HoliSoils WP3

The text below reports the outcome of Task T3.1.d (Uncertainty estimation of GHG reporting) and is a contribution to Chapter X of Deliverable D3.2 (Soil sampling design, monitoring & measurement protocols), specifically on the topic "guideline for error calculations of GHG reporting". The task description copied from the project plan is:

T3.1.d: Uncertainty estimation of GHG reporting (ISRIC, TI)

Datasets are often contaminated by various error sources, such as measurement, analytical, sampling, classification, interpolation, and/or positional error. Errors may also be cross-correlated and correlated in space and time. Errors propagate in subsequent processing steps and through models to GHG exchange estimates. We will list major error sources for all soil properties used as input in models for GHG exchange estimates and review statistical methods for quantification and propagation of error and uncertainty. For this we will mainly rely on IPCC guidelines and common uncertainty analysis methodologies and also consult national reporting to the IPCC and UNFCCC.

Author: G.B.M. Heuvelink (ISRIC)

X Uncertainties in GHG reporting

Reports of greenhouse gas (GHG) emissions released into the atmosphere rely on both GHG measurements and models. However, neither measurements nor models are without imperfections, implying that the reported GHG emissions are mere approximations of the actual emissions and are not error-free. It is of great importance to quantify the errors in the estimation process as this determines the fitness for use of the estimates. In instances where estimation errors are substantial, the estimates should be used with caution by policy and decision makers, as well as other end-users. It may be beneficial to reduce estimation errors, and knowing the extent to which each error source contributes to GHG emission estimation error can be very useful in achieving this objective.

This chapter reviews the major sources of uncertainty of GHG emission models, describes how errors and uncertainties can be characterized statistically (i.e., by probability distributions), and explains how the propagation of uncertainty through GHG emission models can be analysed. Specific attention is paid to how error and uncertainty depend on the spatial and temporal extent over which GHG estimates are made and how uncertainty assessment is addressed in protocols used by IPCC and UNFCCC. Throughout this chapter, the focus is on uncertainty in soil properties and how these propagate through GHG models.

X.1 Sources of uncertainty

Suppose we use a model f to estimate the GHG emission y from model inputs x_i $(i = 1,\ldots,n)$ and model parameters θ_j, $j =,\ldots,m$:

$$\hat{y} = f(x_1,\ldots,x_n,\theta_1,\ldots\theta_m) \quad (1)$$

The estimate $\hat{y}$ will differ from the true GHG emission y for three reasons:

1. *Errors in model inputs*. Model input errors will propagate to the model output. For example, GHG emission depends on soil organic carbon and soil moisture, but these soil properties are rarely known exactly and often derived from soil maps. All maps have errors, and these soil map errors will propagate to the estimated GHG emission.
2. *Errors in model parameters*. Many models have parameters that are calibrated using inverse modelling approaches, but since calibration data sets are finite and have measurement errors,

the estimated parameters have errors too. These errors will also propagate to the GHG emission estimates.

3. *Errors in model structure*. Even if the inputs and parameters of a model were known without error, there would still be a GHG emission estimation error because the model structure is only a simplified representation of the 'true' physical, chemical and biological processes (Heuvelink, 1998a). During modelling, many sub-processes may have purposely been ignored or represented in a simplified way. Reality is too complex to be modelled exactly and hence models are simplified representations of reality.

In this section we take a closer look at all three error sources. But before we do that we present a statistical model of error and uncertainty.

X.1.1 Statistical model of error and uncertainty

Suppose that the organic carbon content of the soil at some location equals 25.8 g kg^{-1}. Suppose further that this value is unknown to us because we did not take a soil sample at the location and analyse it in the laboratory. All that we have is a map that states that the soil organic carbon at the location equals 31.2 g kg^{-1}. Clearly, the soil map is in error, and the error equals 25.8 – 31.2 = − 5.4 g kg^{-1}. Here, error is defined as the difference between the true and estimated value of the soil property.

In practice, we often do not know the error, because we need the true value to calculate it and we do not have the resources to perfectly measure the soil everywhere at all times. In other words, we are often uncertain about the error and the true value. Although we do not know the error and are uncertain, this does not mean that we are completely ignorant. For instance, we might know that the chances are equal that the map error is positive or negative (because we used an unbiased mapping method), we might know that it is very unlikely that the absolute value of the error is greater than 1.50 g kg^{-1}, etc. Thus, it is not unreasonable to assume that we can define a probability distribution of the error. This distribution lists all possible values of the error, and attaches a probability or probability density to each of them (Figure X.1).

The most common probability distribution is the normal distribution, which has two parameters: a mean and a standard deviation. The mean represents the systematic error, whereas the standard deviation is a measure of the random error. To give an example where measurement error is the cause of uncertainty, suppose the pH of a soil sample is measured repeatedly in a laboratory, and let the set of measured values have a mean of 7.24 and standard deviation 0.46. Let it also be known that the 'true' pH is 7.41 (note that in practice the true value may be difficult to get, but a reference value established in a certified laboratory could be asserted as such), then we may characterize the measurement error by a normal distribution with mean of 0.17 and a standard deviation of 0.46. Note that the square of the standard deviation is known as the variance.

Errors in variables that vary in space and time may be correlated in space and time. Also, errors in one variable may be correlated with those in another variable. For instance, if clay is defined as 1–sand–silt then measurement errors in sand and silt will be correlated with the estimation error of clay. It is beyond the scope of this chapter to provide detailed statistical models for all these extensions, instead we refer to Heuvelink et al. (2007).

