CHAPTER 3

UNCERTAINTIES

[Parts shaded in grey – the unchanged text from the 2006 IPCC Guidelines]

Authors

Newton Paciornik

Michael Gillenwater, Riccardo De Lauretis, Daniela Romano

Contributing Authors

Suvi Monni, Stephen Ogle, Cody Alsaker

Contents

Equations

Figures

Tables

Boxes

3 UNCERTAINTIES

3.1 INTRODUCTION

This chapter provides guidance in estimating and reporting uncertainties associated with both annual estimates of emissions and removals, and emission and removal trends over time. It also elaborates on the importance of uncertainty assessment as a means of improving emission inventories over time. It is written from the viewpoint of the inventory compiler and provides, with examples, two approaches for combining category uncertainties into uncertainty estimates for total national net emissions and the trend.

3.1.1 Overview of uncertainty analysis

[Elaboration of Section 3.1.1 of the *2006 IPCC Guidelines*].

Uncertainty assessment is an important part of the effort of compiling an inventory of anthropogenic emissions and removals of GHGs (GHG inventory) and to assess its evolution over time. Since the GPG2000 report, the IPCC has adopted the concept of "*Good Practice*" in developing a GHG inventory, defined as an inventory that "contains neither over- nor under-estimates so far as can be judged, and in which uncertainties are reduced as far as practicable".

The first notion that emerges from this concept is that it is impossible to eliminate uncertainty completely, leading to the immediate conclusion that for every value reported in an inventory there will exist an associated uncertainty. Knowledge of this uncertainty is an integral part of the inventory compilation effort.

The second notion that follows is that, as a priority, effort should focus on accuracy, ensuring that emissions and removals are neither over- nor under-estimated. In short, bias should be eliminated as far as can be judged. Figure 3.2 of the *2006 IPCC Guidelines* gives a good illustration of the difference between accuracy and precision clearly showing that a precise estimate is of limited value if it is not accurate.

The key word is "knowledge". Knowing the processes involved and the information available is key to quantify and reduce uncertainty. While variability is a characteristic of the process and cannot be eliminated, uncertainty, which covers both random error and bias, can be broadly associated with lack of knowledge. Causes of uncertainty are described in section 3.1.5 of the *2006 IPCC Guidelines* and further discussed in section 3.1.5 of this *2019 Refinement*.

Uncertainty calculation is strongly linked to the methods used to estimate emissions and removals. Simple methods are based on the multiplication of activity data (AD) by an emission factor (EF). More generally, both AD and EF can be the result of several different parameters (see section 3.2.3 for a discussion). For some complex systems, models are developed for their description (including spatial-temporal scales), evaluation of emissions and calculation of uncertainty.

Regardless of the complexity of the approach, uncertainty of the results is a function of the uncertainty of data (activity or emission factors) used to compile the inventory. Hence, data collection and uncertainty evaluation are strongly linked. In short, all data collected should have an associated uncertainty assessment (further discussed in section 3.2).

Finally, it is important to point out that producing an uncertainty analysis result (level or trend) for the inventory is not an independent goal. The uncertainty values are not absolute measures of the overall quality of the inventory. Even if they depend on the level of the complexity of the estimation methods and uncertainty calculation approaches, they are also a function of the share of sectors and categories in each country. Moreover, the uncertainty analysis as a whole is an important tool in the process of improvement of the inventory. Together with the *key category* analysis, it helps the inventory compilers in prioritizing the improvements in methodology development and data collection for the different source and sink categories (see section 3.1.2).

3.1.1a Uncertainty assessment as part of inventory management

[New guidance in the *2019 Refinement*].

The uncertainty assessment is one of the instruments that will be used by the inventory compiler in the effort of improving the inventory over time. Regardless of the framework, under which national GHG inventories are developed and reported this will not be a one-time task. Inventories will be reported annually, biannually or over longer periods but will be updated and extended periodically.

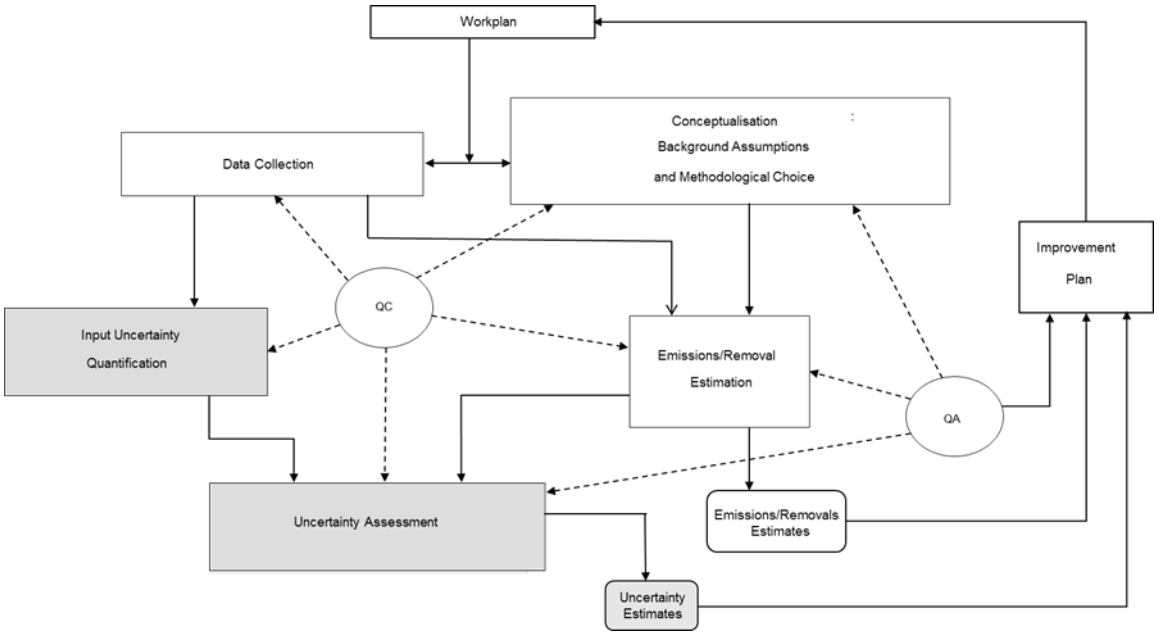
Between two reporting occasions, it is *good practice* to evaluate the data sources, data flows and methods used. Ideally the inventory would have been verified by a third party and recommendations produced (e.g. reviews under the UNFCCC) in accordance with guidance provided in Chapter 4 of the 2006 of Volume 1 of the *2006 IPCC Guidelines*. Figure 1.1 in Chapter 1.1 of Volume 1 of the *2006 IPCC Guidelines* illustrates the steps of a typical inventory cycle and Chapter 1 of this report covers the steps to put in place the institutional arrangements necessary to manage the process, providing the organization and resources for planning and preparation of the inventory. Figure 3.1 below, builds from Figure 3.1 in Chapter 3 of Volume 1 of the *2006 IPCC Guidelines* to show how the uncertainty assessment fits in this improvement cycle.

The process of producing an uncertainty analysis can pragmatically be divided into four parts: (1) the rigorous investigation of the likely causes of data uncertainty and quality; (2) the creation of quantitative uncertainty estimates and parameter correlations; (3) the mathematical combination of those estimates when used as inputs to a statistical model (e.g., first-order error propagation or Monte Carlo method); and (4) the selection of inventory improvement actions (improvement plan) to take in response to the results of the previous three parts.

The improvement plan will assess the opportunities and prioritize the ways to improve the inventory based on the *key category* analysis, the uncertainty assessment, the recommendations from quality assurance and verification processes (including review process) and available resources.

Particularly in relation to the uncertainty analysis, the improvement plan will investigate ways to improve accuracy that would have been identified and ways to enhance precision for categories with high contribution to the overall uncertainty of the inventory. The approach 2 for *key category* analysis is a useful tool for this prioritization.

Figure 3.1 Overall structure of a generic uncertainty analysis



3.1.2 Overall structure of uncertainty analysis

[Elaboration of Section 3.1.2 of the *2006 IPCC Guidelines*].

As part of the planning process, an improvement plan will be developed selecting the categories for which changes would be implemented in the new inventory. The changes would cover both methodological choice and data specification, availability and collection. Often the improvement focuses on getting better data for the same methodology (e.g. collecting country-specific data). The goal is generally to increase the accuracy of the inventory through a better representation of the emissions/removals processes.

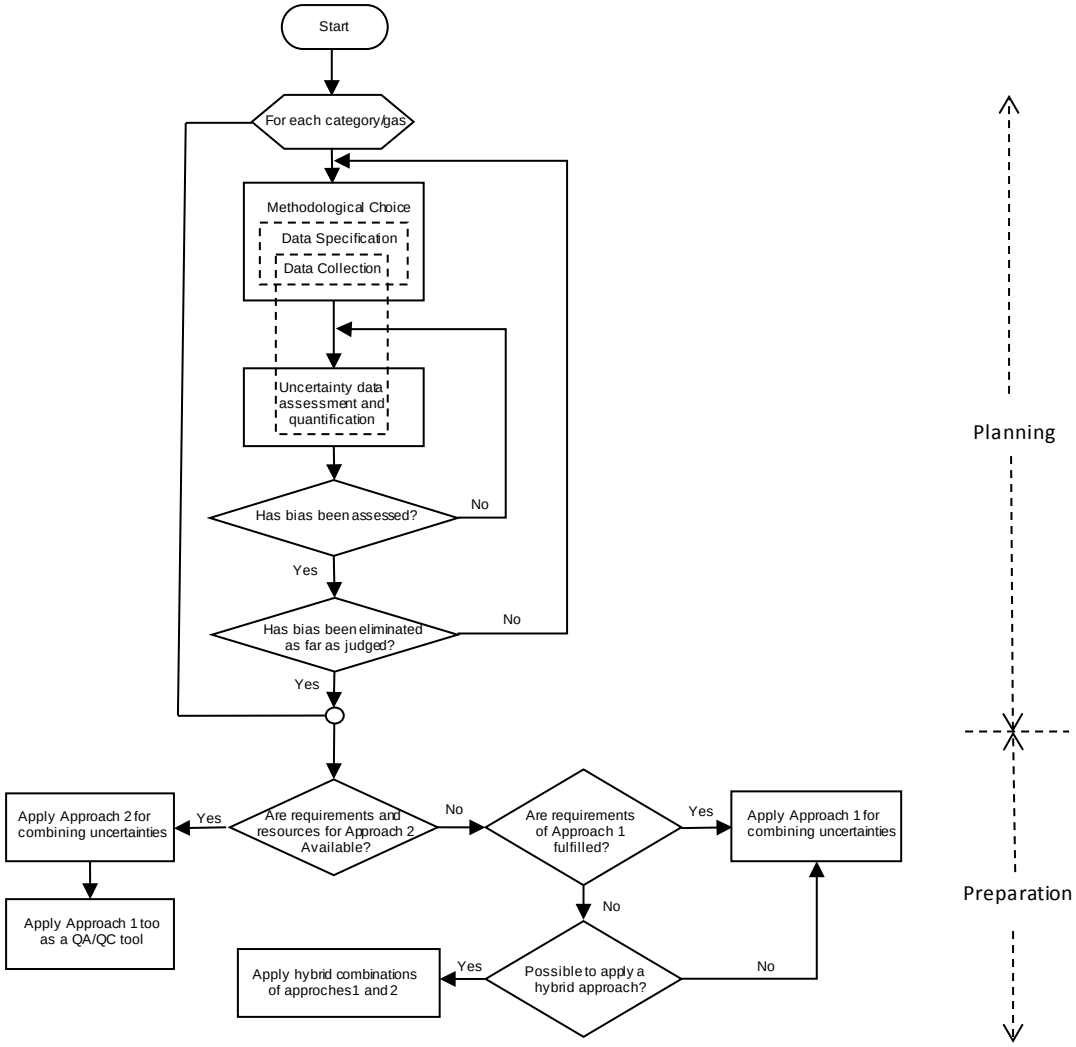
Figure 4.1 of Volume 1 of the *2006 IPCC Guidelines*, shows the steps of methodology choice that will depend on the selection of the category for improvement, the data availability, the possibility of data obtaining and the resources involved.

