overconfidence so as not to underestimate the true uncertainty.

There is also the possibility of more conscious biases:

- *Motivational bias* is a desire by an expert to influence an outcome or to avoid contradicting prior positions on an issue.
- *Expert bias* arises from an unqualified expert's desire to appear as a true expert in the field. This would typically lead to overconfident estimates of uncertainty.

[2] In these cases it may be helpful to consider a numerical method, such as the bootstrap, for characterising sampling distributions. Methods for characterising sampling distributions for the mean are described by Cullen and Frey (1999), Frey and Rhodes (1996), and Frey and Burmaster (1999).

- *Managerial bias* is a situation in which an expert makes judgements that achieve organisational goals, rather than judgements that reflect the actual state of knowledge regarding an inventory input.
- *Selection bias* occurs when the inventory agency selects the expert who tells it what it wants to hear.

The best way to avoid these biases is to be careful in the selection of experts.

Expert judgements can be elicited from individuals or groups. Groups can be useful for sharing knowledge and hence could be part of the motivation, structuring, and conditioning steps of the elicitation. However, group dynamics may introduce other biases. Thus, it is usually preferable to elicit judgement on an individual basis.

A PROTOCOL FOR EXPERT ELICITATION

An example of a well-known protocol for expert elicitation is the Stanford/SRI protocol. Its five steps are described below, and an example of its use is presented in Box 6.1, A Brief Example of Detailed Expert Judgement.

- *Motivating*: Establish a rapport with the expert, and describe the context of the elicitation. Explain the elicitation method to be used and the reason it was designed that way. The elicitor should also try to explain the most commonly occurring biases to the expert, and to identify possible biases in the expert.
- *Structuring*: Clearly define the quantities for which judgements are to be sought, including, for example, the year and country, the emission source category, the averaging time to be used (one year), the focus on uncertainty in the mean value of emission factors, and the structure of the emission inventory model. Clearly identify conditioning factors and assumptions (e.g. emissions should be for typical conditions averaged over a one-year period).
- *Conditioning*: Work with the expert to identify all relevant data, models, and theory pertaining to the quantity for which judgements about uncertainty are required.
- *Encoding*: Request the expert's judgement regarding uncertainty. The next section on encoding describes some alternative methods to use.
- *Verification*: Analyse the expert's response and provide the expert with feedback as to what has been concluded regarding his or her judgement. Is what has been encoded really what the expert meant? Are there inconsistencies in the expert's judgement?

METHODS FOR ENCODING EXPERT JUDGEMENTS

The method to be used in the encoding step should depend upon the expert's familiarity with probability distributions. Some commonly used methods are:

- *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*: This method 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 this would be repeated for the other end of the distribution. 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 because some experts are overconfident about their knowledge of probability distributions.

BOX 6.1

A BRIEF EXAMPLE OF DETAILED EXPERT JUDGEMENT

Suppose that the inventory agency has identified an expert regarding emissions of methane from power plants, and wish to obtain her judgement regarding the uncertainty in annual average emissions for this source category. As part of the motivation step, the elicitor has explained to the expert the general purpose of the analysis and the expert elicitation protocol to be used. In the structuring step, the elicitor works with the expert to set up the specific elicitation protocol. For example, although all the inventory agency may want is an annual average uncertainty estimate, the expert may tell the elicitor that she prefers to provide judgements separately for start-up, part load, and full load operation of the plant, and that these three judgements should be weighted in order to come up with the combined uncertainty for an annual average. After structuring the problem, the elicitor then reviews the expert information relevant to the assessment, such as measurements that may have been made on similar types of power plants or other combustion sources. In the elicitation step, the elicitor might ask the expert for an upper value such that there is only a one in 40 chance (2.5 percent probability) of obtaining a higher value. After getting the value, the elicitor asks the expert to explain the logical basis for this estimate, such as the scenario of operation at the plant that might lead to such a high emission rate. Then the process might be repeated for the lower end of the range, and perhaps for the median, 25^{th} percentile, and 75^{th} percentile. A mixture of fixed value and fixed probability questions might be used. The elicitor should plot these on a graph so that any inconsistencies can be identified and corrected during the time available with the expert. In the verification step, the elicitor would make sure that the expert is comfortable that their judgement has been well represented. The elicitor might also see how the expert would react to the possibility of values outside of the interval for which judgements were provided, so as to ensure that the expert is not being overconfident.

Sometimes the only available expert judgement will consist of a range, perhaps quoted together with a most likely value. Under these circumstances the following rules apply:

- Where experts only provide an upper and a lower limiting value, assume the probability density function is uniform and that the range corresponds to the 95% confidence interval.
- Where experts also provide a most likely value, assume a triangular probability density function using the most likely values as the mode and assuming that the upper and lower limiting values each exclude 2.5% of the population. The distribution need not be symmetrical.

Some other sources of information on expert elicitation include Spetzler and von Holstein (1975), Morgan and Henrion (1990), Merkhofer (1987), Hora and Iman (1989), and NCRP (1996).

The subjective nature of expert judgement increases the need for quality assurance and quality control procedures to improve comparability of uncertainty estimates between countries. Therefore expert judgements should be documented as part of the national archiving process, and inventory agencies are encouraged to review expert judgements, particularly for *key source categories*. Documentation should include:

- Reference number for judgement;
- Date;
- Person(s) involved and affiliation;
- The quantity being judged;
- The logical basis for judgement, including any data taken into consideration;
- The resultant probability distribution, or the range and most likely value and the probability distribution subsequently inferred;
- Identification of any external reviewers;
- Results of any external review;
- Approval by inventory agency, specifying date and person.

6.3 METHODS FOR COMBINING UNCERTAINTIES

Once the uncertainties in the source categories have been determined, they may be combined to provide uncertainty estimates for the entire inventory in any year and the uncertainty in the overall inventory trend over time.

The error propagation equation, as discussed more extensively in Annex 1 of this report, and in Annex I of the *IPCC Guidelines* (Reporting Instructions), yields two convenient rules for combining uncorrelated uncertainties under addition and multiplication:

- *Rule A:* Where uncertain quantities are to be combined by addition, 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 can be derived for the uncertainty of the sum, that when expressed in percentage terms becomes:

EQUATION 6.3

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

  Where:

  U_{total} is the percentage uncertainty in the sum of the quantities (half the 95% confidence interval divided by the total (i.e. mean) and expressed as a percentage);

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

- *Rule B*: Where uncertain quantities are to be combined by multiplication, the same rule applies except that the standard deviations must all be expressed as fractions of the appropriate mean values (this rule is approximate for all random variables).

  A simple equation can also be derived for the uncertainty of the product, expressed in percentage terms:

EQUATION 6.4

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

  Where:

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

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

The greenhouse gas inventory is principally the sum of products of emission factors and activity data. Therefore, Rules A and B can be used repeatedly to estimate the uncertainty of the total inventory. In practice, uncertainties found in inventory source categories vary from a few percent to orders of magnitude, and may be correlated. This is not consistent with the assumptions of Rules A and B that the variables are uncorrelated with a standard deviation of less than about 30% of the mean, but under these circumstances, Rules A and B may still be used to obtain an approximate result. Alternatively, a stochastic simulation (the Monte Carlo method) can be used, that can combine uncertainties with any probability distribution, range, and correlation structure, provided they have been suitably quantified. Thus, two tiers for uncertainty analysis are described below:

- *Tier 1*: Estimation of uncertainties by source category using the error propagation equation via Rules A and B, and simple combination of uncertainties by source category to estimate overall uncertainty for one year and the uncertainty in the trend.
- *Tier 2*: Estimation of uncertainties by source category using Monte Carlo analysis, followed by the use of Monte Carlo techniques to estimate overall uncertainty for one year and the uncertainty in the trend.

Monte Carlo analysis can also be used in a restricted way within Tier 1 to combine activity data and emission factor uncertainties that have very wide or non-normal probability distributions or both. This approach can also help deal with source categories within Tier 1 that are estimated by process models, rather than by the classical 'emission factor times activity data' calculation. The choice between methods is discussed in Section 6.3.1 below.

6.3.1 Comparison between tiers and choice of method

Use of either Tier 1 or Tier 2 will provide far greater insight than previously available into how individual source categories and greenhouse gases contribute to uncertainty in total emissions in any year, and to the trend in total emissions between years.