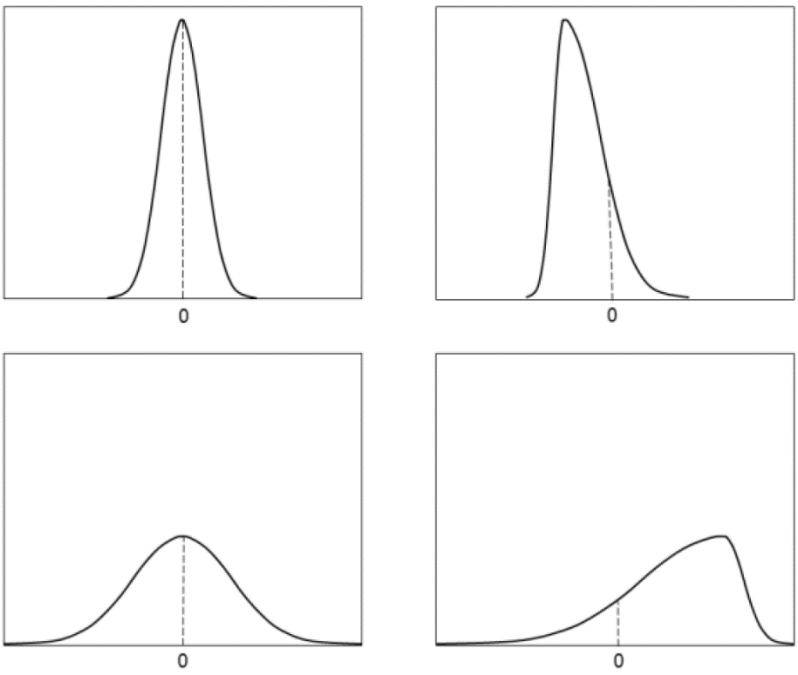

Figure X.1. Examples of probability density functions to characterise uncertainty. Density functions can be narrow (small uncertainty, top) or wide (large uncertainty, bottom). They can also be symmetric around zero (left) or asymmetric and biased (right). From Heuvelink (2014).

X.1.1 Input uncertainty

To quantify uncertainty in inputs to GHG models we first need to consider how each input was derived. If it was derived by a direct measurement in the field or laboratory then the only source of uncertainty is measurement error, which can be quantified by repeated measurements (e.g. van Leeuwen et al., 2021) or through specifications of instrument precision by the manufacturer. Note that this only assesses the random component of the measurement error, not the systematic error. Errors of field measurements that are obtained by expert judgement may be more difficult to assess, but one approach is to let experts quantify their uncertainty (O' Hagan et al., 2006) or have multiple experts independently estimate the soil property of interest (van Leeuwen et al., 2018).

In many cases models are run with inputs that are not directly measured but supplied by maps. This introduces another source of uncertainty. If maps were made using geostatistical interpolation then map interpolation errors are quantified by the kriging standard deviation (Webster and Oliver, 2007). This works for any kriging variant, be it ordinary kriging, regression kriging or lognormal kriging (note that in the latter case the map errors will be lognormally distributed, which might be more realistic than a normal distribution for skewed inputs such as soil organic carbon and precipitation). If maps were made using machine learning it is more difficult to obtain the map errors, but one technique that accomplishes this and is frequently used in digital soil mapping is known as quantile regression forests (Meinshausen, 2006). Usually one derives the 0.05 and 0.95 quantiles to obtain the lower and upper limits of a 90% prediction interval, but in principle all quantiles can be derived, which together define the prediction error probability distribution. Examples of quantifying soil map uncertainty in this way are Vaysse and Lagacherie (2017), Szatmári and Pásztor (2019) and Poggio et al. (2021). Some studies also use a bootstrapping approach to quantify uncertainty in machine learning maps, but this only captures part of the total uncertainty (i.e., a confidence interval instead of a prediction interval) and should not be used to quantify map uncertainty (Wadoux, 2019).

X.1.2 Parameter uncertainty

Model parameters differ from model inputs in a sense that they only have meaning in the context of a model. A typical example is a regression coefficient. We may fit a multiple linear regression model that predicts GHG emission from soil temperature (T) and moisture (M) as:

$$GHG = \beta_0 + \beta_1 \cdot T + \beta_2 \cdot M \tag{2}$$

Here, T and M are inputs that also exist outside the model and can be measured. But the regression coefficients β_i are no physical entities and cannot be measured directly. They can only be estimated using a calibration or inverse modelling approach, which requires paired observations of model inputs and outputs. Many calibration approaches only provide estimates of the model parameters, but our interest is also in the uncertainty of these parameter estimates.

In case of a simple model such as Eq. 2 we can derive the (multivariate) probability distribution of the regression coefficients using analytical methods. Thus, estimates of the regression coefficients, variances of associated estimation errors and their correlation coefficients can be easily derived using standard statistical software packages. For an example see Section 14.4.5 in Heuvelink (2018).

Most GHG emission models are much more complicated than a linear regression model, and in such case it is practically impossible to derive the probability distribution of model parameters in an analytical way. In such case one has to turn to numerical methods, and the most powerful and statistically rigorous method for that is Bayesian calibration (Kennedy and O' Hagan, 2001; van Oijen et al., 2005). The idea of this method is to start by defining a prior distribution for all uncertain parameters. This distribution must reflect the prior knowledge of the modeller. For instance, a modeller may know that a parameter cannot be negative or bigger than some threshold. If the modeller has little to no prior knowledge, the prior distribution may be chosen extremely wide, yielding an uninformative or so-called 'flat' prior. Next, the calibration data are used to update the prior distribution to a posterior distribution. This requires multiplication of the prior by the likelihood, which expresses how 'likely' the data (i.e., the observed model outputs in the calibration set) are, given the parameters. The key equation is:

$$p_{posterior}(\theta|data) \propto p_{prior}(\theta) \cdot p(data|\theta) \tag{3}$$

The likelihood $p(data|\theta)$ involves use of the GHG emission model, because that defines how model parameters influence the model output. Note that the likelihood is a distribution and must therefore include a stochastic element, which either derives from measurement errors in the data (i.e., measurements of GHG emission used for model calibration) or from incorporation of model structural uncertainty (see Section X.1.3 below). Note also that Eq. 3 has a proportionality sign. This obstructs analytical solutions and has led to the development of Markov chain Monte Carlo simulation (Gelman et al., 2013), which solves Eq. 3 using numerical sequential simulation. This method is computationally intensive, which in turn has led to faster approximate solutions, such as the Integrated Nested Laplace Approximation approach (INLA, Lindgren and Rue; 2015).

Bayesian calibration is complicated and requires specific skills but it is also very powerful because it yields the complete multivariate distribution of all model parameters that are treated as uncertain. Thus, it also incorporates correlations between parameter estimation errors, which is of key importance to get a realistic assessment of parameter uncertainty. Examples of Bayesian calibration in GHG emission modelling are Lehuger et al. (2009), Hashimoto et al. (2011), Lehuger et al. (2011) and Rahn et al. (2011).

