Guidance Note on estimating the uncertainty of emission reductions using Monte Carlo simulation

Version 1.0

September 2021

Contents

Guidance Note on estimating the uncertainty of emission reductions using Monte Carlo simulation

Criterion 9 of the FCPF Methodological Framework requires that the uncertainty of the estimate of Emission Reductions be quantified using Monte Carlo methods. The FCPF Guidelines on Uncertainty Analysis of Emission Reductions describe which sources of uncertainty to propagate and provide guidance for conducting Monte Carlo simulations.

ER Programs differ in both the activities they consider and the methods by which they calculate emission factors and activity data. The sources of uncertainty thus differ, and the calculations for correctly combining them will also differ. This document describes the general approach and provides a simple example to illustrate the approach. The general approach described here includes the following steps:

- Step 1. Identify sources of values used in the emission reduction estimates and whether they are independent or shared
- Step 2. Identify the uncertainty associated with each of these variables
- Step 3. Propagate the uncertainties in the estimate of emission reductions using Monte Carlo simulation
- Step 4. Evaluate the contribution of each source to the overall uncertainty

The simple example has been provided in Excel and in R to help users to understand this guidance and each of the steps. Each situation is unique, and the examples given will need to be adapted to the programs in question, but the underlying principles are universal. More details on the example are provided in Annex 1.

This guidance note complements the FCPF Guidelines on Uncertainty Analysis.

Step 1. Identify sources of values used in the emission reduction estimates and whether they are independent or shared.

Following the process by which ER Programs estimated emission reductions, programs should identify all the variables used in the estimation of emissions and removals. Table 1 in the FCPF Guidelines on Uncertainty Analysis provides a list of the main sources of uncertainty that, at minimum, shall be evaluated. In the Monte Carlo simulation, the uncertainties in these variables will be combined by sampling from the likely distributions of their values as determined in Step 2 below.

To combine the uncertainties of multiple variables correctly, requires understanding which variables are used independently and which are shared across multiple calculations.

Variables and their associated uncertainty sources contribute independently to a particular calculation if they are independently derived and are not used for any other variable. For example, tree inventory data are generally collected independently for each stratum or land cover type and are not used in the calculations for other strata or land cover types (shown in blue in Figure 1.1).

Other variables and their associated uncertainties are shared across multiple calculations, for example, carbon fraction (CF), root:shoot ratio (R:S), and tree allometry variables might be used

across multiple forest types (shown in red in Figure 1.1). In that case, carbon densities are calculated by a combination of independent and shared sources and are thus partially correlated (shown in purple in Figure 1.1) and it is important to represent these correlations properly when combining them in error propagation.
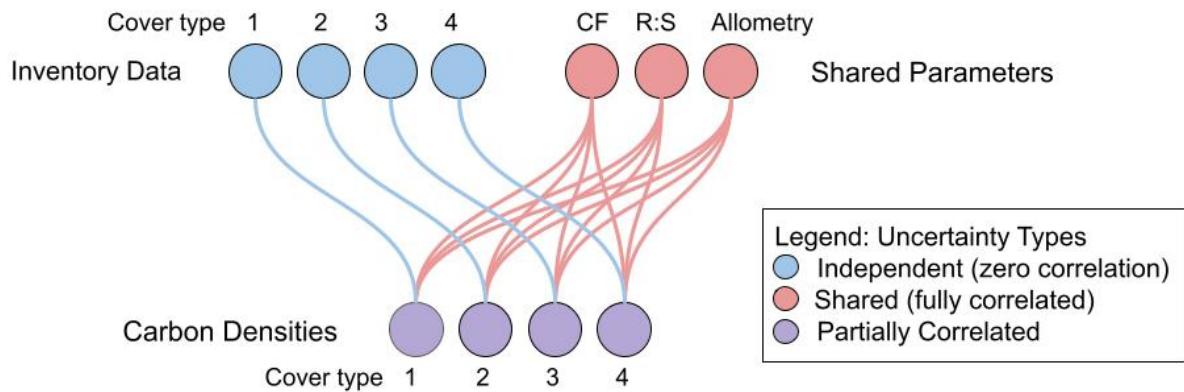

Figure 1.1 In calculating carbon densities, inventory data collected in four types of land cover are independent (blue). If the same values of carbon fraction (CF), root:shoot ratio (R:S), and tree allometry are used across multiple forest types, these are shared (red). Uncertainties calculated from a combination of independent and shared sources will be partially correlated (purple), with correlation coefficients intermediate between 0 (fully independent) and 1 (fully shared).