Figure 3.2 below show the general steps of an uncertainty assessment. It is important to note the strong link among these steps that usually need to be taken in conjunction and frequently reevaluated. This is true between the data definition and collection and between the data collection and the associated data uncertainty.

When assessing data uncertainty, it is essential to identify the causes of uncertainty involving the data estimation. In particular, priority should be given to identifying and correcting causes of bias.

Following the assessment of the uncertainty of the pieces of data used in emissions/removals estimation, the next step is to combine these findings, producing uncertainty assessment for a source or sink category that is subsequently propagated with uncertainties in all categories to determine the uncertainty in the whole inventory. Figure 3.2 show a simple scheme for the choice of approach but it is important to note that choices may vary among categories and usually a hybrid approach would be recommended. It is also important to note that even when requirements for application of approach 1 are not fully present it still can provide useful information about the uncertainty of the inventory. Because of its simplicity when compared with approach 2, it is also recommended to apply approach 1 as a QA/QC tool even when it is possible to apply approach 2.

Figure 3.2 Uncertainty analysis steps description and decision tree



3.1.3 Key concepts and terminology

No refinement.

3.1.4 Basis for uncertainty analysis

No refinement.

3.1.5 Causes of uncertainty

[Elaboration of section 3.1.5 of the *2006 IPCC Guidelines*].

Section 3.1.5 of the *2006 IPCC Guidelines* provides a description of the causes of uncertainty. It covers eight possible causes: lack of completeness, model, lack of data, lack of representativeness of data, statistical random sample error, measurement error, misreporting or misclassification and missing data. Depending on the cause, the result can be biases, random errors or both. Lack of completeness, lack of representativeness of data, misreporting or misclassification will typically lead to bias while model uncertainty and lack of data can lead to both.

For each category, the identification of causes of uncertainty is fundamental for elimination of bias and quantification of random errors. A poor identification step will entirely compromise an uncertainty reducing effort.

3.1.6 Reducing uncertainty

[Elaboration of Section 3.1.6 of the *2006 IPCC Guidelines*].

Uncertainties should be reduced as far as is practicable during the process of compiling an inventory, and it is particularly important to ensure that the model and the data collected are fair representations of the real world. When focusing efforts to reduce uncertainty, priority should be given to those inputs to the inventory that have the most impact on the overall uncertainty of the inventory, as opposed to inputs that are of minor or negligible importance to the assessment as described in Chapter 4, Methodological Choice and Identification of Key Categories. Tools for prioritising where uncertainties should be reduced include *key category* analysis (see Chapter 4) and assessment of the contribution of uncertainties in specific categories to the total uncertainty in the inventory (see Section 3.2.3). Depending on the cause of uncertainty present, uncertainties could be reduced in seven broad ways:

- *Improving conceptualisation*: Improving the inclusiveness of the structural assumptions chosen can reduce uncertainties. An example is better treatment of seasonality effects that leads to more accurate annual estimates of emissions or removals for the AFOLU Sector.
- *Improving models*: Improving the model structure and parameterisation can lead to better understanding and characterisation of the systematic and random errors, as well as reductions in these causes of uncertainty.
- *Improving representativeness*: This may involve stratification or other sampling strategies, as set out in Section 3.2.1.2. This is particularly important for categories in the agriculture, forestry and land use parts of an inventory, but also applies elsewhere, e.g., wherever different technologies are operating within a category. For example, continuous emissions monitoring systems (CEMS) can be used to reduce uncertainty for some sources and gases as long as the representativeness is guaranteed. CEMS produces representative data at the facilities where it is used, but in order to be representative of an entire source category, CEMS data must be available for a sample or an entire set of individual facilities that comprise the category. When using CEMS both GHG emissions concentration and flow will vary, requiring simultaneous observations of both attributes.
- *Using more precise measurement methods*: Measurement error can be reduced by using more precise measurement methods, avoiding simplifying assumptions, and ensuring that measurement technologies are appropriately used and calibrated. See Chapter 2, Approaches to Data Collection.
- *Collecting more data that are measured*: Uncertainty associated with random sampling error can be reduced by increasing the sample size. Both bias and random error can be reduced by filling in data gaps. This applies both to measurements and surveys.

- *Eliminating known risk of bias*: This is achieved by ensuring instrumentation is properly positioned and calibrated (see Section 2.2 in Chapter 2), models or other estimation procedures are appropriate and representative as indicated by the decision trees and other advice on methodological choice in sectoral volumes, and by applying expert judgements in a systematic way.

- *Improving state of knowledge*: Generally, improving the understanding of the categories and the processes leading to emissions and removals can help to discover, and correct for, problems of incompleteness. It is *good practice* to continuously improve emissions and removal estimates based on new knowledge (see Chapter 5, Time Series Consistency).

- *Moving to higher tier method*: For example, Tier 1 emission factors that are considered global defaults may be biased when they are applied in a specific country where emission rates deviate by a specific amount from the global defaults. Moving to a higher tier method in this case, will remove the bias associated with the default emission factor. Applying a higher tier method may also improve the precision of estimates as shown in Box 3.1.

The effort to reduce uncertainty is also one that is tightly integrated with data collection and QA/QC processes. In many ways, it is an in-depth approach to quality management. Both uncertainty analysis and QA/QC processes require rigorous investigation into the causes of data quality problems, especially ones that general QC checks are unlikely to identify. These problems will often involve issues of incomplete data or other systematic biases in the data, which also happen to be key issues for developing a quantitative uncertainty analysis (Gillenwater *et al.*, 2007).

Both QA/QC and uncertainty analysis are part of a learning process. While the uncertainty analysis provides a standalone quantitative assessment of the inventory, its primary function is to understand what produces uncertainty and how to improve inventory quality. Conversely, the outcome of QA/QC procedures may result in a reassessment of individual category or parameter uncertainty estimates. Procedures to check quality and analyse uncertainties overlap and should work together because both processes are intended to understand the causes of uncertainty and identify potential areas of improvement (US-EPA, 2002).

BOX 3.1

EXAMPLE OF REDUCING UNCERTAINTY IN A SOURCE CATEGORY BY ADOPTING HIGHER TIER METHODS

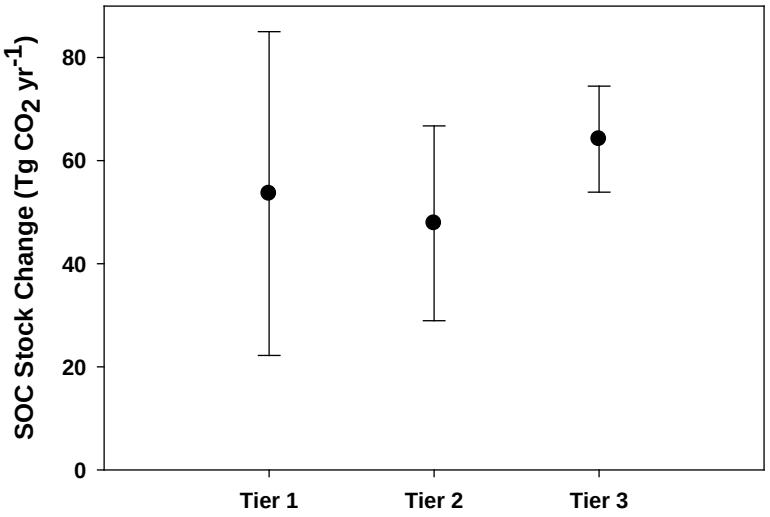
Mineral soil C stock changes for *Cropland Remaining Cropland* have been estimated with all three methodological tiers for the United States, and this box provides information about how uncertainty has been reduced by moving to higher tiers. Each of the methods used a Monte Carlo Analysis for propagating uncertainties addressing key dependencies in the underlying data, such as the relationship among the land use areas. As with other source categories, the Tier 1 method is relatively simple with default emission factors provided in the IPCC guidance, but does require compilation of activity data for a simple classification of the lands, climate and soils. The IPCC guidance (2006) provides uncertainties in emission factors, while uncertainties in land use and management data were derived from the survey data that are used in the inventory. For example, land use data were based on a two-stage survey design that was used to derive joint probability distributions for land use and land use change over the inventory time series. By moving to Tier 2, the compilers derived country-specific emission factors (i.e., stock change factors) that were based on experimental data from the region (Ogle *et al.*, 2003). Specifically, the new factors were derived using a linear mixed-effect modelling approach from 46 experiments evaluating the effect of tillage management on soil C, 19 experiments evaluating the impact of variation in carbon input to soils, and 35 experiments evaluating the impact of land use change between native conditions and long-term cultivation. Compilers also had the option of refining the land representation and activity data into a country-specific set of climate and soil types, in addition to management classes. However, the compilers did not change the classification in this application, and so the uncertainties in activity data were the same for the Tier 1 and 2 methods. Regardless, flexibility in deriving new emission factors improved the precision of the estimates, reducing the confidence interval for the estimated soil C stock changes from ± 59% using the Tier 1 method to a ± 40% for the Tier 2 method (Figure, US-EPA 2017).