X.1.3 Model structure uncertainty

Model structural uncertainty is difficult to quantify. The easiest and most common way is to add a stochastic term to the model that supposedly captures all model structural uncertainty. For instance, in the example of the multiple linear regression given in Eq. 2, we would rewrite this as:

$$GHG = \beta_0 + \beta_1 \cdot T + \beta_2 \cdot M + \varepsilon \quad (3)$$

where ε is typically assumed to be an independent and identically distributed normal random variable with zero mean and constant variance σ^2. This basically means that estimation of the model structural uncertainty is reduced to estimation of a single parameter, i.e. the variance variance σ^2. Extension to dynamic models is done in time series analysis, such as through Auto-Regressive Moving Average (ARMA) models, where model structural error is also additive but where temporal autocorrelation of these errors is incorporated (Sutthichaimethee and Kubaha, 2018; Zhao et al., 2018; Sun et al., 2019). A similar approach is used in state-space modelling, which in addition to a stochastic state equation also defines a stochastic measurement equation, thus enabling conditioning the state variables to measurements through Kalman filtering, smoothing and forecasting (i.e., data assimilation). For examples applying this method to GHG emission see Innes et al. (2015), Henne et al. (2016) and Ding et al. (2017). Incorporating spatial instead of temporal correlation in model structural uncertainty can be done using geostatistical approaches, by defining and quantifying a variogram (Webster and Oliver, 2007).

One other method to characterize model structural uncertainty is through the use of an ensemble of models (Refsgaard et al., 2007; Wagena et al., 2019, Liao et al. 2021). Here, the idea is that the set of models represent a sample from the population of all possible models, so that model uncertainty is quantified by the differences between these models and their outputs. The problem with this approach is that the set of models may not capture the entire population well (because modellers tend to learn and copy from each other and build similar models) and that it is cumbersome because multiple models must be developed and applied.

Finally, it is important to note that quantifying model parameter and model structural uncertainty cannot be done separately and should ideally be done while also accounting for input uncertainty and uncertainty in measurements of model outputs used for model calibration (Quetin et al., 2020; Wadoux et al., 2020). This is because discrepancies between model outputs and independent measurements are the result of uncertainty in model inputs, model parameters, model structure and measurement errors in the independent data and cannot be attributed to just one or a few of these uncertainty sources.

X.2 Uncertainty propagation

Once all three sources of uncertainty as discussed in Section X.1 are quantified (i.e. characterized by probability distributions) it is relatively easy to analyse the propagation of these uncertainties to the model output. Two main methods are often being used for this. The first is the Taylor series approximation method, the second the Monte Carlo simulation method.

X.2.1 Taylor series method

The Taylor series method, also known as first-order analysis, approximates the model f defined in Eq. 1 by a truncated Taylor series around the means of the uncertain inputs and parameters (Heuvelink, 1998b). In case of the first-order approximation, this boils down to a linear model that is fairly close to the true model near the centre of the probability distributions of the uncertain sources and that deteriorates further away from it. Linearizing the model greatly simplifies the uncertainty propagation analysis. It turns out that the variance of the model output can be derived analytically (Taylor, 198; Heuvelink, 1998b):

$$var(y) \cong \sum_{i=1}^{m} var(x_i)^2 \cdot \left(\frac{\partial f}{\partial x_i}\right)^2 + 2 \cdot \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} cov(x_i, x_j) \cdot \left(\frac{\partial f}{\partial x_i}\right) \cdot \left(\frac{\partial f}{\partial x_j}\right) \quad (4)$$

Here, m is the number of uncertain inputs, $\frac{\partial f}{\partial x_i}$ is the partial derivatives of f with respect to the i-th input, which is evaluated at the means of the uncertain inputs. Note that for simplicity Eq. 4 ignores uncertainty in model parameters, but that these can be treated in a similar way by extending the summation over the θ_i. Similarly, if model structure uncertainty is represented by an additive stochastic term (see Section X.1.3) then the variance of that term should simply be added to Eq. 4.

Figure X.2 shows a graphical illustration of the Taylor series method in a case where there is a single input (i.e., $m = 1$). This figure nicely shows that the output uncertainty not only depends on the magnitude of the input uncertainty, but also on the sensitivity of the model to small changes in the input (i.e., the partial derivatives in Eq. 4).

Eq. 4 shows that the uncertainty of the model output depends on the uncertainty of the inputs, but also on the sensitivity of the model output to changes in the model inputs, as expressed by the partial derivatives. If the input uncertainties are uncorrelated the covariance terms are zero and the output variance becomes a sum of terms, each of which represents the contribution of one of the uncertain inputs. This is valuable information because it shows which are the main sources of uncertainty (the weakest links in the chain) and can inform modellers where to focus attention to reduce uncertainty.

Some examples of application of the Taylor series method for uncertainty propagation in environmental modelling are Ferante et al. (1999), Grüneberg et al. 2014, Hoffmann et al. (2014), and Magnussen et al. (2014).

An important disadvantage of the Taylor series method is that it involves an approximation error that can be large. Perhaps an even bigger problem is that it requires that the model is mathematically differentiable with respect to all its inputs and moreover that these mathematical derivatives are known. This makes it difficult to apply this method to complex dynamic models such as process-based GHG emission models. For such models, the Monte Carlo method is a more viable alternative.