Application of Tier 2 to the UK inventory (Eggleston *et al.*, 1998) suggests that the 95% confidence interval is asymmetric and lies between about 7% below and 20% above the mean. Application of Tier 1 (see Appendix 6A.2, Tier 1 Uncertainty Calculation Example) suggests an uncertainty of about ±20%. Since the approximations inherent in Tier 1 mean that it cannot deal with asymmetry, this comparison is encouraging. Physically, the reason for the asymmetry identified under Tier 2 is that the uncertainty range of some very uncertain source categories is constrained by the knowledge that the emissions cannot be less than zero. The Tier 2 method can make use of this extra knowledge, but the Tier 1 method cannot. On trends between years, the Tier 2 study by Eggleston *et al.* suggests that the 95% confidence interval is roughly symmetrical and lies between 5% above and 5% below the mean.[3] The corresponding Tier 1 result gives a range of about ±2%. The lower Tier 1 value is partly because the trend estimated here is for the period 1990 to 1997, whereas the Tier 2 estimate was for 1990 to 2010, but this is unlikely to account for all the differences. Nevertheless, both methods still give similar magnitudes in the trend uncertainty that is less than the uncertainty in total emissions in any year.

Further national comparisons between methods would be very useful in developing understanding. The Tier 1 method, being spreadsheet based, is very easy to apply and would represent hardly any additional effort for an inventory agency also undertaking Tier 2. Therefore, for the present, it is *good practice* for all countries undertaking uncertainty analysis to report Tier 1 results, and for all inventory agencies with sufficient resources and expertise to undertake Tier 2.

6.3.2 Tier 1 – Estimating uncertainties by source category with simplifying assumptions

The Tier 1 analysis estimates uncertainties by using the error propagation equation in two steps. First, the Rule B approximation is used to combine emission factor and activity data ranges by source category and greenhouse gas. Second, the Rule A approximation is used to arrive at the overall uncertainty in national emissions and the trend in national emissions between the base year and the current year.

The Tier 1 approach should be implemented using Table 6.1, Tier 1 Uncertainty Calculation and Reporting that can be set up on commercial spreadsheet software. The table is completed at the source category level using uncertainty ranges for activity data and emission factors consistent with the sectoral *good practice guidance* in Chapters 2 to 5. Different gases should be entered separately as CO_2 equivalents (i.e. the emissions should be multiplied by 100-year GWP values). Trend uncertainties are estimated using two sensitivities:

- *Type A sensitivity*: the change in the difference in overall emissions between the base year and the current year, expressed as a percentage, resulting from a 1% increase in emissions of a given source category and gas in both the base year and the current year.
- *Type B sensitivity*: the change in the difference in overall emissions between the base year and the current year, expressed as a percentage, resulting from a 1% increase in emissions of a given source category and gas in the current year only.

Conceptually, Type A sensitivity arises from uncertainties that affect emissions in the base year and the current year equally, and Type B sensitivity arises from uncertainties that affect emissions in the current year only. Uncertainties that are fully correlated between years will be associated with Type A sensitivities, and

[3] Specifically a fall in emissions of 6 ±5%.

uncertainties that are not correlated between years will be associated with Type B sensitivities. The discussion in Sections 6.2.1 to 6.2.4 above suggests that emission factor uncertainties will tend to have Type A sensitivities, and activity data uncertainties will tend to have Type B. However, this association will not always hold and it is possible to apply Type A sensitivities to activity data, and Type B sensitivities to emission factors to reflect particular national circumstances. Type A and Type B sensitivities are simplifications introduced for the analysis of correlation.

Once the uncertainties introduced into national emissions by Type A and Type B sensitivities have been calculated, they can be summed using the error propagation equation (Rule A) to give the overall uncertainty in the trend.

The columns of Table 6.1, Tier 1 Uncertainty Calculation and Reporting, are labelled A to Q and contain the following information:

- A and B show the IPCC source category and greenhouse gas.
- C and D are the inventory estimates in the base year and the current year[4] respectively, for the source category and gas specified in columns A and B, expressed in CO_2 equivalents.
- E and F contain the uncertainties for the activity data and emission factors respectively, derived from a mixture of empirical data and expert judgement as previously described in this chapter, entered as half the 95% confidence interval divided by the mean and expressed as a percentage. The reason for halving the 95% confidence interval is that the value entered in columns E and F then corresponds to the familiar plus or minus value when uncertainties are loosely quoted as 'plus or minus x%', so expert judgements of this type can be directly entered in the spreadsheet. If uncertainty is known to be highly asymmetrical, enter the larger percentage difference between the mean and the confidence limit.
- G is the combined uncertainty by source category derived from the data in columns E and F using the error propagation equation (Rule B). The entry in column G is therefore the square root of the sum of the squares of the entries in columns E and F.
- H shows the uncertainty in column G as a percentage of total national emissions in the current year. This is a measure of the degree of uncertainty introduced into the national emissions total by the source category in question. The entry in each row of column H is the entry in column G multiplied by the entry in column D, divided by the total at the foot of column D. The total at the foot of column H is an estimate of the percentage uncertainty in total national emissions in the current year, calculated from the entries above using Rule A. This total is obtained by summing the squares of all the entries in column H and taking the square root.
- I shows how the percentage difference in emissions between the base year and the current year changes in response to a one percent increase in source category emissions in both the base year and the current year. This shows the sensitivity of the trend in emissions to a systematic uncertainty in the emissions estimate (i.e. one that is correlated between the base year and the current year). This is the Type A sensitivity as defined above. Appendix 6A.1 provides the derivation for the formula for the entries in column I.
- J shows how the percentage difference in emissions between the base year and the current year changes in response to a one percent increase in source category emissions in the current year only. This shows the sensitivity of the trend in emissions to random error in the emissions estimate (i.e. one, that is not correlated, between the base year and the current year). This is the Type B sensitivity as described above. The formula for the entries in column J is derived in Appendix 6A.
- K uses the information in columns I and F to show the uncertainty introduced into the trend in emissions by emission factor uncertainty, under the assumption that uncertainty in emission factors is correlated between years. If the user decides that the emission factor uncertainties are not correlated between years then the entry in column J should be used in place of that in column I and the result multiplied by $\sqrt{2}$. The formula for the entries in column K is derived in Appendix 6A.
- L uses the information in columns J and E to show the uncertainty introduced into the trend in emissions by activity data uncertainty, under the assumption that uncertainty in activity data is not correlated between years. If the user decides that the activity data uncertainties are correlated between years then the entry in column I should be used in place of that in column J and the $\sqrt{2}$ factor does not then apply. The formula for the entries in column L is derived in Appendix 6A.

[4] The current year is the most recent year for which inventory data are available.

- M is an estimate of the uncertainty introduced into the trend in national emissions by the source category in question. Under Tier 1, this is derived from the data in columns K and L using Rule B. The entry in column M is therefore the square root of the sum of the squares of the entries in columns K and L. The total at the foot of this column is an estimate of the total uncertainty in the trend, calculated from the entries above using the error propagation equation. This total is obtained by summing the squares of all the entries in column M and taking the square root. The formula for the entries in column M and the total at the foot of column M is shown in Appendix 6A.1.
- Columns N to Q are used for indicators and cross referencing to footnotes.
- N contains D, M or R, depending on whether the emission factor uncertainty range is based on default (D) information in source category guidance, measurements (M) made for the purpose or national referenced (R) information.
- O contains D, M or R, depending on whether the activity data uncertainty range is based on default information in sector guidance, measurements made for the purpose or national referenced information.
- P contains the reference numbers of any expert judgements used to estimate uncertainties in this source category.
- Q contains the number of an explanatory footnote to go at bottom of table to identify documentary reference of uncertainty data (including measured data) or other comments relevant to the line.

An example of the spreadsheet with all the numerical data completed is provided in Appendix 6A.2, Tier 1 uncertainty calculation example.