Whether an uncertainty source should be treated as shared or independent depends on how it was collected and how it is used in the calculation. Some programs collect tree allometry data independently for each land cover type. In that case, uncertainties in tree allometry would be independent for each cover type (Figure 1.2). Theoretically, carbon fraction and root:shoot ratios could be determined independently for each cover type, or a single value could be used across forest types. The conversion of carbon to CO2 is treated as a constant without uncertainty, because variability in carbon and oxygen isotope ratios is negligible.
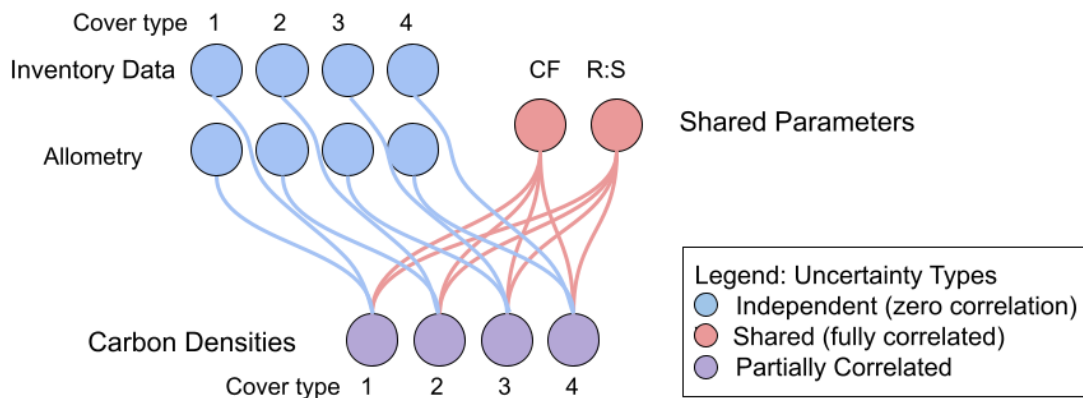
Figure 1 2. Whether a variable is independent or shared depends on how the data were collected. If allometric equations are specific to each cover type, they are independent. Carbon fraction and root:shoot ratio could be determined separately for each forest type, but if they are not, they are shared.

Since the emission reductions are estimated as the difference between the reference level of emissions and the actual monitored emissions, correlation is also relevant between the two estimates. For example, if programs use the same emission factors in the reference and

monitoring periods, they should be treated as shared (as shown in Figure 1.3). In contrast, if inventory data are collected independently for each time period, the emission factors for the two periods are more independent.