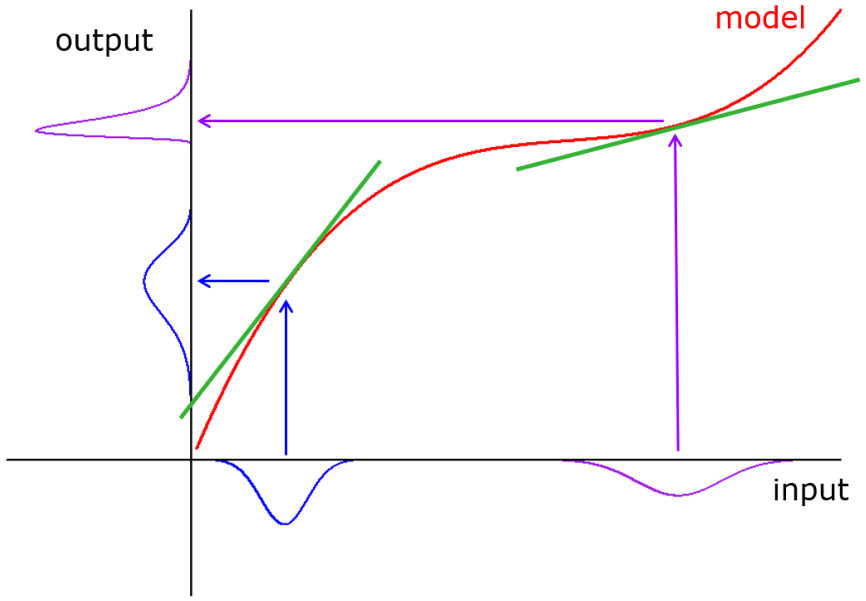

Figure X.2. Graphical illustration of uncertainty propagation and the first order Taylor series method in a case of a single uncertain input. The red line represents the model, the green lines are local linear

approximations of it. In blue a case where the input uncertainty is not very large but the output uncertainty is, because the model is quite sensitive to small changes in the input. In purple the opposite, large input uncertainty but small output uncertainty.

X.2.2 Monte Carlo method

The Monte Carlo method uses an entirely different approach to analyse the propagation of uncertainty. The idea of this method is to compute the model output repeatedly, with input and parameter values that are randomly sampled from their joint probability distribution. The model outputs form a random sample from the probability distribution of the model output, so that parameters of that distribution, such as the mean and variance, can be estimated from the sample. For uncertainty propagation, it is the variance that is of most interest.

The method thus consists of the following steps:

1. Repeat N times:
    a. Generate a set of realisations of the inputs x_i ($i = 1 \ldots n$) and parameters θ_j ($j = 1 \ldots m$).
    b. For this set of realisations, compute and store the model output $y = f(x_1, \ldots, x_n, \theta_1, \ldots \theta_m)$
2. Compute and store sample statistics from the N outputs.

Here, N is the number of Monte Carlo runs, i.e. the Monte Carlo sample size. Next to the variance one can also derive the 0.05 and 0.95 quantiles of the distribution, to obtain a 90% prediction interval of the model output. A histogram or boxplot of the N model outputs also nicely portrays the uncertainty. Realisations of the inputs and parameters are obtained using an appropriate pseudo-random number generator. Note that the inputs and parameters should cannot be simulated independently if they are correlated. They must be simulated simultaneously. This also holds for inputs that are temporally and spatially correlated, in which case temporal and spatial stochastic simulation techniques can be used, such as sequential Gaussian simulation (Webster and Oliver 2007, Chapter 12).

Like the Taylor series method, the Monte Carlo method also involves an approximation error. However, this approximation error can be made smaller by increasing the Monte Carlo sample size (Figure X.3). Although it is case-dependent, in practical applications we often find that at least 100 Monte Carlo runs are required to get reasonable results but that as many as 1,000 or more are needed to get stable results.

The Monte Carlo method works for any model because it only perturbs the model inputs and parameters and leaves the model intact. Another advantage over the Taylor series method is that it provides the entire probability distribution of the model output, not just its variance. These advantages make it the preferred uncertainty propagation method, with numerous applications in the environmental sciences, also for GHG emission modelling (e.g. Nol et al., 2010; Kros et al., 2012; Molto et al., 2013; Stahl et al., 2014; Wojcik-Gront and Gront, 2014; Spafford and MacDougall, 2020; Yanai et al., 2020; Fortin, 2021). An important disadvantage is that it is computationally demanding, because it requires that a model is run many times (but note that the computations can very easily be parallelized).

Uncertainty contributions can also be analysed with the Monte Carlo method. One obvious way of doing that is to repeat the Monte Carlo analysis with one of the inputs or parameters fixed on its deterministic value, and evaluating how much the output variance has decreased by this. More advanced approaches of stochastic sensitivity analysis have also been developed, see Saltelli et al. (2004) for details.

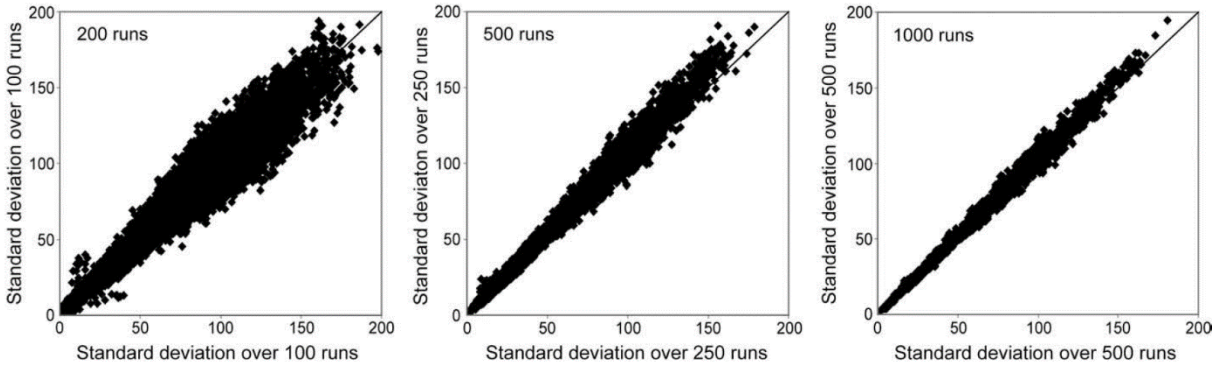

Figure X.3. Scatter plots of standard deviations of N_2O emission (kg N_2O-N ha^{-1} yr^{-1}) over the number of Monte Carlo runs for two Monte Carlo uncertainty propagation analyses, each using a different seed of the random number generator. Left: 100 runs, middle: 250 runs, right: 500 runs. Each diamond represents a grid cell in the study area. The diamonds get closer to the 1:1 line as the number of Monte Carlo runs increases, indicating that both analyses produce a similar standard deviation. Figure and data from Nol et al. (2010).