TABLE 6.1

TIER 1 UNCERTAINTY CALCULATION AND REPORTING

A	B	C	D	E	F	G	H	I	J	K	L	M
IPCC Source category	Gas	Base year emissions	Year t emissions	Activity data uncertainty	Emission factor uncertainty	Combined uncertainty	Combined uncertainty as % of total national emissions in year t	Type A sensitivity	Type B sensitivity	Uncertainty in trend in national emissions introduced by emission factor uncertainty	Uncertainty in trend in national emissions introduced by activity data uncertainty	Uncertainty introduced into the trend in total national emissions
		Input data	Input data	Input data	Input data	$\sqrt{E^2 + F^2}$	$\frac{G \bullet D}{\sum D}$	Note B	$\frac{D}{\sum C}$	$I \bullet F$ Note C	$J \bullet E \bullet \sqrt{2}$ Note D	$\sqrt{K^2 + L^2}$
		Gg CO_2 equivalent	Gg CO_2 equivalent	%	%	%	%	%	%	%	%	%
E.g. 1.A.1. Energy Industries Fuel 1	CO_2											
E.g. 1.A.1. Energy Industries Fuel 2	CO_2											
Etc...	…											
		$\sum C$	$\sum D$				$\sqrt{\sum H^2}$					$\sqrt{\sum M^2}$
Total												

TABLE 6.1 (CONTINUED)

TIER 1 UNCERTAINTY CALCULATION AND REPORTING

A (continued)	B (continued)	N	O	P	Q
IPCC Source category	Gas	Emission factor quality indicator	Activity data quality indicator	Expert judgement reference numbers	Footnote reference number
		Note E	Note E		
E.g. 1.A.1. Energy Industries Fuel 1	CO_2				
E.g. 1.A.1. Energy Industries Fuel 2	CO_2				
Etc...	…				
Total					

Note A If only total uncertainty is known for a source category (not for emission factor and activity data separately), then:

- If uncertainty is correlated across years, enter the uncertainty into column F, and enter 0 in column E;
- If uncertainty is not correlated across years, enter the uncertainty into column E and enter 0 in column F.

Note B

$$\frac{0.01 \bullet D_x + \sum D_i - (0.01 \bullet C_x + \sum C_i)}{(0.01 \bullet C_x + \sum C_i)} \bullet 100 - \frac{\sum D_i - \sum C_i}{\sum C_i} \bullet 100$$

Note C

In the case where no correlation between emission factors is assumed, sensitivity B should be used and the result multiplied by √2:

$$K_x = J_x \bullet F_x \bullet \sqrt{2}$$

Note D

In the case where correlation between activity data is assumed, sensitivity A should be used and the √2 is not required:

$$L_x = I_x \bullet E_x$$

Note E

Please use the following abbreviations:
D – IPCC source category default
M – measurement based
R – national referenced data

6.3.3 Tier 1 aggregation and reporting

Table 6.1, Tier 1 Uncertainty Calculation and Reporting, has one line for each source category, fuel (where appropriate), and greenhouse gas and should be used for reporting.

Although the Tier 1 method allows for correlation over time, as described above, it does not account for correlation and dependency between source categories that may occur because the same activity data or emission factors may be used for multiple estimates. Often one gas dominates the source category and this reduces the effect of any correlation. However, correlation and dependency may be significant for fossil fuels because a given fuel is used with the same emission factor across several sub-categories, and if (as is sometimes the case) total consumption of a fuel is better known than consumption disaggregated by source category, hidden dependencies will exist within the statistics because of the constraint provided by overall consumption. Dependency and correlation can be addressed by aggregating the source categories to the level of overall consumption of individual fuels before the uncertainties are combined. This entails some loss of detail in reporting on uncertainties but will deal with the dependencies where they are thought to be significant (e.g. where the uncertainties in fossil fuel emissions when aggregated from the source category level are greater than expected). The example Tier 1 calculation using UK data provided in Appendix 6A.2 has fossil fuel categories aggregated in this way. This has the advantage of allowing compatibility with the categories suggested in Chapter 7 for the *key source category* analysis.

6.4 TIER 2 – ESTIMATING UNCERTAINTIES BY SOURCE CATEGORY USING MONTE CARLO ANALYSIS

Under Tier 2, the simplifying assumptions required for Tier 1 can be relaxed. Tier 2 uses Monte Carlo analysis to combine source category uncertainties.

The principle of Monte Carlo analysis is to select random values of emission factor and activity data 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. Monte Carlo analysis can be performed at the source category level, for aggregations of source categories or for the inventory as a whole.

Monte Carlo analysis can deal with probability density functions of any physically possible shape and width, can handle varying degrees of correlation (both in time and between source categories) and can deal with more

complex models (e.g. the 1st order decay for CH_4 from landfills) as well as simple 'emission factor times activity data' calculations.

Eggleston *et al*. (1998) provide an example of Monte Carlo analysis, applied to a national greenhouse gas inventory and used to estimate uncertainties both in overall emissions and emissions trends. Another example of the use of Monte Carlo analysis is given in McCann *et al*. (1994). A general description of the Monte Carlo method can be found in Fishman (1996).

Like all methods, Monte Carlo analysis only provides satisfactory results if it is properly implemented. This requires the analyst to have scientific and technical understanding of the inventory. Of course, the results will only be valid to the extent that the input data, including any expert judgements, are sound.

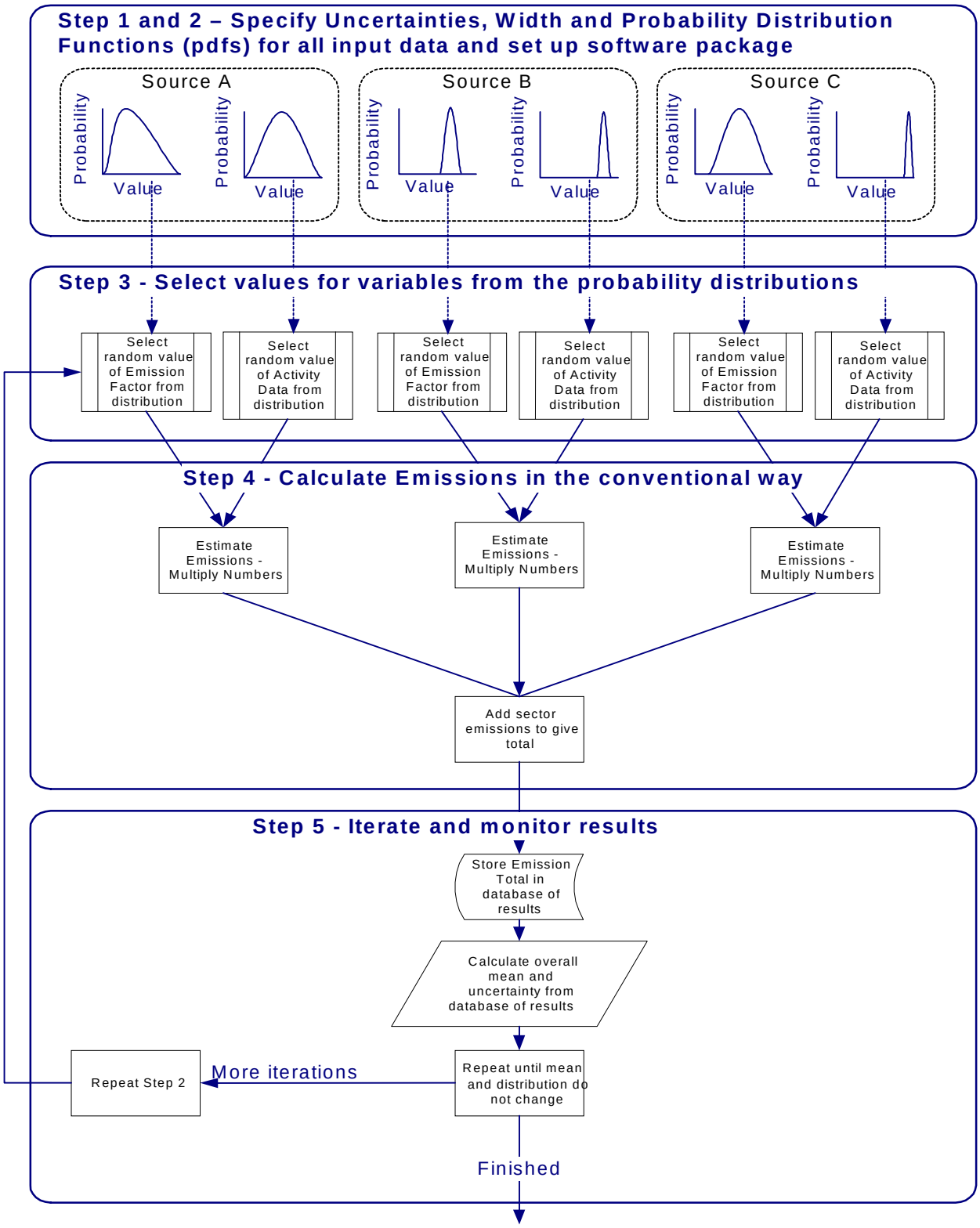
The Monte Carlo approach consists of five clearly defined steps shown in Figure 6.1. Only the first two of these require effort from the user, the remainder being handled by the software package. Section 6.5.3 contains a short discussion of various software packages.