The compilers further improved the inventory for *Cropland Remaining Cropland* by developing a Tier 3 method. This method was based on applying the Century Ecosystem Model, and later the DayCent Ecosystem Model (Ogle *et al.* 2010, US-EPA 2017). These models incorporate a more mechanistic representation of the processes influencing soil organic matter dynamics, including water flows through the soil, crop production, organic matter decomposition, and nutrient cycling (Parton *et al.* 1987). With a more advanced representation of processes, the inventory was able to capture a broader suite of drivers influencing the change in soil C stocks. In addition, the inventory incorporated more detailed information on activity data and environmental variables, such as weather, soils, and management practices. There were additional uncertainties associated with these activity data, such as the variability in specific N fertilisation rates. Several of the main datasets, however, such as land use and cropping histories, did not differ across the three methods. In theory, Tier 3 methods allow compilers to develop a methodology that is more specific to national circumstances, and ultimately an approach meeting *good practice* that is working towards the goal of neither over nor under-estimating emissions (or removals) as far as can be judged. To address uncertainty in the emission rates (i.e., analogous to the emission factors for the Tier 1 and 2 methods), the compilers evaluated uncertainty (i.e., bias and precision) in the Century/DayCent model predictions of soil C stock changes by comparing results to independent measurement data. They used these data comparisons to develop an empirical model to adjust for biases and assess precision in the inventory results (Ogle *et al.* 2007). The Tier 3 inventory reduced uncertainty in soil C stock change estimates over 5 years from a ± 40% with the Tier 2 method to ± 16% for the Tier 3 method (see Figure below).

It is difficult to know that the improvements produce more accurate estimates because the compilers do not have an omniscient view of the emissions (if they did, then estimation would be unnecessary). However, incorporating data specific to a country for estimating Tier 2 emission factors will better represent the population of emission sources in the country. The Tier 1 factors are based on samples from a larger global population, which has considerably more variation in climates, soils and other variables driving soil organic matter dynamics, and all of this variation is not relevant for an individual country. Of course, the accuracy of the Tier 2 factors also depends on an adequate sample of emission measurements in a country. For the Tier 3 method, the compilers incorporated scientific understanding of soil organic matter dynamics using the Century/DayCent model, which embodies key processes and structure that influence soil C stock changes. In turn, the compilers could estimate management impacts on soil C stock changes with more specificity to physical and biogeochemical conditions of the plant-soil environments in the country than is possible with the lower tier methods. The compilers quantified the accuracy and precision of the model estimates based on independent data to ensure that confidence intervals incorporated model prediction error.

Emissions estimates will not always be more precise or even accurate with higher tier methods. Ultimately, the level of improvement will depend on the precision of the model inputs, representativeness of the model and/or emissions data, and implementation, and requires sufficient data and testing.

Figure: Estimates and 95% confidence intervals for mineral soil C stock changes in *Cropland Remaining Cropland* in the United States using Tier 1, 2 and 3 methods.



3.1.7 Implications of methodological choice

No refinement.

3.2 QUANTIFYING UNCERTAINTIES

[Elaboration of section 3.2 of the *2006 IPCC Guidelines*].

Regardless of the methodology used to estimate emissions/removals for a category, the evaluation will be based on the underlying data. The overall uncertainty of the emissions/removals will depend on the uncertainty associated with each and every piece of data that is used to inform the inventory. As such, *good practice* uncertainty assessment begins with *good practice* in data collection. Uncertainty consideration will need to be an integral part of the data collection effort, including selection of data sources and choice of methods following the guidance in Chapter 2 of Volume 1 of the *2006 IPCC Guidelines.*

Section 3.2 of Volume 1 of the *2006 IPCC Guidelines* covers the different techniques for quantifying uncertainties depending on the availability of information and ways of data collection. These include measured data, published information, model outputs and expert judgement. Usually, the pragmatic approach will be a combination of the techniques.

Again, regardless of the approach, it is *good practice* to follow strictly the procedures for QA/QC according to the guidance in Chapter 6 of Volume 1 of the *2006 IPCC Guidelines.* This will be fundamental in preventing mistakes and misreporting and misclassification errors and approach deviations.

Ultimately, the measure of uncertainty will be a 95 percent confidence interval around a point estimate for the value. In order to develop this information a probability density function (PDF) will be associated with each quantity. The development of that PDF is an essential part of the uncertainty assessment. Section 3.2.2.4 of Volume 1 of the *2006 IPCC Guidelines* provide guidance on how to select the PDF. The representativeness of the PDF will depend on the characteristics of the quantity, including domain (e.g., if it can have both positive or negative values, or only non-negative values), range (e.g., is the range narrow or does it cover orders-of-magnitude) and shape (e.g., symmetry). The same characteristics will be fundamental when the approaches for combining uncertainties are selected.

Where the PDF is believed to be symmetrical the confidence interval can be conveniently expressed as plus or minus half the confidence interval width divided by the estimated value of the variable (e.g., ± 10%). Where the PDF is not symmetrical upper and lower limits of the confidence interval need to be specified separately (e.g., -30%, +50%). In both cases, the understanding is that the confidence interval has a 95 percent probability of enclosing the true but unknown value of the emission factor, parameter or activity data.

BOX 3.2

DIFFERENCE BETWEEN STANDARD DEVIATION AND STANDARD ERROR

In case of data assumed normally distributed, the 95% confidence interval may be derived considering the standard deviation (σ) or the standard error (*SE*) around our point estimate. The uncertainty of our estimate (μ) may be expressed as:

Uncertainty = ± (1.96 σ/μ) *100%, where:

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

or

Uncertainty = ± (1.96 *SE*/μ) *100%, where:

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

Where:

n is the number of observations;

x_i are the observed values.

In the formula for the calculation of *SE*, the denominator is assumed to be "*n*" instead of "*n-1*" as an approximation for large samples and to be consistent with the calculation spreadsheet.

Some practical examples may help the inventory compiler choose between these two statistics.

The standard deviation is a measure of variability. The standard deviation of a sample can be used as an estimate of the variability of the population from which the sample was drawn. For data with a normal distribution, about 95% of individuals will have values within 1.96 standard deviations of the mean, the other 5% being equally scattered above and below these limits.

When the sample mean is available, the interest of the inventory compiler is usually not in the mean of that particular sample, but in the mean of the population from which the sample is drawn. For instance, for a sectoral category in the inventory, in order to estimate a specific parameter (e.g. emission factor, carbon stock change factor or AD), data are usually collected with the aim to generalize from them and use the sample mean as an estimate of the average parameter for the whole category.

The sample mean will vary from sample to sample; the way this variation occurs is described by the "sampling distribution" of the mean. The variability of the mean is calculated by the standard deviation of this sampling distribution, which is defined as the standard error of the mean.

The standard error falls as the sample size increases but the standard deviation will not tend to change.

In summary, to calculate the uncertainty of the parameter of concern, the first step is to establish if it derives from:

1). The variability of the population (i.e. how much values of the population are spread), which is measured by the standard deviation; or

2). From the variability of the mean of the samples (i.e. how much the mean values of the samples taken from the population are spread), which is measured by the standard error.

The first case occurs when the mean value is used to estimate an individual of the population (e.g. the average C stock of a forest to infer the C stock a single portion of that forest). The second one occurs when the mean value is used to estimate the entire population (e.g. the average C stock of a forest to infer the C stock of the entire forest).

The following examples are provided for emission factors.

Case 1:

Availability of annual information to derive country specific emission factors of a specific category/gas/fuel. Data are yearly collected from the whole population or a representative sample(s) of the relevant category.

This situation may occur in case data are collected from facilities.

In this case, the annual emission factor is calculated as the average emission factor from repeated measurements in the specific year and it may change over the years. Inventory compilers are therefore interested in the variability of this average annual value.

Assuming a normal distribution of the data collected, the 95% confidence interval may be expressed with the standard error and the uncertainty of the estimated emission factor as:

Uncertainty = ± (1.96 SE/μ) *100%

Case 2:

Availability of irregular information to derive country specific emission factors of a specific category/gas/fuel. Data are not regularly collected and the result of data collected for one single year for a specific category is used for a longer period of the time series.

This situation may occur in case data are sporadically collected from facilities, e.g. methane emissions and relevant activity data and parameters from landfills.

In this case, the 95% confidence intervals can be calculated using the standard deviation of the point estimate because, assuming the value representative of other years, the variability of the population has to be considered. The uncertainty will be:

Uncertainty = ± (1.96 σ/μ) *100%

Case 3:

Availability of annual information to derive country specific emission factors at an upper level than actually used.

This situation may occur if for instance, an average emission factor is available and this country specific value is applied to a specific portion of area, e.g. carbon stock per hectare of deforested area.

As in case 2, the variability of the individuals should be considered to derive the 95% confidence intervals and the uncertainty is to be estimated as:

Uncertainty = ± (1.96 σ/μ) *100%

3.2.1 Sources of data and information

No refinement.

3.2.1.1 UNCERTAINTIES ASSOCIATED WITH MODELS

No refinement.

3.2.1.2 EMPIRICAL DATA FOR SOURCES AND SINKS AND ACTIVITY

This section describes sources of empirical data, and their implications for uncertainty, and is relevant to measured emissions data, data obtained from literature, and activity data.

UNCERTAINTY ESTIMATES OBTAINED FROM MEASURED EMISSIONS/REMOVALS DATA

No refinement

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

No refinement

UNCERTAINTIES ASSOCIATED WITH ACTIVITY DATA

Activity data are often more closely linked to economic activity than are emission factors are. However, unlike emission factor data, there is typically no statistical sample of alternative activity data estimates readily available to fit distributions and estimate uncertainty. 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 than emission factor data. Activity data are often collected and published regularly by national statistical agencies, which may have already assessed the uncertainties associated with their data as part of their data collection procedures. These previously developed uncertainty estimates can be used to construct PDFs. This information will not necessarily have been published, so it is recommended to contact the statistical agencies directly. Since economic activity data are not usually collected for the purpose of estimating greenhouse gas emissions and removals, it is *good practice* to assess the applicability of the uncertainty estimates before using them.