X.3 Upscaling uncertainties

Most GHG emission models are dynamic and hence predict emissions over time. Model outputs can also be spatialized by running the model at all locations (usually the nodes of a fine grid) in the area of interest. All this means that the GHG emission estimates are distributed in space and time. Often, users are not interested in the GHG emission at points but in the average or total emission for an area (e.g., a field, region, entire country or the globe) and/or time period (e.g., a day, month, year or decade). Upscaling GHG estimates is easy, but how about the associated uncertainties? It turns out that this is much more difficult and cannot be done without accounting for spatial and temporal correlations of the GHG emission estimation errors. Moreover, this problem seems largely ignored by the scientific community.

First, note that spatial and temporal aggregation leads to a decrease of uncertainty. This is because errors partly cancel out. The uncertainty decrease is largest if errors are uncorrelated. Theoretically, if GHG emission estimation errors at the points in the area of interest have zero mean (i.e., have only a random error component and no systematic error) and are uncorrelated then the uncertainty decreases to zero. All errors will cancel out. Mathematically, this follows from:

$$var\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} var(\varepsilon_i) + \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} cov(\varepsilon_i, \varepsilon_j) \quad (5)$$

where the ε_i are the GHG estimation errors at points and n is the total number of points in space and/or time over which the average is taken. In case errors are uncorrelated, the covariance terms are zero so that the second term on the right-hand side of Eq. 5 is zero, while the first term approaches zero as n becomes large. Eq. 5 also shows that the uncertainty of the average crucially depends on the covariances of the estimation errors. This is often ignored or grossly simplified in natural resource inventories, see for instance Lugato et al. (2014), Harris et al. (2021) and Plaza et al. (2018).

Time series models and geostatistics can quantify the spatio-temporal correlations of input errors. If Monte Carlo uncertainty propagation accounts for these correlations then the uncertainty of spatio-temporal averages and totals can be properly addressed. Szatmári et al. (2021) used such approach to derive the uncertainty of the soil organic carbon stock change over time for Hungary at multiple spatial scales. The study confirmed that uncertainty decreases as the area over which is aggregated increases. At point scale none of the estimated soil organic carbon changes between 1992 and 2010 was statistically significant, while at the county and country scale they were.

X.4 Uncertainty reporting in IPCC and UNFCCC guidelines

There are many reports and websites that describe and explain how the IPCC and UNFCCC deal with uncertainties in greenhouse gas emission inventories. In this section we limit ourselves to a review of IPCC documents from the '2006 IPCC Guidelines for National Greenhouse Gas Inventories'[1] and the '2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories'[2]. Note that the UNFCCC adopts the IPCC guidelines[3]. Examples of implementations and applications of these methods as published in the peer-reviewed international scientific literature are Monni et al. (2004), Pichancourt et al. (2018), Calvo et al. (2019), Jonas et al. (2019), Yanai et al. (2020), and Fortin (2021).

X.4.1 IPCC chapters 'Uncertainties'

Chapter 3 of Volume 1 of Eggleston et al. (2006) is exclusively devoted to uncertainty assessment of GHG emission inventories. It starts with definitions and among others explains the difference between accuracy and precision. Accuracy is associated with unbiasedness (i.e., a lack of systematic error) and precision with the degree of random error. Uncertainty is defined as "the lack of knowledge of the true value of a variable that can be described as a probability density function". Thus, in their terminology, estimates and predictions must be both accurate and precise to have low uncertainty. Uncertainty is quantified by the limits of a 95% confidence or prediction interval. This agrees with the statistical description of uncertainty presented in Section X.1.

Note that the IPCC definition of accuracy is opposed to a more conventional definition of accuracy in for example dictionaries, as "the degree to which the results of a measurement, calculation, or specification conforms to the correct value or a standard". Under this definition, accuracy includes both systematic and random error.

The report continues with considering eight broad causes of uncertainty and discusses how each of these could be reduced:

3. Lack of completeness;
4. Model (models are a simplification of real systems);
5. Lack of data;
6. Lack of representativeness of data;
7. Statistical random sampling error;
8. Measurement error;
9. Misreporting or misclassification;
10. Missing data.

The chapter also describes how quantification of the eight causes of uncertainty above can be achieved using empirical data, expert judgement and from published references, and proposes which types of probability distributions can be used. It does not consider the spatial extensions of these distributions. Thus, it ignores quantification of spatial correlations, although it recognises "scale mismatches" and that "uncertainty will tend to increase as the geographic scope decreases". This confirms that uncertainty strongly depends on the spatial and/or temporal support of the estimates and predictions (Section X.3). Temporal autocorrelation of uncertainties is specifically addressed by referring to the use of time series models.

---

[1] https://www.ipcc-nggip.iges.or.jp/public/2006gl/

[2] https://www.ipcc-nggip.iges.or.jp/public/2019rf/index.html

[3] https://www.eea.europa.eu/data-and-maps/indicators/transport-emissions-of-greenhouse-gases/good-practice-guidance-and-uncertainty

The chapter does address statistical dependence and correlations between uncertainties among inputs and how these can be incorporated in Monte Carlo uncertainty propagation methods. It does not present the Taylor series method in full but is restricted to simplified versions of it that address addition, subtraction and multiplication.

The focus of the chapter is on uncertainties in input data, but there is some attention for dealing with model uncertainty. Model uncertainty is typically characterised by an additive residual whose probability distribution can be derived from comparing model predictions with independent measurements of the response variable.

Finally, the chapter provides guidelines for reporting and documentation and provides two concrete examples of uncertainty estimates for inventories.

Chapter 3 of Volume 1 of Calvo et al. (2019) presents refinements to the 2006 IPCC Guidelines. It only addresses uncertainty in data. It starts by stating that an uncertainty assessment consists of four parts:

1. Rigorous investigation of the likely causes of data uncertainty;
2. Development of quantitative uncertainty estimates and parameter correlations;
3. Mathematical combination of those estimates when used as inputs to a statistical model;
4. Selection of inventory improvement actions in response to the results of the previous three parts.

Here, steps 1 and 2 refer to the probabilistic modelling of uncertainty as discussed in Section X.1 of this report, while step 3 addresses uncertainty propagation methods discussed in Section X.2.

The chapter pays much attention to reduction of uncertainties and mentions that this can be achieved in seven broad ways:

1. Improving conceptualisation;
2. Improving models;
3. Improving representativeness;
4. Using more precise measurement methods;
5. Collecting more data that are measured;
6. Eliminating known risk of bias;
7. Improving state of knowledge;
8. Moving to higher tier method.