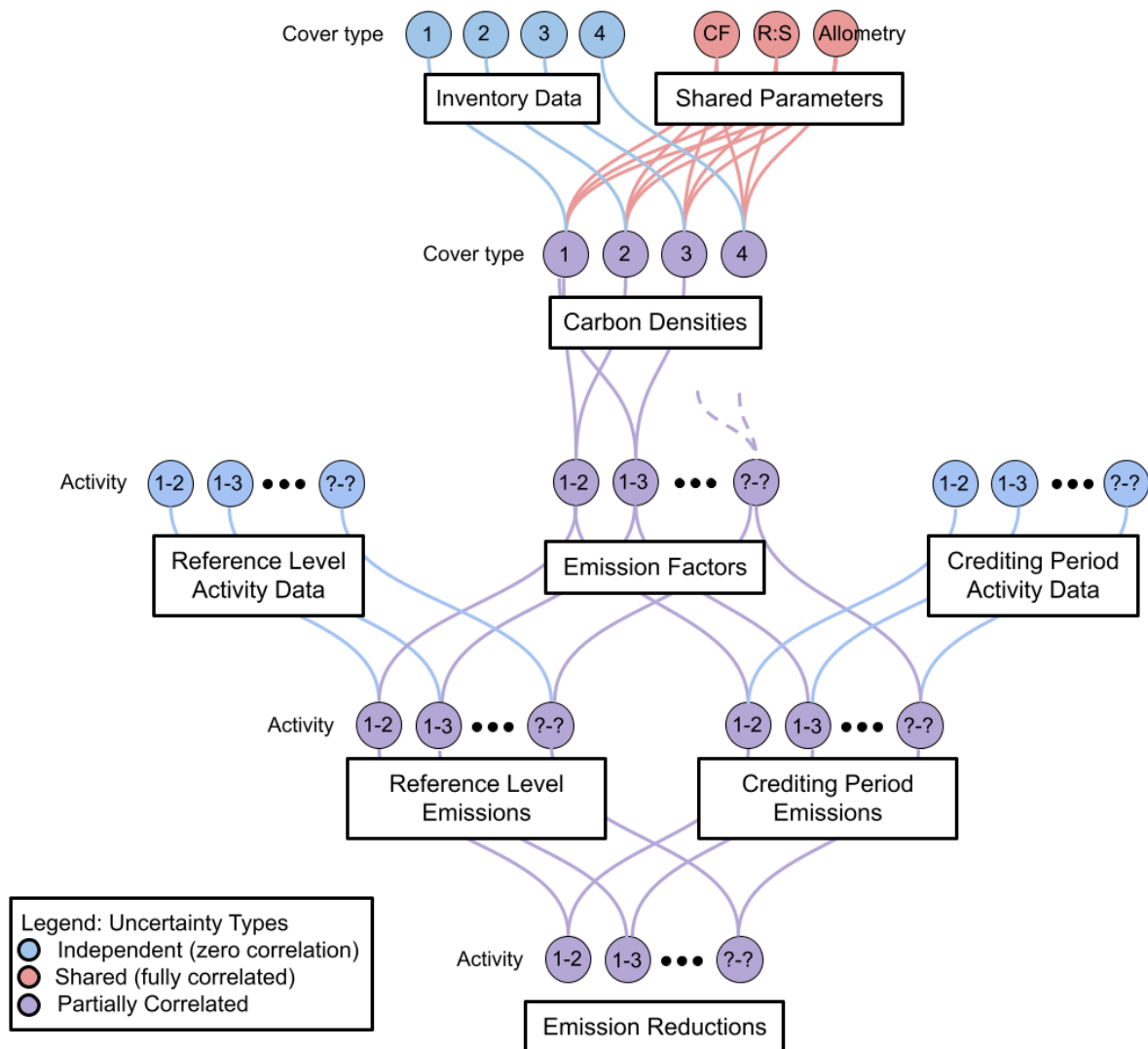**Figure 1 3. Emission factors are derived from carbon densities of the categories of land cover in transition. Many such transitions are possible and the number and degree of correlation of such activities will differ by program depending on the design. Emission factors may be shared between the reference level and monitoring period, as shown here. In estimating the uncertainty in the emission reductions, shared sources of uncertainty will share a randomly sampled value at each iteration of the Monte Carlo simulation. Uncertainties associated with results based on both shared and independent sources will be partially correlated (shown in purple).**

The degree of partial correlation (i.e., the value of the correlation coefficient) that arises from combining a mix of shared and independent uncertainty sources can be estimated analytically (not covered in this guidance) or by examining the results of Monte Carlo simulation. It is not important to know the correlation coefficient of intermediate variables in the calculation. However, if the calculation were to begin with partially correlated inputs, such as carbon densities, the correlation coefficients would need to be estimated to be propagated correctly (see Section 3.2.3).

Potential Pitfalls

Treating uncertainty sources as independent when they are shared will underestimate the true combined uncertainty. In the Simple Example provided in Annex 1, the uncertainty in the ER should be 416%. If the emission factors are treated as independent between the reference level and the monitoring period, instead of shared, then a value of 403% is obtained, which is incorrect.

Step 2. Identify the uncertainty associated with each of these variables.

Using the Monte Carlo approach to error propagation requires defining the distributions of the variables used in the calculation. There are two ways to generate random samples that mimic the likely distribution of a variable. The first is to sample from a defined distribution (a probability density function or PDF). The second is to randomly sample values of the variable from a data set (bootstrapping).