There are several approaches that may be helpful in assessing the uncertainty of activity data in particular circumstances:

***Activity data based on complete samples* (*censuses*):** Census data are activity data that are based, in principle, on counting every instance of a particular activity. Census typically includes both systematic and random errors. Systematic errors arise through systematic undercounting or double counting. Random errors are typically the sum of a range of commonplace errors. Random errors usually can be expected to be normally distributed and serially uncorrelated. Because activity data are usually collected by the same people, using the same processes, for each observation, systematic errors are likely to take approximately the same value each year. There are several approaches to identifying the potential uncertainty of activity data for complete samples. These approaches are often an integrated part of a QA/QC plan:

- To check for the size of random errors, look for fluctuations over time, and differential fluctuations in series that ought to be highly correlated with the data of interest.
- To check for bias errors, cross-check the data of interest with other, related information. One might, for instance, look up and down the supply chain for fuels, comparing coal production, coal import/export, and reported consumption. Or, one might study activities for which data are collected independently but which ought to be highly correlated with the data of interest, for instance reported fuel input vs. electricity output. One might also look at activity data of different frequencies (e.g., monthly, annual), if they are collected using different approaches.
- Interpretation of statistical differences, within, for instance, national energy data are an example of cross-checking. The comparison between energy-related carbon dioxide emissions derived from the IPCC reference approach is a formal cross-check with emissions estimates derived from other sources.

Census-based activity data are often 'precise but inaccurate' in the taxonomy shown in Figure 3.2, the random errors are small, but there may be larger bias errors. Cross-checking can suggest upper and lower bounds for possible bias errors, and sometimes will permit an actual estimate of the bias error. A possible bias error lurking within these bounds may often be characterised as a truncated uniform distribution: cross-checking shows that the unobservable true value must lie within a particular range, but there may be no reason to think any point within that range is more or less likely. However, because the bias errors in activity data are likely to be highly correlated, the difference between the reported value and the unknown true value is likely to be about the same every year, and this characteristic should be taken into account when estimating trend uncertainty.

***Activity data based on random samples*:** Some kinds of activity data are derived from sample surveys, for instance consumer surveys, land use surveys, or forest cover surveys. In these cases, the data will be subject to sampling errors, that are normally distributed. The agency conducting the sample will normally be able to advise on sampling error. If this information is unavailable, it may be possible to identify or infer the sample and population sizes and calculate sampling error directly.

The most common survey designs are simple random sampling, systematic sampling, stratified sampling, and two-stage sampling. For a simple random sampling design, a sample of n elements are selected without replacement from a population of N total elements with equal probability. For example, a survey may sample the fuel usage from 2,000,000 vehicles in a country with 80,000,000 total vehicles by randomly selecting vehicles to be included in the sample. Each sampled vehicle is multiplied by a weight of 4 (i.e., total number of vehicles divided by the number that are sampled) and summed to estimate the total fuel usage. This design is commonly used when there is little additional information known about the population.

With systematic sampling, an initial sample element is randomly selected then subsequent sampling elements are selected at equal increments, such as geographic distances apart. For example, a survey may be determining the

amount of biomass C in forestlands by sampling 50 forest stands from a population 1000 stands in a country. A random location is selected for the first sample, and then additional samples are spaced at 20 km apart across all of the forestland in a country. The biomass C for each forest is multiplied by a weight of 20 in this example (total number of forest stands divided by the number in the sample), and then summed to obtain the total biomass C for forestlands in the country. Systematic sampling is used to ensure a wide dispersion of samples in a geographical region.

Stratified sampling designs subdivide population into separate groups, referred to as strata. Individual stratum may be sampled using simple random sampling or systematic sampling. The differences among strata should be as heterogeneous as possible, whilst the subpopulation within a stratum should be as homogeneous as possible. For example, farms may be sampled to determine the amount of livestock manure N production by stratifying the farms according to the production systems in a country. If there are 15 production systems, the surveyor may have funds to sample 100 farms in each production system for a total of 1500 farms. If 10 farms are sampled in each production system, then the total amount of manure N production is estimated by multiplying each farm's value by a weight of 10 (total number of farms in a stratum divided by the number in the sample). The national total is the sum of the manure N production for the 15 production systems. In addition, individual stratum can have different sample sizes, and the weight would change in this case based on the total number of farms and number sampled in each stratum.

With a two-stage sampling design, the population is first divided into primary sampling units, and each primary sampling unit is further divided into secondary sampling units. The primary sampling units are typically selected using simple random sampling, stratified or systematic sampling, while secondary sampling units within the sampled primary sampling units are typically selected using simple random sampling. Total estimates are made for each primary sampling unit, and then combined to estimate the total for the entire population. For example, the amount of waste transported to landfills may be determined by creating primary sampling units based on random selection of provinces, and then municipalities within provinces are randomly selected for the secondary sampling units. The total amount of waste is determined for the individual provinces in the first step given the total number of municipalities in a province and the number of municipalities that are sampled. In the second step, the total waste production for the entire country is determined based on the total number of provinces and the number of provinces that are sampled. This type of sampling design may be the best approach for optimizing the precision of activity data with limited funding.

The variance calculations for total estimates from each survey design are given below based on Särndal *et al.* (1992).

EQUATION 3.0
VARIANCE CALCULATION FOR SIMPLE RANDOM SAMPLE DESIGN

$$\frac{N^2}{n}\left(1-\frac{n}{N}\right)\frac{\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}{n-1}$$

Where:

N = number of elements in the population;

n = number of elements in the sample *n*;

y_i = value of the i^{th} element in the sample;

$\overline{y}$ = average of elements in the population.

EQUATION 3.0A
VARIANCE CALCULATION FOR SYSTEMATIC SAMPLE DESIGN

$$\frac{N^2}{n}\left(1-\frac{1}{\frac{N}{n}}\right)\frac{\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}{n-1}$$

Where:

N = number of elements in the population;

n = number of elements in the sample;

y_i = value of the i^{th} element in the sample;

$\overline{y}$ = average of elements in the population.

EQUATION 3.0B
VARIANCE CALCULATION FOR RANDOM STRATIFIED DESIGN

$$\sum_{h=1}^{H} N\frac{2}{h}\left(\frac{1}{n_h}-\frac{1}{N_h}\right)S\frac{2}{h}$$

Where:

N_h = number of elements in the population in strata *h*;

n_h = number of elements in the sample in strata *h*;

H = total number of strata;

$S\frac{2}{h}$ = sample variance of stratum *h*.

EQUATION 3.0C
VARIANCE CALCULATION FOR TWO-STAGE DESIGN

$$\frac{N}{n}\left(N-n\right)S\frac{2}{u}+\frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\left(M_i-m_i\right)S\frac{2}{i};$$

$$S\frac{2}{u}=\frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{t}_i-n^{-1}\sum_{i=1}^{n}\hat{t}_i\right)^2;$$

$$S\frac{2}{i}=\frac{1}{m_i-1}\sum_{j=1}^{m_i}\left(y_{ij}-\overline{y}_i\right)^2.$$

Where:

N = number of primary units in the population;

M_i = number of secondary elements in the population in the i^{th} primary unit;

n = number of primary units sampled in the first stage;

m_i = number of secondary elements sampled in the 2nd stage in the i^{th} primary unit;

573 $\hat{t}_i$ = estimated total value for the i^{th} primary unit;

574 $S\frac{2}{u}$ = sample variance of the *i* estimated total values;

575 $S\frac{2}{i}$ = sample variance of the *j* elements in the i^{th} primary unit;

576 y_{ij} = value of the j^{th} element in the i^{th} primary unit;

577 $\overline{y_i}$ = average value of the *j* elements in i^{th} primary unit.

578

579 3.2.1.3 EXPERT JUDGEMENT AS A SOURCE OF INFORMATION

580 No refinement.

581 3.2.2 Techniques for quantifying uncertainties

582 No refinement.

583 3.2.3 Methods to combine uncertainties

584 [Update of Section 3.2.3 of the *2006 IPCC Guidelines*].

585 It further elaborates on the two approaches to combine uncertainties: Approach 1, simple propagation of error
586 equations, and Approach 2, Monte Carlo simulation. A tool for the implementation of Approach 1 is also included
587 as an addendum.

588 Once the uncertainties in activity data, emission factor or other parameters for a category have been determined,
589 they may be combined to provide uncertainty estimates for the category emissions. Once the uncertainties for the
590 categories have been determined, they may be combined to provide uncertainty estimates for the entire inventory
591 in any year and the uncertainty in the overall inventory trend over time.

592 Two approaches for the estimation of combined uncertainties are presented in the following sections: Approach 1
593 uses simple error propagation equations, while Approach 2 uses Monte Carlo or similar techniques. Either
594 Approach may be used for emission sources or sinks, subject to the assumptions and limitations of each Approach
595 and availability of resources.

596 Figure 3.2 flowchart shows a basic step-by-step suggestion on how the choice of approach could be made. In
597 practice, however, the options are not always straightforward.

598 Approach 1 is simpler to apply but requires assumptions that frequently are not entirely met, such as lack of
599 significant correlations among the quantities used in the inventory, uncertainties that are less than ± 30% of the
600 quantity value or uncertainties that are symmetrically distributed. Approach 2 requires more information on the
601 probability distributions of the data involved in the calculations. As such, it also involves assumptions and more
602 information on the underlying processes and its application depends on the capacity to acquire this information.
603 In turn, approach 2 may provide a more representative confidence interval for the uncertainty in the category.

604 Approach 2 will be particularly appropriate to use when uncertainties are large, their distribution are non-Gaussian
605 and algorithms are complex functions.

606 Biases should be addressed prior to applying either Approach 1 or 2, as these approaches focus on quantifying the
607 random component of the uncertainty of the inventory results where known sources of bias have been removed.

3.2.3.1 APPROACH 1: PROPAGATION OF ERROR

Approach 1 is based upon error propagation and is used to estimate uncertainty in individual categories, in the inventory as a whole, and in trends between a year of interest and a base year. The key assumptions, requirements, and procedures are described here.

Approach 1 should be implemented using Table 3.1, Approach 1 Uncertainty Calculation. A tool set up on a commercial spreadsheet software is provided, as an addendum to this chapter, to facilitate the implementation of Table 3.1. The table is completed at the category level using uncertainty ranges for activity data and emission factors consistent with the sectoral *good practice guidance*[1]. Different gases should be entered separately as CO_2 equivalents.