The chapter also explains the crucial difference between a standard deviation (measure of variability) and standard error (measure of uncertainty). With regard to uncertainty in activity data, it pays attention to how uncertainty due to sampling variability is assessed under different probability sampling designs, such as simple random sampling, stratified sampling, systematic sampling and two-stage sampling. There is also refinement to the Taylor series uncertainty propagation methodology, by incorporating correlations between uncertain inputs, as in Eq. 4 above. However, this is not incorporated in the equations provided in the chapter and neither is an extension made to models that are not additive, subtractive or multiplicative.

X.4.2 IPCC chapters 'Quality Assurance / Quality Control and Verification'

Chapter 6 of Volume 1 of Eggleston et al. (2006) and Chapter 6 of Volume 1 of Calvo et al. (2019) present QA/QC and verification procedures to be used in the development of national greenhouse gas inventories.

Quality Control (QC) is defined as a system of routine technical activities to assess and maintain the quality of the inventory and designed to:

1. Provide routine and consistent checks to ensure data integrity, correctness and completeness;

2. Identify and address errors and omissions;
3. Document and archive inventory material and record all QC activities.

Quality Assurance (QA) is a planned system of review procedures conducted by personnel not directly involved in the inventory process. It refers to reviews performed by independent parties. Verification refers to activities and procedures that help establish the reliability of an inventory. It involves the use of independent data and comparisons with inventory estimates made by other bodies or through alternative methods. It is noted that it is important to distinguish verification, as defined by the IPCC guidelines, from the term verification used in carbon markets, which is synonymous with an independent audit.

The chapters address practical considerations in developing QA/QC and Verification systems and explain the key elements of these systems. There is also reference to ISO standards related to quality management systems. Next, the chapters address category-specific QC procedures in detail, such as used for emissions or removals data QC and activity data QC. It also covers calculation-related QC, among others paying attention to checks of the calculation algorithm that will safeguard against duplication of inputs, unit conversion errors and other calculation errors.

The chapters compare the QA/QC process with uncertainty analyses procedures. For instance, the QA/QC process could identify critical components of the inventory estimates and data sources that contribute to the uncertainty and which should therefore be included in the uncertainty assessment. Conversely, the uncertainty assessment can provide insight into uncertainty propagation and hence inform the QA/QC system.

Chapter 6 of Volume 1 of Calvo et al. (2019) presents a detailed refinement on the components needed for GHG emission inventory verification using atmospheric measurements. It lists the key elements needed and works this out for four target gases (methane, carbon dioxide and nitrous oxide and fluorinated gases). It also presents a refinement on the use of complimentary observations and global modelling products for verification. Here, the use of inverse modelling and satellite observations become important. The chapter also addresses a gap in the 2006 IPCC Guidelines related to the development and use of models. It makes a case for using more complex models although it also notes some well-known adverse effects of modelling. It provides a good practice use of models that can protect against these adverse effects.

X.4.3 IPCC chapters ‘Forest land’

Chapter 4 of Volume 4 of Eggleston et al. (2006) and Chapter 4 of Volume 4 of Calvo et al. (2019) provide methods for estimating greenhouse gas emissions and removals due to changes in biomass, dead organic matter and soil organic carbon on forest land and land converted to forest land. Both chapters include sections on uncertainty assessment, which consider source-specific uncertainties for biomass, dead organic matter and soil carbon.

The source-specific uncertainties for biomass are emission and removal factors and activity data. It is recognised that, due to limited data availability, uncertainty in the estimates of changes of carbon stock in dead organic matter is generally larger than that of the estimates of changes in carbon stock in biomass. For soil carbon inventories, three broad sources of uncertainty are distinguished:

1. Uncertainties in land-use and management activity and environmental data;
2. Uncertainties in reference soil carbon stocks if using tier 1 or 2 approaches;
3. Uncertainties in carbon stock change/emission factors for tier 1 and 2 approaches, and model structure and parameter error for tier 3 model-based approaches, and measurement error and sampling errors associated with tier 3 measurement-based inventories.

For uncertainty propagation analysis, both chapters rely on simple uncertainty propagation methods (i.e., as derived from the Taylor series method), useful for tier 1 and tier 2 approaches, and Monte Carlo simulation, appropriate for tier 3 approaches.

The chapters specifically address the case where land is converted to forest land, which has the same uncertainty sources as before but in addition also uncertainties related to the land use conversion itself.

Chapter 4 of Volume 4 of Calvo et al. (2019) presents tables with quantified uncertainty about variables such as the ratio of below-ground biomass to above-ground biomass, above-ground biomass and above-ground net biomass growth, but it is not explained how these figures were derived and neither is it explained to what spatial support these figures refer.

X.5 Conclusions

Three main sources of uncertainty in GHG estimates are uncertainty in input data, model parameters and model structure. These uncertainties will propagate to the model output. There is a rich scientific literature on uncertainty quantification and uncertainty propagation, also in applications to GHG emission modelling.

Uncertainty in model inputs can best be represented by probability distributions. This is not an easy task in GHG accounting because to do this properly it must also account for cross-correlation, temporal correlation, and spatial correlation. Uncertainty in model parameters is also represented by probability distributions and can be assessed if measurements of the model output are available, so that parameter uncertainty can be quantified using Bayesian calibration. Uncertainty in model structure is difficult to assess and usually represented by an additive noise term.

Once uncertainties in model inputs, parameters and model structure are quantified it is fairly easy to analyse the uncertainty propagation. This can be done with the Taylor series method and the Monte Carlo method. Both methods also allow to assess the contributions of individual uncertainty sources to the overall output uncertainty.

Uncertainty in GHG estimates depends on the spatial and temporal support for which estimates are derived. Uncertainty tends to be smaller for larger supports, such as when averages or totals for large regions or countries are made. This effect is seldom properly assessed.

IPCC and UNFCCC guidelines for national greenhouse gas inventories also represent uncertainty by probability distributions and use the Taylor series method and Monte Carlo simulation to trace the propagation of uncertainties. These guidelines are summaries of methods from the scientific literature although they also provide concrete recipes for specific cases.