- **Step 1 – Specify source category uncertainties.** Specify the uncertainties in the basic data. This includes emission factors and activity data, their associated means and probability distribution functions, and any cross correlation between source categories. Take account of the information in Sections 6.2.1 to 6.2.5.
- **Step 2 – Set up software package.** The emission inventory calculation, the probability density functions and the correlation values should be set up in the Monte Carlo package.

The software automatically performs the subsequent steps.

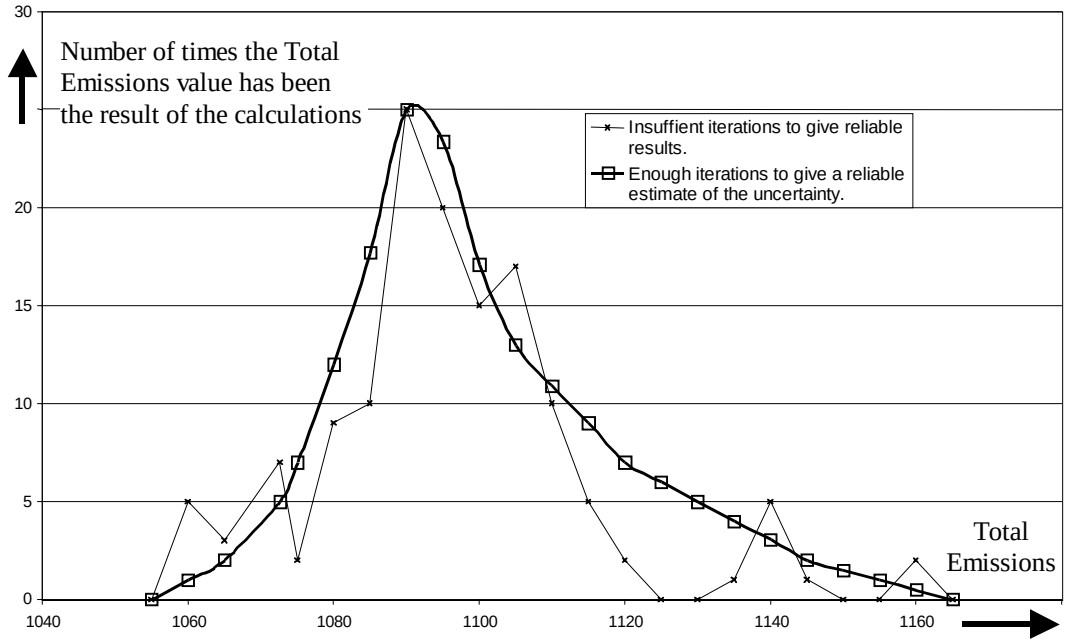
- **Step 3 – Select random variables.** This is the start of the iterations. For each input data item, emission factor or activity data, a number is randomly selected from the probability density function of that variable.
- **Step 4 – Estimate emissions.** The variables selected in Step 3 are used to estimate total emissions. The example given in Figure 6.1 assumes three source categories, each estimated as activity multiplied by an emission factor, and then summed to give total emissions. The calculations can be more complex. Emissions by gas can be multiplied by GWP values, in order to obtain total national emissions in CO_2 equivalent. Correlations of 100% are easy to incorporate, and good Monte Carlo packages allow other correlations to be included. Since the emission calculations should be the same as those used to estimate the national inventory, the Monte Carlo process could be fully integrated into the annual emission estimates.
- **Step 5 – Iterate and monitor results.** The calculated total from step 4 is stored, and the process then repeats from step 3. The mean of the totals stored gives an estimate of the total emission. Their distribution gives an estimate of the probability density function of the result. As the process repeats, the mean approaches the final answer. When the mean no longer changes by more than a predefined amount, the calculation can be terminated. When the estimate for the 95% confidence range is determined to within ± 1%, then an adequately stable result has been found. Convergence can be checked by plotting a frequency plot of the estimates of the emission. This plot should be reasonably smooth (Figure 6.2, Example frequency plots of the results of a Monte Carlo simulation). These actions should be handled by the software, with the user specifying either a number of iterations or convergence criteria.

Figure 6.1 Illustration of Monte Carlo Method



This example assumes three emission source categories each where the emission is calculated as Activity Data • Emission Factor.

Figure 6.2 Example Frequency Plots of the Results of a Monte Carlo Simulation



6.4.1 Tier 2 uncertainties in trends

The Tier 2 Monte Carlo method can be used to estimate uncertainties in the trend as well as in the absolute emission value in a given year. The procedure is a simple extension of that described in the previous section.

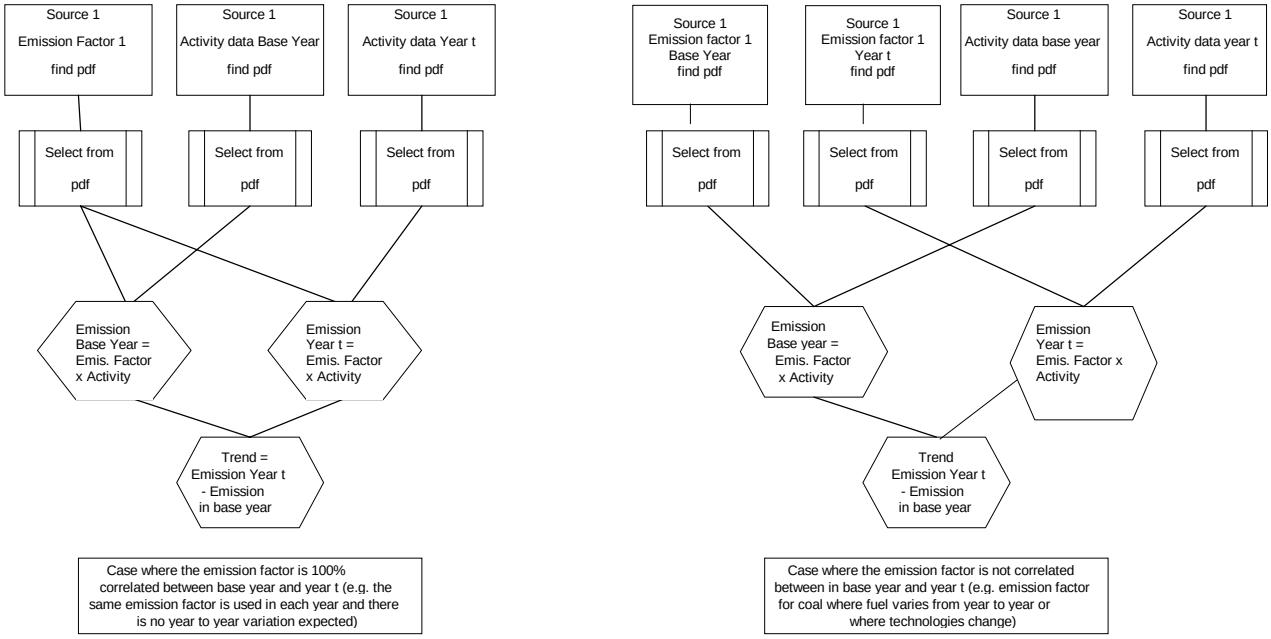
The trend is defined here as the difference between the base year and the year of interest (year t). Therefore, the Monte Carlo analysis needs to be set up to estimate both years simultaneously. The procedure is:

- **Step 1 – Specify source category uncertainties.** Determine the probability density functions for each emission factor and activity. This is the same process as described above except that it needs to be done for both the base year and the current year, and relationships between the data need to be considered. For many source categories, the same emission factor will be used for each year (i.e. the emission factors for both years are 100% correlated). In these cases, one distribution is described and the value selected from it is used for each year in step 3. Changes in the technologies or practices will alter the emission factor over time. In this case, two emission factors should be used, that have a lower or zero correlation. If the emission factors contain a random element or vary unpredictably from year to year, then separate emission factors should also be used (e.g. with fossil fuel carbon content that can change according to the market supply of the fuel and also contains its own uncertainty). Generally, activity rates are assumed to be uncorrelated between years, and so two distributions should be input, even if their parameters are the same, so that two different random selections from these distributions will be generated in step 3. The computer package used may well enable other correlations to be set up and these capabilities could be used if sufficient information is available. However, this will probably be necessary in only a few cases.
- **Step 2 – Set up software package.** The software package should be set up as previously described, except that the probability distribution functions will need to capture the relationship between emissions in two years, and for trend calculations there need to be two separate but simultaneous calculations of the emissions in the base year and year t. In cases where the input data is assumed to be 100% correlated (mainly some emission factors) care needs to be taken that the same random number selected from the probability distribution function is used in estimating both years. A final calculation is then needed to find the difference between the two years.

Subsequent steps are, in most cases, performed automatically by the software package.

- **Step 3 – Select random variables**. The computer program will proceed as previously described, taking account of any correlation between probability density functions (PDF). Figure 6.3, below, shows the calculation scheme for trend analysis
- **Step 4 – Estimate Emissions.** As in the previous description, the variables selected in Step 3 will be used to estimate the total emissions.
- **Step 5 – Results.** The emissions total calculated in step 4 is stored in a data file. The process then repeats from step 3 until there is adequate convergence of the results. Considerations for this are the same as described above. A range of results is estimated at the same time including total and sectoral emissions for the base year, total and sectoral emissions for year *t*, and the differences (trends) between these for the total and any sectors of interest.

Figure 6.3 Calculation Scheme for Monte Carlo Analysis of the Absolute Emissions and the Trend of a Single Source Category, Estimated as Emission Factor Times an Activity Rate



6.4.2 Reporting Tier 2 uncertainty analysis

The following data format is suitable for reporting the results of the Monte Carlo simulation for emissions by source category, by fuel (where appropriate) and by greenhouse gas expressed as CO_2 equivalent. In Table 6.2, the overall uncertainty in the national emissions trend appears at the foot of columns I and J. Inventory agencies performing a Tier 2 analysis should also report the results of a Tier 1 analysis using Table 6.1, as set out in Section 6.3.1, Comparison between tiers and choice of method.

TABLE 6.2

TIER 2 UNCERTAINTY REPORTING

A	B	C	D	E	F	G	H	I	J
IPCC Source category	Gas	Base year emissions	Year t emissions	Uncertainty in year t emissions as % of emissions in the category		Uncertainty introduced on national total in year t	% change in emissions between year t and base year	Range of likely % change between year t and base year	
		(Gg CO_2 equivalent)	(Gg CO_2 equivalent)	% below (2.5 percentile)	% above (97.5 percentile)	(%)	(%)	Lower % (2.5 percentile)	Upper % (97.5 percentile)
e.g. 1.A.1 Energy Industries Fuel 1	CO_2								
e.g. 1.A.2 Energy Industries Fuel 2	CO_2								
Etc…	…								
Total									

6.5 PRACTICAL CONSIDERATIONS IN USING MONTE CARLO SIMULATION

Monte Carlo simulation requires that the analyst specify probability distributions for each model input for which the uncertainty is to be quantified. The assumption is that the simulation is a reasonable representation of the real world. The probability distributions may be obtained by a variety of methods, including statistical analysis of data, or the elicitation of expert judgement. A key consideration is to develop the distributions for the inputs so that they are all based upon the same underlying assumptions regarding averaging time, location, and other conditioning factors relevant to the particular assessment (e.g. climatological conditions influencing agricultural greenhouse gas emissions). For this reason, one should not assume that an uncertainty distribution from another country is directly applicable as an input to an inventory.

6.5.1 Specifying probability distributions for inventory inputs

Monte Carlo simulation requires identification of the model inputs for which probability distributions are to be assigned, and development of the corresponding probability distributions. Methods for developing distributions based upon elicitation of expert judgement have already been reviewed in this chapter. Methods for developing distributions based upon statistical analysis of data are described and illustrated by Cullen and Frey (1999). Other useful references include Hahn and Shapiro (1967), Ang and Tang (1975), D'Agostino and Stephens (1986), Morgan and Henrion (1990), and USEPA (1996, 1997, 1999). Some examples of probabilistic analyses applied to emission inventories are given by Frey *et al.* (1998) and Frey *et al.* (1999).

In order to use data as a basis for developing distributions, the first critical step is to determine if the data are a random, representative sample, in the case of a sample from a population. Some key questions to ask regarding the data include:

- Are the data representative of the variety of conditions pertaining to the emission or activity factors specific to national circumstances?
- Are the data a random sample?
- What is the averaging time associated with the data set, and is it the same as for the assessment (which will be for annual emissions in a given year)?

If the data are a random, representative sample, then the distribution can be established directly using classical statistical techniques, even if the sample size is small. It may be necessary to convert data using an appropriate averaging time. General advice on choosing probability density functions is provided in Annex 1, Conceptual Basis for Uncertainty Analysis, Section 2.5, Good Practice Guidance for Selecting a Probability Density Function.

In the ideal case, available data will represent an annual average for an emission factor or an annual total for activity data. In this case, the data would represent a single sample from a population distribution of annual average values. The estimated standard deviation of the population would be an appropriate measure of uncertainty in annual emissions. In other cases, the data may represent an exhaustive census of the sum of all activity (e.g. total energy use for a particular fuel). In this case, information regarding errors in the measurements or survey instruments would form a basis for assessing uncertainty. The range of uncertainty of activity data might be bounded using independent methods or consistency checks. For example, fuel consumption data can be compared with estimates of production, including estimates of production via different methods.

In the case of a population sample, the most critical aspect to evaluate is whether the data are random and representative of the population. If these conditions are met, classical statistical methods can be used to define the distribution. If not, then some combination of data analysis and expert elicitation of distributions will be required. In the former case, Cullen and Frey (1999) suggest exploration of the data set using summary statistics and graphics to evaluate essential features (e.g. central tendency, range of variation, skewness). The insights obtained by examining the data, combined with knowledge of the processes that generated the data, should be considered when selecting a mathematical or numerical representation of the distribution for input into the Monte Carlo simulation.

Once a particular distribution is selected as a candidate for fitting to the data set, techniques such as 'maximum likelihood estimation[5]' or the 'method of matching moments[6]' can be used to estimate the parameters of the distribution. The goodness-of-fit of the distribution can be evaluated in numerous ways, including comparison of the fitted cumulative distribution function (CDF) with the original data set, probability plots, and goodness-of-fit tests (e.g. Cullen and Frey, 1999). It is important that the selection of a parametric distribution to represent a data set should be based not solely upon goodness-of-fit tests, but upon similarities in processes that generated the data versus the theoretical basis for a distribution (e.g. Hahn and Shapiro, 1967).

If the data are averaged over less than one year, it may be necessary to extrapolate the uncertainty over the year. Consider an example in which the data set represents variability in daily average emissions measurements for a particular source category. One approach, described in detail by Frey and Rhodes (1996), is to fit a parametric distribution to the data set for daily variability, use a numerical technique known as bootstrap simulation to estimate uncertainty in the parameters of the distribution, and use Monte Carlo simulation to simulate randomised annual averages of the emission factor. Using bootstrap simulation, the uncertainty in the sampling distribution for the parameters for the fitted distribution can be simulated (e.g. Efron and Tibshirani, 1993; Frey and Rhodes, 1996; Frey and Burmaster, 1999).