KEY ASSUMPTIONS OF APPROACH 1

In Approach 1 uncertainty in emissions or removals can be propagated from uncertainties in the activity data, emission factor and other estimation parameters through the error propagation equation (Mandel, 1984, Bevington and Robinson, 1992). If correlations exist, then either the correlation can be included explicitly or data can be aggregated to an appropriate level such that correlations become less important. Approach 1 also theoretically requires that the standard deviation divided by the mean value is less than 0.3. In practice, however, the approach will give informative results even if this criterion is not strictly met and some correlations remain. Approach 1 assumes that the relative ranges of uncertainty in the emission and activity factors are the same in the base year and in year *t*. This assumption is often correct or approximately correct. If any of the key assumptions of Approach 1 do not apply, then either an alternative version of Approach 1 can be developed (e.g., see Section 3.4) or Approach 2 can be used instead.

Where the standard deviation divided by the mean is greater than 0.3 the reliability of Approach 1 can be improved. The section 'Dealing with Large and Asymmetric Uncertainties in the Results of Approach 1' in this section describes how to do this.

KEY REQUIREMENTS OF APPROACH 1

In order to quantify uncertainty using Approach 1, estimates of the uncertainty for each input are required, as well as the equation through which all inputs are combined to estimate an output. The simplest equations include statistically independent (uncorrelated) inputs. When inputs are known to be fully (or mostly) correlated, modified equations should be used or a preliminary step should be performed to combine these inputs before the application of the basic rules.

Uncertainty of the inputs will represent a 95 percent confidence interval expressed as a percentage of the point estimate of the input (e.g. ± 20%). When the probability distribution function is known to be asymmetrical, upper and lower limits of the confidence interval need to be specified separately (e.g., -10%, +20%). In this case, approach 1 will provide only a rough approximation and in order to be used the interval needs to be replaced by a symmetrical interval built using the larger of the two quantities (e.g. ± 20%). When uncertainties are known to be large and asymmetrical, more elaborated techniques may be applied as described in Section 3.7.3 of the *2006 IPCC Guidelines.*

PROCEDURE OF APPROACH 1

The Approach 1 analysis estimates uncertainties by using the error propagation equation in two steps. First, the Equation 3.1 approximation is used to combine emission factor, activity data and other estimation parameter ranges by category and greenhouse gas. Second, the Equation 3.2 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.

Uncertainty of an Annual Estimate

[1] Where estimates are derived from models, enter the uncertainty associated with the activity data used to drive the model, and enter the uncertainty associated with the model parameters instead of the emission factor uncertainty. It may be necessary to use expert judgement, or error propagation calculations associated with the model structure. If it is impractical to separate the uncertainty estimate obtained from a model for a category into separate activity and emission factor components, then enter the total uncertainty for the category in the emission factor column and assign zero uncertainty to the activity factor column.

The error propagation equation[2] yields two convenient rules for combining uncorrelated uncertainties under addition and multiplication:

Where uncertain quantities are to be combined by multiplication a simple equation (Equation 3.1) can then be derived for the uncertainty of the product, expressed in percentage terms[3]. This rule is approximate for all random variables. Under typical circumstances, this rule is reasonably accurate as long as the percentage uncertainty is less than approximately 30%. This rule is not applicable to division.

EQUATION 3.1
COMBINING UNCERTAINTIES – APPROACH 1 – MULTIPLICATION

$$U_{total} = \sqrt{U\frac{2}{1} + U\frac{2}{2} + ... + U\frac{2}{n}}$$

Where:

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

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

Where uncertain quantities are to be combined by addition or subtraction, a simple equation (Equation 3.2) can be derived for the uncertainty of the sum, expressed in percentage terms. This rule is exact for uncorrelated variables.

EQUATION 3.2
COMBINING UNCERTAINTIES – APPROACH 1 – ADDITION AND SUBTRACTION

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

Where:

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

x_i = quantities to be added;

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

The GHG Inventory is principally the sum of products of emission factors, activity data and other estimation parameters. Therefore, Equations 3.1 and 3.2 can be used repeatedly to estimate the uncertainty of the total inventory. In practice, uncertainties found in inventory categories vary from a few percent to orders of magnitude and may be correlated. This is not consistent with the assumptions of Equations 3.1 and 3.2 that the variables are uncorrelated, and with the assumption of Equation 3.2 that the coefficient of variation is less than about 30 percent, but under these circumstances, Equations 3.1 and 3.2 may still be used to obtain an approximate result.

Applying approach 1 (level) in practice

Simple methods for estimation of the emissions of a category are usually based on the multiplication of activity data (AD) by an emission factor (EF). In many cases, it will be a reasonable assumption that these values are uncorrelated. The uncertainty associated with the emissions can then be calculated by Equation 3.2a:

[2] As discussed more extensively in Annex 1 of the *Good Practice Guidance and Uncertainty Management* (*GPG2000*, IPCC, 2000), and in Annex I of the *Revised 1996 IPCC Guidelines* (Reporting Instructions) (*1996 IPCC Guidelines*, IPCC, 1997).

[3] The option for expressing uncertainties in percent terms allows the results to be presented in a user-friendly way. However, caution should be exercised in the interpretation of the results in cases where the point estimate is very small when compared with the size of the confidence interval (e.g. a sector or inventory where removals and emissions are of similar sizes).

EQUATION 3.2A
COMBINING UNCERTAINTIES – APPROACH 1 – AD • EF

$$U_{emissions} = \sqrt{U^2_{AD} + U^2_{EF}}$$

More generally, both AD and EF can be result of several different parameters and this frequently occurs for the EF (e.g. EF = a • b • c). The uncertainty of the EF will be calculated as:

EQUATION 3.2B
COMBINING UNCERTAINTIES – APPROACH 1 – EF = A • B • C

$$U_{EF} = \sqrt{U^2_a + U^2_b + U^2_c}$$

The uncertainties associated with the emissions for each subcategory will be combined to obtain the uncertainty associated with a whole category and further combined to obtain the uncertainty of the whole inventory. In these steps the uncertainties as the quantities are combined through addition, Equation 3.2 should be applied.

Particular attention should be given to the correlation in this step. The subcategories can be highly correlated, because either the ADs are derived from the same source or the EFs have parameters in common. A special situation occurs when an input is entirely dependent on a set of other inputs. As noted in the *2006 IPCC Guidelines* this could occur, for example, if residential fuel is estimated as the difference between total consumption and usage in the transportation, industrial, and commercial sectors. Similarly, in the AFOLU sector, when land transitions are assessed, total area transitions depend on the total area of the country, resulting in less degrees of freedom for the variables.

Approach 1 has limitations to the consideration of correlation as it only allows for full correlation or independency between the variables. Still broad sensibility can be implemented, either for correlation between variables in the same year or different years. This flexibility is included in the tool described in section 3.6.2. It is important to note that in the case of full correlation among categories, aggregation of these categories is the recommended procedure. When information is lacking for either uncertainties of AD or EF for subcategories of a category, pre-processing by expert judgement may be necessary to either provide individual values to the subcategories or recommend their aggregation. Where partial correlations are known to exist and are relevant, approach 2 is recommended.

BOX 3.3

EXAMPLE OF UNCERTAINTY CALCULATION: CH_4 EMISSIONS FROM MANURE MANAGEMENT

In accordance with the Tier 1A methodology described in Chapter 10 (section 10.4) of this methodology report CH_4 emissions from manure management are estimate applying the equation below:

$$CH_{4(mm)} = \left[\sum_{T,S} \left(N_{(T)} \bullet VS_{(T)} \bullet AWMS_{(T,S)} \right) \bullet EF_{(T,S)} / 1000 \right]$$

Where:

$CH_{4(mm)}$ = CH_4 emissions from Manure Management in the country, kg CH_4 yr^{-1};

$N_{(T)}$ = number of head of livestock species/category *T* in the country;

$VS_{(T)}$ = annual average VS excretion per head of species/category *T*, kg VS $animal^{-1}$ yr^{-1};

$AWMS_{(T,S)}$ = fraction of total annual VS for each livestock species/category *T* that is managed in manure management system *S* in the country, dimensionless;

$EF_{(T,S)}$ = emission factor for direct CH_4 emissions from manure management system *S*, by animal species/category in the country, g CH_4 kg VS^{-1} in manure management system *S*.

In addition, $VS_{(T)}$ is evaluated by the equation:

$$VS_{(T)} = VS_{rate(T)} \bullet \frac{TAM}{1000} \bullet 365$$

Where:

$VS_{rate(T)}$ = default VS excretion rate, kg VS (1000 kg animal mass)$^{-1}$ day^{-1};

$TAM_{(T)}$ = typical animal mass for livestock category *T*, kg $animal^{-1}$.

If the choice is to apply a Tier 2 methodology both parameters VS and EF are evaluated through the equations:

$$VS = \left[GE \bullet \left(1 - \frac{DC}{100} \right) + \left(UE \bullet GE \right) \right] \bullet \left[\left(\frac{1 - ASH}{18.45} \right) \right]$$

Where:

GE = gross energy intake, MJ day^{-1};

DC% = digestibility of the feed in percent (e.g. 60%);

(UE • GE) = urinary energy expressed as fraction of GE.

ASH = the ash content of manure calculated as a fraction of the dry matter feed intake (e.g., 0.08 for cattle). Use country-specific values where available.

$$EF_{(T)} = \left(VS_T \bullet 365 \right) \left[B_{O(T)} \bullet 0.67 \bullet \sum_{S,k} \frac{MCF_{S,k}}{100} \bullet AWMS_{(T,S,k)} \right]$$

Where:

$EF_{(T)}$ = annual CH_4 emission factor for livestock category *T*, g CH_4 kg VS^{-1};

$B_{o(T)}$ = maximum methane producing capacity for manure produced by livestock category *T*, m^3 CH_4 kg^{-1} of VS excreted;

$MCF_{(S,k)}$ = methane conversion factors for each manure management system *S* by climate region *k*, %.

Essentially, by these equations, the CH_4 emissions are estimated by a sum of products of parameters and, as such, Equations 3.1 and 3.2 apply and can be successively used, always under usual assumptions. The parameters may be classified as AD or EF, although this is not really necessary and sometimes artificial.

In order to estimate the uncertainty, a point estimate and a confidence interval are necessary for each of the parameters. Note that for some of them this may be a complex task as the equations for calculation are not always linear. This is the case for gross energy intake (GE). In this situation, application of the approach 2 for the estimation of the confidence interval of the parameter is recommended.