Acknowledgements

This review was prepared in the framework of the HoliSoils project, with funding from the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101000289).

I thank Niels H. Batjes for his editorial comments.

References

Calvo et al. (Eds.) (2019), *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*. IPCC, Switzerland. https://www.ipcc.ch/report/2019-refinement-to-the-2006-ipcc-guidelines-for-national-greenhouse-gas-inventories/

Ding, J., K. Miyazaki, V.R. Johannes, B. Mijling, J.I. Kurokawa, S. Cho, G. Janssens-Maenhout, Q. Zhang, F. Liu and P.F. Levelt (2017), Intercomparison of NOx emission inventories over East Asia. *Atmospheric Chemistry and Physics* 17, 10125-10141.

Eggleston, S., L. Buendia, k. Miwa and K. Tanabe (2006), *2006 IPCC Guidelines for National Greenhouse Gas Inventories*. IPCC, Switzerland. https://www.ipcc-nggip.iges.or.jp/public/2006gl/.

Ferrante, M. and T.C.J. Yeh (1999), Head and flux variability in heterogeneous unsaturated soils under transient flow conditions. *Water Resources Research* 35, 1471-1479.

Fortin, M. (2021), Comparison of uncertainty quantification techniques for national greenhouse gas inventories. *Mitigation and Adaptation Strategies for Global Change* 36, 7.

Gelman, A., J.B. Carlin, H.S. Stern and D.B. Rubin (2013). *Bayesian Data Analysis. Texts in Statistical Science* (3rd ed.). Boca Raton, FL: Chapman and Hall/CRC.

Grüneberg, E., D. Ziche and N. Wellbrock (2014), Organic carbon stocks and sequestration rates of forest soils in Germany. *Global Change Biology* 20, 2644-2662.

Harris, N.L., D.A. Gibbs, A. Baccini, R.A. Birdsey, S. De Bruin, M. Farina, L. Fatoyinbo, M.C. Hansen, M. Herold, R.A. Houghton et al. (2021). Global maps of twenty-first century forest carbon fluxes. *Nature Climate Change* 11 , 234–240.

Hashimoto, S., T. Morishita, T. Sakata, S. Ishizuka, S. Kaneko and M. Takahashi (2011), Simple models for soil CO2, CH4, and N2O fluxes calibrated using a Bayesian approach and multi-site data. *Ecological Modelling* 222, 1283-1292.

Henne, S., D. Brunner, B. Oney, M. Leuenberger, W. Eugster, I. Bamberger, F. Meinhardt, M. Steinbacher and L. Emmenegger (2016), Validation of the Swiss methane emission inventory by atmospheric observations and inverse modelling. *Atmospheric Chemistry and Physics* 16, 3683-3710.

Heuvelink, G.B.M. (1998a), Uncertainty analysis in environmental modelling under a change of spatial scale. *Nutrient Cycling in Agro-ecosystems* 50, 255-264.

Heuvelink, G.B.M. (1998b), *Error Propagation in Environmental Modelling with GIS*. Taylor & Francis, London.

Heuvelink, G.B.M. (2014), Uncertainty quantification of GlobalSoilMap products. In: *GlobalSoilMap. Basis of the Global Spatial Soil Information System*. D. Arrouays, N. McKenzie, J. Hempel, A. Richer de Forges and A.B. McBratney (Eds.), pp. 335-340.

Heuvelink, G.B.M. (2018), Uncertainty and uncertainty propagation in soil mapping and modelling. In: *Pedometrics*. A.B. McBratney, B. Minasny and U. Stockmann (eds.). Springer.

Heuvelink, G.B.M., J.D. Brown and E.E. Van Loon (2007), A probabilistic framework for representing and simulating uncertain environmental variables. *International Journal of GIS* 21, 497-513.

Hoffmann, U. T. Hoffmann, E.A. Johnson and N.J. Kuhn (2014), Assessment of variability and uncertainty of soil organic carbon in a mountainous boreal forest (Canadian Rocky Mountains, Alberta). *Catena* 113, 107-121.

Inness, A., A.M. Blechschmidt, I. Bouarar, S. Chabrillat, M. Crepulja, R.J. Engelen, H. Eskes, J. Flemming, A. Gaudel, F. Hendrick, V. Huijnen, L. Jones, J. Kapsomenakis, E. Katragkou, A. Keppens, B. Langerock, M. de Maziere, D. Melas, M. Parrington, V.H. Peuch, M. Razinger, A. Richter, M.G. Schultz, M. Suttie, V. Thouret, M. Vrekoussis, A. Wagner and C. Zerefos (2015), Data assimilation of satellite-retrieved ozone, carbon monoxide and nitrogen dioxide with ECMWF's Composition-IFS. *Atmospheric Chemistry and Physics* 15, 5275 5303.

Jonas, M., R. Bun, Z. Nahorski, G. Marland, M. Gusti and O. Danylo (2019), Quantifying greenhouse gas emissions. *Mitigation and Adaptation Strategies for Global Change* 24, 839-852.

Kennedy M.C. and A. O' Hagan (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B* 63, 425-464.

Kros, J., G.B.M. Heuvelink, G.J. Reinds, J.P. Lesschen, V. Ioannidi and W. De Vries (2012), Uncertainties in model predictions of nitrogen fluxes from agro-ecosystems in Europe. *Biogeosciences* 9, 4573-4588.

Lehuger, S., B. Gabrielle, M. van Oijen, D. Makowski, J.-C. Germon, T. Morvan and C. Henault (2000), Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model. *Agriculture, Ecosystems and Environment* 133, 208-222.

Lehuger, S., B. Gabrielle, P. Laville, M. Lamboni, B. Loubet and P. Cellier (2011), Predicting and mitigating the net greenhouse gas emissions of crop rotations in Western Europe. *Agricultural and Forest Meteorology* 151, 1654-1671.

Liao, K.H., L.G. Lv, X.M. Lai and Q. Zhu (2021), Toward a framework for the multimodel ensemble prediction of soil nitrogen losses. *Ecological Modelling* 456, 109675.

Lindgren, F. and H. Rue (2015), Bayesian spatial modelling with R-INLA. *Journal of Statistical Software* 63, 1-25.