A simple form of bootstrap simulation works as follows: from the fitted distribution, a random synthetic data set of the same sample size as the original dataset is simulated using Monte Carlo simulation. The synthetic data set is referred to as a *bootstrap sample*. For the bootstrap sample, any statistic or parameter can be calculated, such as a mean or parameters of a new distribution fitted to the synthetic data set. A statistic or parameter estimated from a bootstrap sample is referred to as a *bootstrap replicate* of that statistic or parameter. This process is then repeated many times (typically 500 to 1,000), generating a corresponding number of bootstrap samples and duplicated statistics. The statistics will take on different values each time because the bootstrap samples are randomised versions patterned after the original data set. Thus, this method is a numerical technique for estimating sampling distributions for any statistic for any type of distribution for which the statistic exists. In essence, bootstrap simulation is a numerical technique for simulating random sampling error. The 500 to 1,000 bootstrap samples imply a corresponding number of alternative plausible distributions from which the original data set could have been a random sample. For these alternative distributions, each of which reflects daily variability in the example, one can simulate a year's worth of emissions estimates (i.e. 365 random samples of daily emissions summed to give an annual total or averaged to give an annual average emission factor), thereby yielding 500 to 1,000 estimates of annual mean or annual total emissions. The distribution of these estimates will describe uncertainty in the annual case based upon random sampling error. A key assumption in this example is that there is no autocorrelation among the daily values, and that the daily values are representative of annual conditions – that there are for example no seasonal effects that the sample fails to capture.

6.5.2 How much effort is needed to characterise uncertainty in an inventory input?

Ideally, the amount of effort devoted to characterising uncertainty in an inventory input should be proportional to its importance to the overall uncertainty assessment. It would not be a good use of limited resources to spend large amounts of time exhaustively collecting data and expert judgements for a source category that has little effect on overall uncertainty. Similarly, it would be a shortcoming of an assessment not to devote reasonable resources to quantifying uncertainty in inputs to which the overall uncertainty in the inventory is highly sensitive. Thus, many analysts who perform probabilistic simulation suggest an iterative approach to performing the simulation. In the first iteration of an uncertainty analysis, preliminary assessments of uncertainty of inputs may be made and propagated through the inventory solely for the purpose of making a preliminary identification of what the main sources of uncertainty are. Methods for assessing the importance of each input are described in references such as Morgan and Henrion (1990), Frey and Cullen (1999), and others. An example of one relatively simple technique is to calculate the correlation coefficient between the simulated numerical values of the inventory output distribution and the numerical values simulated for each input distribution. This correlation reflects the strength of the linear relationship between the two. The larger the magnitude of the correlation

[5] The method of maximum likelihood selects as estimates the values of the parameters that maximise the likelihood (the joint probability function or joint density function) of the observed sample.

[6] The method of moments finds estimators of unknown parameters by equating corresponding sample and population moments. The method is easy to employ and provides consistent estimators. In many cases the method of moments estimators are biased (Wackerly, Mendenhall III and Scheaffer, 1996; pp. 395-397).

coefficient, the stronger is the relationship between uncertainty in the input and uncertainty in the output, indicating that the input should be considered as a 'sensitive' one. Many software packages will perform the sensitivity calculation automatically for the user and present them graphically.

After identifying the sensitive inputs, efforts can then be directed as appropriate to improve the estimate of uncertainty for only those inputs. Then, the final uncertainty analysis can be performed with greater confidence on the basis that important inputs have received proportionally greater attention than insensitive inputs.

Another point regarding iteration pertains to the longer-term aspects of performing uncertainty analyses. It can be daunting to set up a Monte Carlo simulation for the first time. However, as the inventory agency gains experience with these analyses, the agency will likely find it easier to improve the analysis in the future. Monte Carlo analysis is typically a learning process for all involved, because it motivates critical and important questions regarding the basis for and quality of an emission inventory. Thus, over time, Monte Carlo simulation will help in deciding where to focus data collection activities that will result in improved confidence regarding the inventory.

6.5.3 Choosing a simulation technique and a simulation sample size

There are several commercially available software tools that can be used to perform Monte Carlo simulation. Examples of these include Crystal Ball, @Risk, Analytica, and Mathematica. The first two are add-ins for commonly used spreadsheet programs. Many software tools offer an option of different sampling methods, including random Monte Carlo simulation and variations of Latin Hypercube Sampling (LHS). LHS can produce 'smoother' looking model output distributions for sample sizes of only a few hundred samples. The disadvantage of using LHS is that one must decide ahead of time how many iterations to use. This is because two or more LHS simulations cannot be combined since they will use overlapping strata, leading to difficulties in interpreting results. In some cases, LHS can yield underestimates of the higher moments of probability distributions, since the stratification method also can preclude clustering of very high or low values as can occur in random data sets. The overall suggestion is to use random Monte Carlo simulation as the default method, because it will give flexibility to continue a random simulation to larger and larger simulation sample sizes if necessary until the model output distribution converges. Cullen and Frey (1999) provide more information on the comparison of LHS and Monte Carlo simulation (pp. 207-213).

6.5.4 Dependence and correlation among inventory inputs

A key issue often considered by analysts when setting up a probabilistic analysis is whether there are dependencies or correlations among model inputs. Ideally, it is preferable to define the model so that the inputs are as statistically independent of each other as possible. Rather than to try to estimate activity data for many sub-categories for which data are derived at least in part by differences, it may be better to assign uncertainties to better known aggregate measures of activity. For example, residential fuel use might be estimated as the difference between total consumption and usage in the transportation, industrial, and commercial sectors. In this case, the estimate of uncertainty in residential fuel use is negatively correlated with the uncertainties in fuel use in the other sub-categories, and may even be very large compared to the uncertainty in total consumption. Thus, rather than try to estimate uncertainties separately for each sub-category, it would be more practical to estimate uncertainty for total consumption, for which good estimates and cross-checks may be available.

6.5.5 Does correlation matter?

An important point to remember is that dependencies, even if they exist, may not be important to the assessment of uncertainties. Dependencies among inputs will matter only if the dependencies exist between two inputs to which the uncertainty in the inventory is sensitive and if the dependencies are sufficiently strong. In contrast, weak dependencies among inputs, or strong dependencies among inputs to which the uncertainty in the inventory is insensitive, will be of relatively little consequence to the analysis.

6.5.6 Some methods for dealing with dependencies or correlation

When dependencies among inputs are judged to be of importance, then a variety of techniques can be considered for incorporating them into the analysis. Examples include: (i) modelling the dependence explicitly; (ii) stratifying or aggregating the source categories to minimise the effect of the dependencies; (iii) simulating correlation using restricted pairing methods (that are included in many software packages); (iv) use of resampling techniques in cases where multivariate datasets are available; (v) considering bounding or sensitivity cases (e.g. one case assuming independence and another case assuming complete positive correlation). More discussion and examples of these types of methods are given in Cullen and Frey (1999), Morgan and Henrion (1990), and USEPA (1996). These documents also contain reference lists with citations to relevant literature.

6.5.7 Specifying correlation in inventory inputs

Many software packages allow specification of a correlation between model inputs (e.g. Iman and Conover, 1982). In some cases, these packages may offer this feature only with LHS, while in others it may be available also in conjunction with random Monte Carlo simulation. There is a detail regarding these methods that advanced users will want to note, which is that these software tools can induce a rank correlation between any two or more distributions, but not a sample correlation. However, there are methods that one can use to specify sample correlation in some types of distributions, such as for multivariate normal distributions (see Morgan and Henrion, 1990 or Cullen and Frey, 1999 for examples).

6.5.8 Analysing inventory outputs

Many software packages allow the user to display probability density functions (PDF), cumulative distribution functions (CDF) and provide output of summary statistics for a given model input. Typically, the CDF will appear to be a smoother curve than the PDF for any given case. Furthermore, the CDF allows for quantitative interpretations of the median, 95 percent confidence interval, or any other percentile of the distribution. Thus, for practical purposes, the CDF is often the most useful representation of the model output. The PDF is useful for obtaining qualitative insights regarding the output, such as whether it is positively skewed.

6.5.9 Encouragement of use of appropriate techniques

The guidance offered here is not meant to preclude the use of improved methods as they become available. Furthermore, this document does not cover all situations that may be faced by an analyst. Therefore, the inventory agency is encouraged to refer to the references cited below for additional suggestions on how to perform uncertainty analyses.