As an example, the formulas are applied for the Tier 1 method for calculation of methane emissions from manure management from dairy cows. Data are from Volume 4, Chapter 10 and (Monni *et al*., 2007). Three types of manure management systems (pasture, slurry and solid storage) are considered.

$$CH_4 = \left[\sum_{i=1}^{3} \left(N_d \bullet VS_d \bullet AWMS_{(d,i)} \bullet EF_{d,i} \right) / 1000 \right]$$

Data:	N_{dairy}	=350 000	(-3%, +3%)
	$V_{Srate,dairy}$	= 7.1 kg/t animal mass/day	(-20%,+20%)
	TAM_{dairy}	= 570 kg	(-4%,+4%)
	$EF_{dairy,pasture}$	= 0.60 g CH_4/kg VS	(-30%,+30%)
	$EF_{dairy,\ slurry}$	= 34 g CH_4/kg VS	(-30%,+30%)
	$EF_{dairy,\ solid}$	= 3.2 g CH_4/kg VS	(-30%,+30%)
	$AWMS_{dairy,pasture}$	= 0.28	(-20%,+20%)
	$AWMS_{dairy,\ slurry}$	= 0.25	(-20%,+20%)
	$AWMS_{dairy,\ solid}$	= 0.47	(-20%,+20%)

It is important to note that $AWMS_i$ are not independent quantities, as $AWMS_1$+$AWMS_2$ +$AWMS_3$=1. This is an example where the variables have one less degree of freedom. Before calculating the uncertainty, $AWMS_{dairy,3}$ would need to be replaced by (1-$AWMS_{dairy,1}$-$AWMS_{dairy,2}$).

However, the terms of the resulting equation will not be all independent and this contradicts the assumptions behind Equations 3.1 and 3.2. To correctly consider the correlation between the values of $AWMS_i$, Approach 2 is recommended to be used.

Aware of the implying approximation, the results of application of Approach 1 are shown below:

Point estimates for CH_4:

$CH_{4,pasture}$ = 0.09 Gg $CH_{4,slurry}$ = 4.39 Gg $CH_{4,solid}$ = 0,78 Gg

$CH_{4,Total}$ = 5,26 Gg

Recalling that:

$$CH_{4,\ pasture} = \left[N_d \bullet VS_{rate,d} \bullet TAM_d \bullet AWMS_{(d,\ pasture)} \bullet EF_{d,\ pasture} \bullet 365 / 10^6 \right]$$

$$U(CH_{4,\ pasture}) = \sqrt{U^2{}_{N_d} + U^2{}_{VSrate,d} + U^2{}_{TAM_d} + U^2{}_{AWMS_d,\ pasture} + U^2{}_{EF_d,\ pasture}}$$

In the example:

$$U\left(CH_{4,\ pasture}\right) = \sqrt{9 + 400 + 16 + 400 + 900} = 41.5\%$$

Similarly: $U\left(CH_{4,slurry}\right) = U\left(CH_{4,solid}\right) = 41.5\%$

And then: $U_{CH_4} = \frac{\sqrt{\left(41.5 \bullet 0.09\right)^2 + \left(41.5 \bullet 4.39\right)^2 + \left(41.5 \bullet 0.78\right)^2}}{5.26} = 35.25\%$

In order to compare this result with the result of Approach 2 two cases of Monte Carlo simulation have been developed, assuming normal distribution for all parameters. In the first one the correlation between the share of systems (AWMS) was disregarded. In the second one the correlation between the systems was taken into consideration. The results obtained were:

Case without correlations: U_{MC} = 37.03 Case with correlations: U_{MC2} = 36.24

The results show that if correlation is disregarded the uncertainty result is higher than when the correlation is considered. The results also show that, in this example, the Approach 1 underestimates the uncertainty.

However, it is interesting to note that, although the result of Approach 2 will be more accurate than the result of Approach 1, the result of Approach 1 is not too far apart from the result of Approach 2. Therefore, Approach 1 can be still qualified as a tool for QA/QC and for directing priorities of improving the inventory if there are not enough data and resources for using Approach 2.

Uncertainty in the Trend

The trend of the net emissions of a category is expressed as a percentage calculated in relation to the emissions in the base year. The uncertainty in the trend will be a function of the uncertainties of the emissions in both the base year and the current year. As a direct consequence, the uncertainty of the trend will be a function of the uncertainties of the activity data and the emission factors at both these points in time.

Similar to the level uncertainty, Approach 1 for the trend uncertainty applies a simple propagation method based on the uncertainties of the input data (activity data and emission factors) for both the base year and the current year. In addition to the assumptions already described, the approach for calculating the trend uncertainty requires assumptions on data correlation between the base year and the current year.

In general, emission factors (and other estimation parameters) uncertainties will tend to be correlated between years while activity data will tend to be uncorrelated between years. The basic approach presented assumes full correlation between emission factors in the base year and the current year and independence between activity data in the base year and the current year. The method allows for change in case the activity data for a category is full correlated between years or emission factor for a category is independent between years reflecting national circumstances. However, as for the level approach, the method does not provide for partial correlations.

The uncertainty in the trend in total emissions from the country is estimated as:

EQUATION 3.2C
APPROACH 1 - TREND UNCERTAINTY

$$U_T = \sqrt{\sum_i \left(U^2_{Te,i} + U^2_{Ta,i}\right)}$$

Where:

U_T = uncertainty in the trend in total emissions from the country;

$U_{Te,i}$ = trend uncertainty introduced by the uncertainty associated with the emission factor of the category/gas *i*;

$U_{Ta,i}$ = trend uncertainty introduced by the uncertainty associated with the activity data of the category/gas *i*.

It is important to note that while the level uncertainty is reported as a confidence interval expressed as percentage uncertainties in relation to the point estimate, the uncertainty of the trend is reported as a confidence interval expressed in percentage points to be added or subtracted to the trend estimation.

In order to know how the uncertainty of the emission factors and activity data affects the trend in the emissions we will need to develop type A and type B sensitivities as follows:

- *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 percent increase in emissions or removals of a given category and gas in both the base year and the current year.

EQUATION 3.2D
CALCULATION OF TYPE A SENSITIVITY

$$A_x = \left| \frac{0.01 \bullet E_{x,t} + \sum_i E_{i,t} - \left(0.01 \bullet E_{x,BY} + \sum_i E_{i,BY} \right)}{\left(0.01 \bullet E_{x,BY} + \sum_i E_{i,BY} \right)} \bullet 100 - \frac{\sum_i E_{i,t} - \sum_i E_{i,BY}}{\sum_i E_{i,BY}} \bullet 100 \right|$$

Where:

A_x = the type A sensitivity for category/gas *x*;

$E_{i,t}$ = emissions/removals for category/gas *i* in the year *t*;

$E_{i,BY}$ = emissions/removals for category/gas *i* in the base 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 percent increase in emissions or removals of a given category and gas in the current year only.

EQUATION 3.2E
CALCULATION OF TYPE B SENSITIVITY

$$B_x = \left| \frac{E_{x,t}}{\sum_i E_{i,BY}} \right|$$

Where:

B_x = the type B sensitivity for category/gas *x*;

$E_{x,t}$ = emissions/removals for category/gas *x* in the year *t*;

$E_{i,BY}$ = emissions/removals for category/gas *i* in the base year.

Under the assumption that the emission factors are fully correlated, to a variation in the base year emission factor will correspond the same variation in the current year emission factor. Hence, the emission factor uncertainty will be propagated to the trend through a Type A sensibility.

EQUATION 3.2F
TREND UNCERTAINTY DUE TO EMISSION FACTOR

$$U_{Te,i} = A_i \bullet U_{EF,i}$$

Where:

$U_{Te,i}$ = trend uncertainty introduced by the uncertainty associated with the emission factor of the category/gas *i*;

A_i = the type A sensitivity for category/gas *i*;

$U_{EF,i}$ = uncertainty of the emission factor for category/gas *i*.

Under the assumption that the activity data in the base year and the current year are independent both the uncertainties have to be taken into consideration. Hence, the activity data uncertainty will be propagated to the trend through a Type B sensibility that shows the sensitivity to a random uncertainty error in the emissions estimate. The additional factor of √2 is introduced because an uncorrelated uncertainly might affect either the base year or the current year.

EQUATION 3.2G
TREND UNCERTAINTY DUE TO ACTIVITY DATA

$$U_{Ta,i} = B_i \bullet U_{AD,i} \bullet \sqrt{2}$$

Where:

$U_{Ta,i}$ = trend uncertainty introduced by the uncertainty associated with the activity data of the category/gas *i*;

B_i = the type B sensitivity for category/gas *i*;

$U_{AD,i}$ = uncertainty of the activity data for category/gas *i*.

Worksheet for Approach 1 Uncertainty Calculation

The columns of Table 3.1, Approach 1 Uncertainty Calculation, are labelled A to Q and contain the following information, of which the derivation of key equations is given in Section 3.7.1 in Section 3.7, Technical Background Information.

- A shows the sector of the IPCC category.
- B shows the code of the IPCC category.
- C shows the name of the IPCC category.
- D shows the greenhouse gas.
- E and F are the inventory estimates in the base year and the current year[4] respectively, for the category and gas specified in Columns C and D, expressed in CO_2 equivalents.
- G and I 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 percent confidence interval divided by the mean and expressed as a percentage. The reason for halving the 95 percent confidence interval is that the value entered in Columns G and I corresponds to the familiar plus or minus value when uncertainties are loosely quoted as 'plus or minus *x* percent', so expert judgements of this type

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

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.