Lugato, E., P. Panagos, F. Bampa, A. Jones and L. Montanarella, L. (2014). A new baseline of organic carbon stock in European agricultural soils using a modelling approach. *Global Change Biology*, 20 , 313–326

Magnussen, S., M. Kohl and K. Olschofsky (2014), Error propagation in stock-difference and gain-loss estimates of a forest biomass carbon balance. *European Journal of Forest Research* 133, 1137-1155.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research 7*, 983-999.

Molto, Q., V. Rossi and L. Blanc (2013), Error propagation in biomass estimation in tropical forests. *Methods in Ecology and Evolution* 4, 175-183.

Monni, S., S. Syri and I. Savolainen (2004), Uncertainties in the Finnish greenhouse gas emission inventory. *Environmental Science & Policy* 7, 87-98.

Nol, L., G.B.M. Heuvelink, A. Veldkamp, W. De Vries and J. Kros (2010). Uncertainty propagation analysis of an N2O emission model at the plot and landscape scale. *Geoderma* 159, 9 23.

O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, J., Garthwaite, P., Jenkinson, D., Oakley, J. and T. Rakow (2006). *Uncertain Judgements: Eliciting Experts*. Wiley, Chichester.

Pichancourt, J.B., R. Manso, F. Ningre and M. Fortin (2018), A carbon accounting tool for complex and uncertain greenhouse gas emission life cycles. *Environmental Modelling & Software* 107, 158-174.

Plaza, C., C. Zaccone, K. Sawicka, A.M. Mendez, A. Tarquis, G. Gasco, G.B.M. Heuvelink, E.A. Schuur and F.T. Maestre (2018), Soil resources and element stocks in drylands to face global issues. *Scientific Reports* 8 , 1-8.

Quetin, G.R., A.A. Bloom, K.W. Bowman and A.G. Konings (2020), Carbon flux variability from a relatively simple ecosystem model with assimilated data is consistent with terrestrial biosphere model estimates. *Journal of Advances in Modeling Earth Systems* 12, e2019MS001889.

Rahn, K.H., K. Butterbach-Bahl and C. Werner (2011), Selection of likelihood parameters for complex models determines the effectiveness of Bayesian calibration. *Ecological Informatics* 6, 333-340.

Refsgaard, J.C., J.P. van der Sluijs, A.L. Højberg and P.A. Vanrolleghem (2007), Uncertainty in the environmental modelling process: a framework and guidance. *Environmental Modelling & Software* 22, 1543-1556.

Saltelli, A., S. Tarantola, F. Campolongo and M. Ratto (2004), *Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models.* Wiley, Chicester.

Spafford, L. and A.H. MacDougall (2020), Quantifying the probability distribution function of the transient climate response to cumulative CO2 emissions. *Environmental Research Letters* 15, 034044.

Stahl, G., J. Heikkinen, H. Petersson, J.R. La and S. Holm (2014), Sample-based estimation of greenhouse gas emissions from forests - A new approach to account for both sampling and model errors. *Forest Science* 60, 3-13.

Sun, Q. B. Liao and Q.Y. Tao (2019), Ecological agriculture development and spatial and temporal characteristics of carbon emissions of land use. *Applied Ecology and Environmental Research* 17, 11045-11053.

Sutthichaimethee, J. and K. Kubaha (2018), Forecasting energy-related carbon dioxide emissions in Thailand's construction sector by enriching the LS-ARIMAXi-ECM model. *Sustainability* 10, 3593.

Szatmári, G. and L. Pásztor (2019), Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms. *Geoderma* 337, 1329-134.

Szatmári, G., L. Pásztor and G.B.M. Heuvelink (2021), Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics. *Geoderma* 403, 115356.

Taylor, J.R. (1982), *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. 2nd Edition. University Science Books, Mill Valley.

Van Leeuwen, M.M.W.J., G.B.M. Heuvelink, J. Wallinga, I.J.M. De Boer, J.C. Van Dam, E.A. Van Essen, S.W. Moolenaar, F.P.M. Verhoeven, J.J. Stoorvogel and C.R. Stoof (2018), Visual soil evaluation: reproducibility and correlation with standard measurements. *Soil & Tillage Research* 178, 167-178.

Van Leeuwen, C.C.E., V.L. Mulder, N.H. Batjes and G.B.M. Heuvelink (2022), Statistical modelling of measurement error in wet chemistry soil data. *European Journal of Soil Science* 73, e13137.

Van Oijen, M., J. Rougier and R. Smith (2005), Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiology* 25, 915-927.

Vaysse, K. and P. Lagacherie (2017), Using quantile regression forest to estimate uncertainty of digital soil mapping products. *Geoderma* 291, 55-64.

Wadoux, A.M.J.-C. (2019). Using deep learning for multivariate mapping of soil with quantified uncertainty. *Geoderma* 351, 59-70.

Wadoux, A.M.J-C., G.B.M. Heuvelink, R. Uijlenhoet and S. de Bruin (2020), Optimization of rain gauge sampling density for river discharge prediction using Bayesian calibration. *PeerJ* 8, e9558.

Wagena, M.B., G. Bhatt, E. Buell, A.R. Sommerlot, D.R. Fuka and Z.M. Fuka (2019), Quantifying model uncertainty using Bayesian multi-model ensembles. *Environmental Modelling & Software* 117, 89-99.

Webster, R. and M.A. Oliver (2007). *Geostatistics for Environmental Scientists* (2nd ed.). Wiley, Chichester.

Wojcik-Gront, E. and D. Gront (2014), Assessing uncertainty in the Polish agricultural greenhouse gas emission inventory using Monte Carlo simulation. *Outlook on Agriculture* 43, 61-65.

Yanai, R.D., C. Wayson, D. Lee, A.B. Espejo, J.L. Campbell, M.B. Green, J.M. Zukswert, S.B. Yoffe, J.E. Aukema and A.J. Lister (2020), Improving uncertainty in forest carbon accounting for REDD plus mitigation efforts. *Environmental Research Letters* 15, 124002.

Zhao, X., M. Han, L.L. Ding and A.C. Calin (2018), Forecasting carbon dioxide emissions based on a hybrid of mixed data sampling regression model and back propagation neural network in the USA. *Environmental Science and Pollution Research* 5, 2899-2910.