6.6 CONCLUSION

The methods described in this chapter should enable inventory agencies to estimate and report the uncertainty in total emissions in any year, and the uncertainty in the trend between years, together with the contribution that each source category makes to these overall uncertainties. This information should help prioritise efforts to improve the precision of inventories in future, and can show how the overall and trend uncertainties respond as the uncertainties in individual source categories are reduced.

APPENDIX 6A.1
DERIVATION OF THE FORMULAE IN TABLE 6.1 (TIER 1)

EXPLANATION OF THE VARIABLES

C_x = Value of an entry in column C and row x

n = number of emission categories (rows)

$\sum C_i$ = Sum over all emissions categories (rows) of the inventory from i=1 to i=n

COLUMN A-F

Input data

COLUMN G

Combined uncertainty using error propagation equation

$$G_x = \sqrt{E_x^{\,2} + F_x^{\,2}}$$

COLUMN H

Combined uncertainty as a percentage of the total emissions in year t.

$$H_x = \frac{G_x \bullet D_x}{\sum D_i}$$

The total in column H (total emission uncertainty) is obtained using the error propagation equation:

$$\text{Total of column H} = \frac{\sqrt{\sum_x \left[\left(\sum_x D_x \right)^2 \bullet (H_x)^2 \right]}}{\sum_x D_x} = \sqrt{\sum_x H_x^{\,2}}$$

COLUMN I

Entries in column I show how the difference in emissions between the base year and the year t changes in response to a 1% increase in emissions of source category *x* emissions in the base year and year t. This shows the sensitivity of the trend in emissions to a systematic uncertainty in the emission estimate – i.e. one that is correlated between the base year and year t. This sensitivity is described as type A sensitivity.

I_x = percentage trend if source category x is increased by 1% in both years – percentage trend without increase

$$= \frac{0.01 \bullet D_x + \sum D_i - (0.01 \bullet C_x + \sum C_i)}{(0.01 \bullet C_x + \sum C_i)} \bullet 100 - \frac{\sum D_i - \sum C_i}{\sum C_i} \bullet 100$$

COLUMN J

Entries in column J show how the difference in emissions between the base year and year t changes in response to a 1% increase in the emissions of source category *x* in year t only. This shows the sensitivity of the trend in emissions to random uncertainty error in the emissions estimate – i.e. one that is not correlated between the base year and year Y. This sensitivity is described as type B sensitivity.

J_x = percentage trend if source category x is increased by 1% in year t – percentage trend without increase

$$= \frac{0.01 \bullet D_x + \sum D_i - \sum C_i}{\sum C_i} \bullet 100 - \frac{\sum D_i - \sum C_i}{\sum C_i} \bullet 100$$

$$= \frac{D_x}{\sum C_i}$$

COLUMN K

Under the assumption that the same emission factor is used in both years and the actual emission factors are fully correlated, the % error introduced by it is equal in both years. Therefore the formula for the uncertainty introduced on the trend by the emission factor is:

$$K_x = \text{sensitivity A} \bullet \text{uncertainty of emission factor}$$
$$= I_x \bullet F_x$$

In case no correlation between emission factors is assumed, sensitivity B should be used and the result needs to be increased by $\sqrt{2}$ for the reason given below in the main derivation for column L:

$$K_x = \text{sensitivity B} \bullet \text{uncertainty of emission factor} \bullet \sqrt{2}$$
$$= J_x \bullet F_x \bullet \sqrt{2}$$

COLUMN L

The trend is the difference between the emissions in the base year and in the year t. Therefore the uncertainty of the activity data of the base year and t has to be taken into account. The two uncertainties combined using the error propagation equation and the assumption that the uncertainty is the same in the base year and year t is:

$$= \sqrt{(\text{uncertainty (activity data, base year)})^2 + (\text{uncertainty (activity data, year t)})^2}$$
$$\approx \sqrt{(\text{uncertainty (activity data, year t)})^2 \bullet 2}$$
$$= E_x \bullet \sqrt{2}$$

Since activity data in both years are assumed to be independent, column L equals:

L_x = sensitivity B • combined uncertainty of activity data of both years

$$= J_x \bullet E_x \bullet \sqrt{2}$$

In case correlation between activity data is assumed, sensitivity A should be used and the √2 factor does not apply.

$$L_x = I_x \bullet E_x$$

COLUMN M

In column M figures the combined uncertainty introduced on the trend by the uncertainty in the activity data and the emissions factor.

$$M_x = \sqrt{K_x^{\,2} + L_x^{\,2}}$$

The entries M_i in column M are combined to obtain the total uncertainty of the trend using the error propagation equation as follows:

$$\text{Total of column M} = \frac{\sqrt{\sum_x \left[\left(\sum_x (D_x - C_x) \right)^2 \bullet (M_x)^2 \right]}}{\sum_x (D_x - C_x)} = \sqrt{M_1^{\,2} + M_2^{\,2} + \ldots + M_n^{\,2}}$$

APPENDIX 6A.2
TIER 1 UNCERTAINTY CALCULATION EXAMPLE

The following spreadsheet shows an example calculation for the national greenhouse gas inventory of the United Kingdom.

TABLE 6.3

TIER 1 UNCERTAINTY CALCULATION AND REPORTING EXAMPLE

	A	B	C	D	E	F	G	H	I	J	K	L	M
	IPCC Source Category	Gas	Base year emissions 1990	Year t emissions 1997	Activity data uncertainty	Emission factor uncertainty	Combined uncertainty	Combined uncertainty as % of total national emissions in year t	Type A sensitivity	Type B sensitivity	Uncertainty in trend in national emissions introduced by emission factor uncertainty	Uncertainty in trend in national emissions introduced by activity data uncertainty	Uncertainty introduced into the trend in total national emissions
			Gg CO2 equivalent	Gg CO2 equivalent	%	%	%	%	%	%	%	%	%
1A	Coal	CO2	238 218	142 266	1.2	6	6.1	1.2	-0.0966	0.1840	-0.58	0.31	0.66
1A	Oil	CO2	208 684	196 161	1	2	2.2	0.6	0.0076	0.2538	0.02	0.36	0.36
1A	Natural Gas	CO2	111 052	181 691	2	1	2.2	0.6	0.1039	0.2351	0.10	0.66	0.67
1A	Other (waste)	CO2	138	741	7	20	21.2	0.0	0.0008	0.0010	0.02	0.01	0.02
1B	Solid Fuel Transformation	CO2	2 573	1 566	1.2	6	6.1	0.0	-0.0010	0.0020	-0.01	0.00	0.01
1B	Oil & Natural Gas	CO2	8 908	6 265		14	14.0	0.1	-0.0024	0.0081	-0.03	0.00	0.03
2A1	Cement Production	CO2	6 693	6 157	1	2	2.2	0.0	0.0001	0.0080	0.00	0.01	0.01
2A2	Lime Production	CO2	1 192	1 703	1	5	5.1	0.0	0.0008	0.0022	0.00	0.00	0.01
2A3	Limestone & Dolomite use	CO2	1 369	1 551	1	5	5.1	0.0	0.0004	0.0020	0.00	0.00	0.00
2A4	Soda Ash Use	CO2	116	120	15	2	15.1	0.0	0.0000	0.0002	0.00	0.00	0.00
2B	Ammonia Production	CO2	1 358	814		5	5.0	0.0	-0.0005	0.0011	0.00	0.00	0.00
2C1	Iron&Steel Production	CO2	3 210	1 495	1.2	6	6.1	0.0	-0.0019	0.0019	-0.01	0.00	0.01
5D	Land Use Change & Forestry	CO2	31 965	27 075	5	54	54.2	2.1	-0.0027	0.0350	-0.14	0.25	0.29
6C	MSW Incineration	CO2	660	29	7	20	21.2	0.0	-0.0007	0.0000	-0.01	0.00	0.01
		CO2 Total	616 137	567 634									
1A	All Fuel	CH4	2 507	1 975	1.2	50	50.0	0.1	-0.0004	0.0026	-0.02	0.00	0.02
1B1	Coal Mining	CH4	17 188	6 687	1	13	13.0	0.1	-0.0116	0.0087	-0.15	0.01	0.15
	Solid Fuel Transformation	CH4	215	173	6	50	50.4	0.0	0.0000	0.0002	0.00	0.00	0.00
1B2	Natural Gas Transmission	CH4	8 103	7 301	2	15	15.1	0.2	-0.0001	0.0094	0.00	0.03	0.03
	Offshore Oil& Gas	CH4	2 402	1 957	10	26	27.9	0.1	-0.0003	0.0025	-0.01	0.04	0.04
2C	Iron & Steel Production	CH4	16	13	1.2	50	50.0	0.0	0.0000	0.0000	0.00	0.00	0.00
4A	Enteric Fermentation	CH4	19 177	18 752	1	20	20.0	0.5	0.0016	0.0243	0.03	0.03	0.05
4B	Manure Management	CH4	2 338	2 325	1	30	30.0	0.1	0.0003	0.0030	0.01	0.00	0.01
4F	Field Burning	CH4	266	0	25	50	55.9	0.0	-0.0003	0.0000	-0.02	0.00	0.02
6A	Solid Waste Disposal	CH4	23 457	17 346	15	46	48.4	1.2	-0.0052	0.0224	-0.24	0.48	0.53
6B	Wastewater Handling	CH4	701	726	15	48	50.3	0.1	0.0001	0.0009	0.01	0.02	0.02
6C	Waste Incineration	CH4	1	1	7	50	50.5	0.0	0.0000	0.0000	0.00	0.00	0.00
		CH4 total	76 371	57 257									