- H indicates if the uncertainty in activity data is correlated across years
- J indicates if the uncertainty in emission factor is correlated across years
- K is the combined uncertainty by category derived from the data in Columns G and I using the error propagation equation (Equation 3.2). The entry in Column K is therefore the square root of the sum of the squares of the entries in Columns G and I.
- L shows the uncertainty in Column K as a percentage of total national emissions in the current year. The entry in each row of Column L is the square of the entry in Column K multiplied by the square of the entry in Column F, divided by the square of total at the foot of Column F. The value at the foot of Column L is an estimate of the percentage uncertainty in total national net emissions in the current year, calculated from the entries above using Equation 3.1. This total is obtained by summing the entries in Column L and taking the square root.
- M shows how the percentage difference in emissions between the base year and the current year changes in response to a one percent increase in category emissions/removals for both the base year and the current year. This shows the sensitivity of the trend in emissions to a systematic uncertainty in the estimate (i.e., one that is correlated between the base year and the current year). This is the Type A sensitivity as defined above.
- N shows how the percentage difference in emissions between the base year and the current year changes in response to a one percent increase in category emissions/removals in the current year only. This shows the sensitivity of the trend in emissions to random error in the 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.
- O shows the uncertainty introduced into the trend in emissions by emission factor uncertainty. If the uncertainty in emission factors is correlated between years (J = Y) the result is the product of the information in Columns M and I. If the emission factor uncertainties are not correlated between years (J = N) then the entry in Column N should be used in place of that in Column M and the result multiplied by $\sqrt{2}$.
- P shows the uncertainty introduced into the trend in emissions by activity data uncertainty. If the uncertainty in activity data is not correlated between years (H = N) the result is the product of the information in Columns N and G multiplied by $\sqrt{2}$. If the activity data uncertainties are correlated between years (H = Y) then the entry in Column M should be used in place of that in Column N and the $\sqrt{2}$ factor does not then apply.
- Q is an estimate of the uncertainty introduced into the trend in national emissions by the category in question. Under Approach 1, this is derived from the data in Columns O and P using Equation 3.2. The entry in Column Q is therefore the sum of the squares of the entries in Columns O and P. 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 entries in Column Q and taking the square root. The uncertainty in the trend is a *percentage point* range relative to the inventory trend. For example, if the current year emissions are 10 percent greater than the base year emissions, and if the trend uncertainty at the foot of Column Q is reported as 5 percent, then the trend uncertainty is 10%±5% (or from 5% to 15% increase) for the current year emissions relative to the base year emissions.
- Explanatory footnotes go at the bottom of the table and give documentary references of uncertainty data (including measured data) or other relevant comments.

An example of the spreadsheet with all the numerical data completed is provided in Section 3.6, Approach 1 uncertainty calculation example. Details of this approach are given in Section 3.7.1 and derivation of the uncertainty in the trend is in Section 3.7.2.

945

TABLE 3.2
APPROACH 1 UNCERTAINTY CALCULATION

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Inventory sector	IPCC category code	IPCC category name	GHG	Base year emissions or removals	Year *t* emissions or removals	Activity data uncertainty	AD uncertainty correlated across years?	Emission factor / estimation parameter uncertainty	EF uncertainty correlated across years?	Combined uncertainty	Contribution to Variance by Category in Year *t*	Type A sensitivity	Type B sensitivity	Uncertainty in trend in national emissions introduced by emission factor / estimation parameter 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 Note A	Input data Default: N	Input data Note A	Input data Default: Y	$\sqrt{G^2 + I^2}$	$\frac{(K \bullet F)^2}{(\sum F)^2}$	Note B	$\left\|\frac{F}{\sum E}\right\|$	If J = Y $M \bullet I$ If J = N $N \bullet I \bullet \sqrt{2}$	If H = N $N \bullet G \bullet \sqrt{2}$ If H = Y $M \bullet G$	$O^2 + P^2$
				Gg CO_2 equivalent	Gg CO_2 equivalent	%	Y/N	%	Y/N	%		%	%	%	%	%
e.g. Energy	e.g. 1.A.1	e.g. Energy Industries Fuel 1	CO_2													
e.g.	e.g. 1.A.1	e.g. Energy Industries Fuel 2	CO_2													
Etc...	Etc.	Etc...	…													
Total				$\sum E$	$\sum F$						$\sum L$					$\sum Q$
										Percentage uncertainty in total inventory:	$\sqrt{\sum L}$				Trend uncertainty:	$\sqrt{\sum Q}$

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

- If uncertainty is correlated across years, enter the uncertainty into Column I, and enter 0 in Column G; it is suggested to assume correlation across years if some of the parameters used in the estimates are the same in both years or derived from the same source.
- If uncertainty is not correlated across years, enter the uncertainty into Column G, and enter 0 in Column I; it is suggested to assume no correlation between years if the estimates for the two years are independent from each other, for example based on independent measurements.

Note B: Absolute value of:

$$\frac{0.01 \bullet F_x + \sum F_i - \left(0.01 \bullet E_x + \sum E_i\right)}{0.01 \bullet E_x + \sum E_i} \bullet 100 - \frac{\sum F_i - \sum E_i}{\sum E_i} \bullet 100$$

Where:

E_x, F_x = entry from row *x* of the table from the corresponding column, representing a specific category;

$\sum E_i, \sum F_i$ = sum over all categories (rows) of the inventory of the corresponding column.

DEALING WITH LARGE AND ASSYMMETRIC UNCERTAINTIES

No refinement.

3.2.3.2 APPROACH 2: MONTE CARLO SIMULATION

No refinement.

3.2.3.3 HYBRID COMBINATIONS OF APPROACHES 1 AND 2

No refinement.

3.2.3.4 COMPARISON BETWEEN APPROACHES

No refinement.

3.2.3.5 GUIDANCE ON CHOICE OF APPROACH

No refinement.

3.3 UNCERTAINTY AND TEMPORAL AUTOCORRELATION

No refinement.

3.4 USE OF OTHER APPROPRIATE TECHNIQUES

No refinement.

3.5 REPORTING AND DOCUMENTATION

No refinement.

3.6 EXAMPLES

[Elaboration of section 3.6 of the *2006 IPCC Guidelines*].

Chapter 3 of Volume 1 of the *2006 IPCC Guidelines* includes two examples of uncertainty estimates for inventories, both based upon the Finnish 2003 greenhouse gas inventory.

These examples illustrate that the results from Approaches 1 and 2 can be very similar when the overall uncertainty is relatively small. However, Approach 2 is a more flexible approach that enables quantification of asymmetry in probability ranges, such as for the year *t* inventory.

Step-by-step example for Approach 2 based on the Italian GHG Inventory (CH_4 emissions from enteric fermentation in the Agriculture sector) is provided below. This example focusses on the process of obtaining the data for all parameters involved and the analysis of results. CH_4 emissions are estimated by a Tier 2 approach.

Step 1

A list of selected parameters used in the CH_4 emission factors estimation process is indicated below. For each parameter, the choices of distributions and underlying assumptions are described and whether they can be modeled by Monte Carlo.

Parameter	Description	MCA	Range	Source
Animal number	Average annual population within a country by animal species (include all livestock categories)	Yes	The uncertainty associated with populations vary depending on the source, but should be within +-20%. The National Institute of Statistics has estimated an uncertainty of 5-6% associated with data. Expert judgment (ISPRA) assumed 10% uncertainty	IPCC, 2006; National Institute of Statistics; ISPRA
Milk production	Total average annual milk production (dairy and buffalo)	Yes	Expert judgment, assuming the same value as for animal number (10%)	ISPRA
Methane conversion factor (Y_m)	Y_m is the fraction of gross energy in feed converted to methane (dairy cattle and buffalo)	Yes	IPCC expert group judgment assumed for dairy cattle and buffalo a conversion factor equal to 6.5%$\pm$1%	IPCC, 2006
Weight	Live-weight data should be collected for each animal sub-category, and the data should be based on weight measurements of live animals (dairy cattle, buffalo, and non-dairy)	Yes	Expert judgment, assuming the same value as for animal number (10%)	ISPRA
% animal grazing	Animals graze open range land or hilly terrain and expend significant energy to acquire feed (dairy cattle and buffalo)	Yes	Expert judgment; for emission estimates 10% of grazing animals were assumed while based on actual statistics it has been calculated to be around 5% (uncertainty 50%)	ISPRA
Fat content	Average fat content of milk is required for dairy cattle and buffalo	Yes	Expert judgment, assuming the same value as for animal number (10%)	ISPRA

% giving birth	Percent of females that give birth in a year for dairy cattle and buffalo	Yes	Expert judgment, assumed 5%	ISPRA
Feed digestibility (DE)	The proportion of energy in the feed not excreted in the feces is known as feed digestibility, expressed as a percentage (dairy cattle and buffalo)	Yes	Default 12 to 20%. Expert judgement 18%	IPCC, 2006 ISPRA
EF for Tier 1 approach	The EF is assumed for an animal category for an entire year (365 days): Swine (sows and other swine), sheep, goats, horses, mules and asses, rabbits	Yes	All estimates have an uncertainty of +30-50%. EFs estimated using the Tier 1 method are unlikely to be known more accurately than +30% and may be uncertain to +50%. Assumed 50%.	IPCC, 2006 ISPRA
Dry matter intake (DMI)	DMI establishes the amount of nutrients available to an animal for health and production. Important for the formulation of diets	Yes	The same error in estimating DE between 12-20%. Assumed 20%	IPCC, 2006 ISPRA
Coefficient for NE_m (CF_i)	Coefficient for calculating NE_m	No		
Weight gain (kg/d)	Average weight gain (or loss) per day, kg/d (for cattle and buffalo)	No		
NE_m	= net energy required by the animal for maintenance (Equation 4.1 IPCC 2000), MJ/day	No		
NE_a	= net energy for animal activity (Equations 4.2a and 4.2b IPCC 2000), MJ/day	No		
NE_g	= net energy needed for growth (Equations 4.3a and 4.3b IPCC 2000), MJ/day	No		
NE_l	= net energy for lactation, MJ/day	No		
Gross energy (GE)	Amount of energy (MJ/day) an animal needs to perform activities such as growth, lactation, and pregnancy	No		
….				
…..				

For each parameter, the choice of distribution and distribution parameters (mean, median, range etc.) is based on actual information if available (literature, distribution of measurements, past data information) or/and expert judgment. The shape of distribution may vary from the classical normal or lognormal distributions to more sophisticated ones. Whenever assumptions or constraints on variables are known, this information is reflected on the choice of type and shape of distributions (e.g. variability, asymmetry and multimodal).

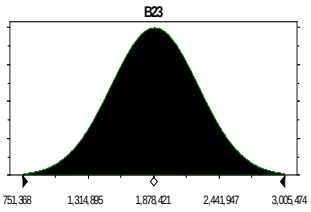
Examples of selected distributions for some parameters are shown in the following figure.