TABLE 6.3

TIER 1 UNCERTAINTY CALCULATION AND REPORTING EXAMPLE

	A	B	C	D	E	F	G	H	I	J	K	L	M
	IPCC Source Category	Gas	Base year emissions 1990	Year t emissions 1997	Activity data uncertainty	Emission factor uncertainty	Combined uncertainty	Combined uncertainty as % of total national emissions in year t	Type A sensitivity	Type B sensitivity	Uncertainty in trend in national emissions introduced by emission factor uncertainty	Uncertainty in trend in national emissions introduced by activity data uncertainty	Uncertainty introduced into the trend in total national emissions
			Gg CO2 equivalent	Gg CO2 equivalent	%	%	%	%	%	%	%	%	%
1A2& 1A4& 1A5	Other Combustion	N2O	3 865	3 562	1.2	195	195.0	1.0	0.0001	0.0046	0.01	0.01	0.01
1A3	Transport	N2O	1 300	3 645	1.4	170	170.0	0.9	0.0032	0.0047	0.54	0.01	0.54
1B2	Oil & Natural Gas	N2O	3	2	10	110	110.5	0.0	0.0000	0.0000	0.00	0.00	0.00
2B	Adipic Acid Production	N2O	25 136	17 766	0.5	15	15.0	0.4	-0.0067	0.0230	-0.10	0.02	0.10
2B	Nitric Acid Production	N2O	4 383	3 723	10	230	230.2	1.2	-0.0004	0.0048	-0.08	0.07	0.11
4B	Manure Management	N2O	1 583	1 559	1	509[a]	509.0	1.1	0.0002	0.0020	0.08	0.00	0.08
4D	Agricultural Soils	N2O	29 472	29 098	1	509	509.0	21.0	0.0029	0.0376	1.47	0.05	1.47
4F	Field Burning	N2O	78	0	10	230	230.2	0.0	-0.0001	0.0000	-0.02	0.00	0.02
6B	Wastewater Handling	N2O	153	157	1	100	100.0	0.0	0.0000	0.0002	0.00	0.00	0.00
6C	Waste Incineration	N2O	115	11	7	230	230.1	0.0	-0.0001	0.0000	-0.03	0.00	0.03
		N2O Total	66 089	59 525									
2	Industrial Processes	HFC	11 374	18 447	2	25	25.1	0.7	0.0104	0.0239	0.26	0.07	0.27
3	Industrial Processes	PFC	2 281	661	5	19	19.6	0.0	-0.0018	0.0009	-0.03	0.01	0.04
4	Industrial Processes	SF6	724	1 170	10	8	12.8	0.0	0.0007	0.0015	0.01	0.02	0.02
			Halocarbon & SF6 Total										
	Total Emissions	GWP weighted total	772 976	704 693									
	Total Uncertainties				Overall uncertainty in the year (%)			21.3			Trend uncertainty (%)		2.0

[a] Uncertainty estimated from lognormal distribution used in Monte Carlo simulation. I.e. (97.5 percentile-mean)/mean *100.

REFERENCES

Ang, A.H.S. and W.H. Tang (1975). *Probability Concepts in Engineering Planning and Design, Volume 1*. John Wiley and Sons, New York.

Cullen, A.C. and H.C. Frey (1999*). Probabilistic Techniques in Exposure and Risk Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. Plenum Press, New York.

D'Agostino, R.B. and M.A. Stephens (1986). *Goodness of Fit Techniques*. Marcel Dekker, New York.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, New York.

Eggleston, S. *et al.* (1998). *Treatment of Uncertainties for National Greenhouse Gas Emissions.* Report AEAT 2688-1 for DETR Global Atmosphere Division, AEA Technology, Culham, UK.

Fishman G.S. (1996). *Monte Carlo: Concepts, Algorithms, and Applications*. Springer-Verlag, New York.

Frey, H.C. and D.E. Burmaster (1999). 'Methods for Characterization of Variability and Uncertainty: Comparison of Bootstrap Simulation and Likelihood-Based Approaches'. *Risk Analysis*, 19(1): 109-129.

Frey, H.C. and D.S. Rhodes (1996). 'Characterizing, Simulating, and Analyzing Variability and Uncertainty: An Illustration of Methods Using an Air Toxics Emissions Example'. *Human and Ecological Risk Assessment*, 2(4), pp. 762-797.

Frey, H.C., Bharvirkar, R., Thompson, R. and Bromberg, S. (1998). 'Quantification of Variability and Uncertainty in Emission Factors and Inventories'. *Proceedings of Conference on the Emission Inventory*, Air & Waste Management Association, Pittsburgh, PA, USA.

Frey, H.C., Bharvirkar, R. and Zheng, J. (1999). *Quantitative Analysis of Variability and Uncertainty in Emissions Estimation, Final Report*. Prepared by North Carolina State University for U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA, July 1999. Available at www4.ncsu.edu/~frey/

Hahn, G. J. and Shapiro, S. S. (1967). *Statistical Models in Engineering*, John Wiley and Sons, New York.

Hora, S.C., and R.L. Iman (1989). 'Expert opinion in risk analysis: The NUREG-1150 methodology'. *Nuclear Science and Engineering*, 102, pp. 323-331.

Iman, R. L. and Conover, W. J. (1982). 'A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables'. *Communications in Statistics*, B11(3), pp. 311-334.

McCann, T.J. & Associates and Nosal, M. (1994). *Report To Environment Canada Regarding Uncertainties In Greenhouse Gas Emission Estimates.* Calgary, Canada.

Merkhofer, M.W. (1987). 'Quantifying judgmental uncertainty: methodology, experiences, and insights'. *IEEE Transactions on Systems, Man, and Cybernetics*, 17(5), pp. 741-752.

Morgan, M.G. and M. Henrion (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, New York.

NCRP (National Council on Radiation Protection and Measurements) (1996). *A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination*. NCRP Commentary No. 14, Bethesda, MD, USA.

Robinson, J.R. (1989). 'On Uncertainty in the Computation of Global Emissions for Biomass Burning'. *Climatic Change*, 14, pp. 243-262.

Spetzler, C.S., and von Holstein, S. (1975). 'Probability encoding in decision analysis'. *Management Science*, 22(3), pp. 340-358.

USEPA (1996). *Summary Report for the Workshop on Monte Carlo Analysis*. EPA/630/R-96/010, Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC, USA.

USEPA (1997). *Guiding Principles for Monte Carlo Analysis.* EPA/630/R-97/001, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC, USA, March 1997. Available on World Wide Web at: http://www.epa.gov/ncea/mcpolicy.htm.

USEPA (1999). *Report of the Workshop on Selecting Input Distributions For Probabilistic Assessments.* EPA/630/R-98/004, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC, January 1999. Available on World Wide Web at: http://www.epa.gov/ncea/input.htm.

Wackerly, D. D., Mendenhall III, W. and Scheaffer, R. L. (1996). *Mathematical Statistics with Applications.* Duxbury Press, USA.