Assumption: number of dairy cattle

Normal distribution with parameters:

Mean	1,878,421
Standard Dev.	375,684

Selected range is from -Infinity to +Infinity

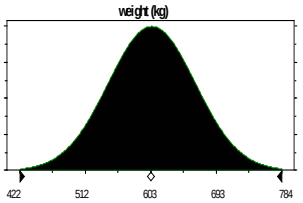


Assumption: weight (kg)

Normal distribution with parameters:

Mean	603
Standard Dev.	60

Selected range is from 0 to +Infinity

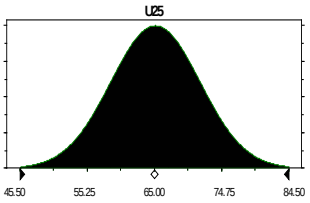


Assumption: digestibility of feed

Normal distribution with parameters:

Mean	65
Standard Dev.	6.5

Selected range is from 0 to 84.44

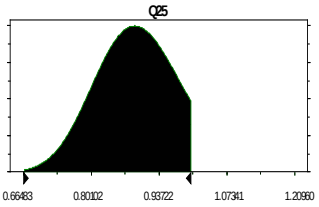


Assumption: portion of cows giving birth

Lognormal distribution with parameters:

Mean	0.90123
Standard Dev.	0.09012

Selected range is from 0.00000 to 1.00000



Step 2

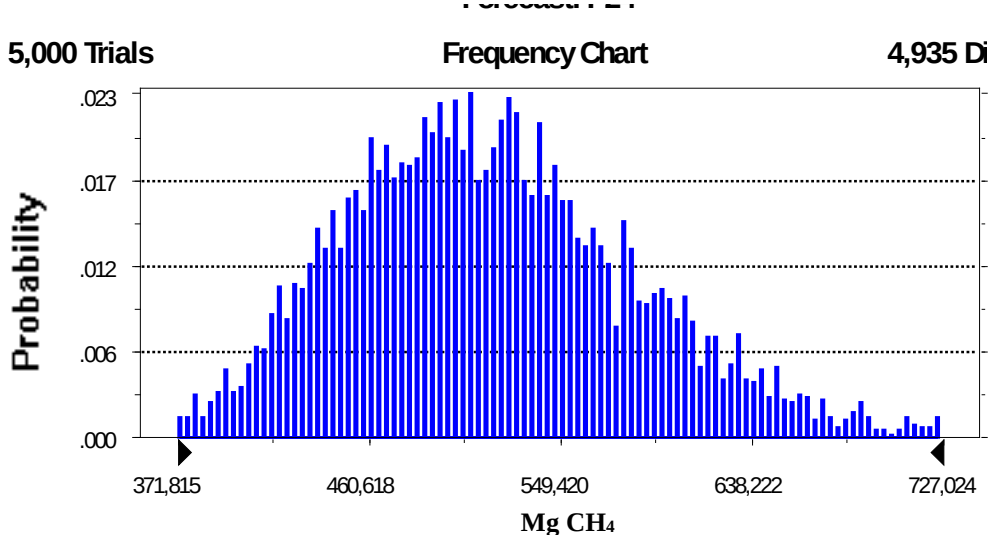
A description of the statistics resulting from the Monte Carlo analysis is reported.

Statistics of the Monte Carlo analysis for CH_4 emissions from enteric fermentation, year 2009

	Value
Trials	5000
Mean	519,226
Median	512,480
Standard Deviation	71,264
Range Minimum	340,639
Range Maximum	869,092
Uncertainty (%)	**-21.8; +31.7**

The application of Approach 1 to this category results in an uncertainty equal to 20.2%.

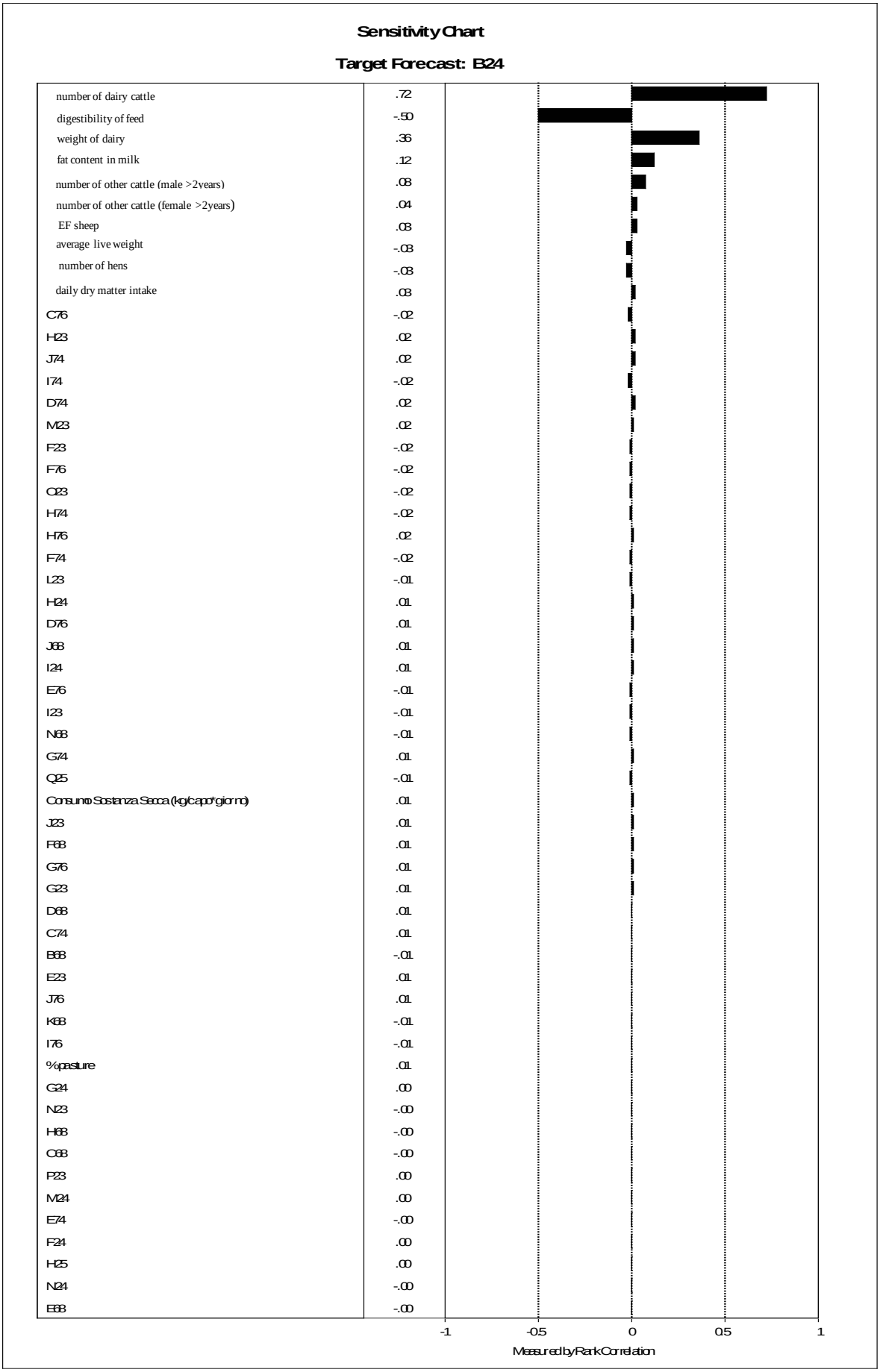
The probability density function resulting from the Monte Carlo assessment is shown in the following figure.



Step 3

The most relevant parameters for the uncertainty of CH_4 emissions from enteric fermentation, measured by the rank correlation coefficient have been individuated from the application of Monte Carlo. These are the number of dairy cattle, digestibility and the weight of animals. As far as feasible, it is important to reduce the associated uncertainty.

The results of this analysis are shown in the following chart.


Sensitivity Chart
Target Forecast: B24
number of dairy cattle .72
digestibility of feed -.50
weight of dairy .36
fat content in milk .12
number of other cattle (male >2years) .08
number of other cattle (female >2years) .04
EF sheep .03
average live weight -.03
number of hens -.03
daily dry matter intake .03
C76 -.02
H23 .02
J74 .02
I74 -.02
D74 .02
M23 .02
F23 -.02
F76 -.02
C23 -.02
H74 -.02
H76 .02
F74 -.02
L23 -.01
H24 .01
D76 .01
J68 .01
I24 .01
E76 -.01
I23 -.01
N68 -.01
G74 .01
Q25 -.01
Consumo Sostanza Secca (kg/capo*giorno) .01
J23 .01
F68 .01
G76 .01
G23 .01
D68 .01
C74 .01
B68 -.01
E23 .01
J76 .01
K68 -.01
I76 -.01
%pasture .01
G24 .00
N23 -.00
H68 -.00
C68 -.00
P23 .00
M24 .00
E74 -.00
F24 .00
H25 .00
N24 -.00
E68 -.00
-1
-0.5
0
0.5
1
Measured by Rank Correlation

3.7 TECHNICAL BACKGROUND INFORMATION

No refinement.

3.7.1 Approach 1 variables and equations

No refinement.

3.7.2 Approach 1 – details of the equations for trend uncertainty

No refinement.

3.7.3 Dealing with large and asymmetric uncertainties in the results of Approach 1

No refinement.

3.7.4 Methodology for calculation of the contribution to uncertainty

No refinement.

References

Bevington, P.R. and Robinson, D.K. (1992). Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill: New York.

EPA, U. (2002) Procedures manual for Quality Assurance/Quality Control and Uncertainty Analysis.

EPA, U. (2017) Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2015.

Gillenwater, M., Sussman, F., Cohen, J (2007) Practical Policy Applications of Uncertainty Analysis for National Greenhouse Gas Inventories. *Water, Air &Soil Pollution: Focus* **7**: 451-474.

Mandel, J. (1984) The Statistical Analysis of Experimental Data, Dover Publications New York, USA, p.410.

Monni, S., Perälä, P., Kristina, R. (2007) Uncertainty in Agricultural CH_4 and N_2O Emissions from Finland - Possibilities to Increase Accuracy in Emission Estimates. *Mitigation and Adaptation Strategies for Global Change*. **12**:545-571.

Ogle, S. M., Breidt, F.J., Easter, M., Williams, S., and Paustian, K. (2007) An Empirically based approach for estimating uncertainty associated with modeling carbon sequestration in soils. *Ecological Modelling* **2005**: 453-463.

Ogle, S. M., Breidt, F.J., Easter, M., Williams, S., Killian, K., and Paustian, K. (2010) Scale and Uncertainty in modeled soil organic carbon stock changes for US croplands using a process-based model. *Global Change Biology* **16**: 810-820.

Ogle, S. M., Eve, M.D, Breidt, F.J., Paustian, K. (2003) Uncertainty in estimating land use and management impacts on soil organic carbon storage for US agroecosystems between 1982 and 1987. *Global Change Biology* **9**: 1521-1542.

Parton, W.J., Schimel, D.S., Cole, C.V., Ojima, D.S. (1987). Analysis of factors controlling soil organic matter levels in Great Plains grasslands. *Soil Sci. Soc. Am. J.* **51**, 1173-1179.

Särndal, C-E, B. Swensson, J. Wretman (1992) Model assisted survey sampling. Springer Series in Statistics, Springer-Verlag, New York.