2.1 Decide between PDF and bootstrap

It is easier to sample from a defined distribution than to sample from a data set, especially in Excel. Sampling from a data set has the advantage that no assumptions are required about the nature of the distribution. If the distribution is not normal, then bootstrapping would be more accurate, unless the data are not representative. The figure below provides a simple decision tree to decide between PDF and bootstrap to generate random samples that mimic the likely distribution of a variable.
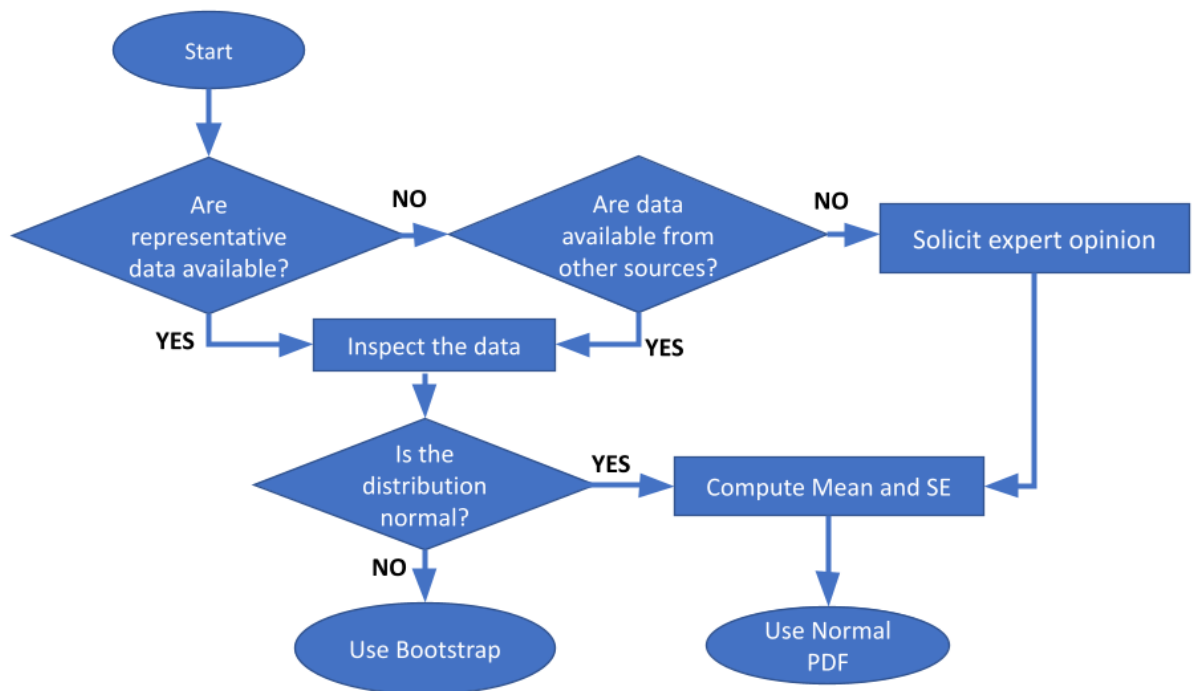

Figure 2.1. Decision tree for choosing whether to sample from data or from a PDF.

2.2 Describing uncertainty with a probability density function (PDF)

A mathematical function can be used to describe the distribution of possible values of a variable. Commonly, in the absence of information to the contrary, normal distributions are used. A distribution other than normal could be selected if the possible values are not normally distributed. For some applications, beta, binomial, gamma, weibull, or lognormal distributions best describe the distribution of observations. However, if there are enough data to determine that the distribution is not normal, then bootstrapping is also an option.

In REDD+ carbon accounting, uniform distributions have sometimes been used to describe possible values of the carbon fraction (IPCC 2006 Table 4.3) and root:shoot ratio (IPCC 2019 refinement of 2006 Table 4.4). For this reason, we provide guidance on how to use a uniform PDF. However, it seems implausible that there is a zero probability of a value outside these ranges. Instead, we recommend that the mean and standard error of the available data be used to define a normal distribution. Alternatively, the values could be sampled by bootstrapping.

In the absence of reliable data, expert judgement may be used to define a PDF. The FCPF Guidelines on Uncertainty Analysis of Emission Reductions recommends independently consulting at least three experts when the parameter estimate is not available or is not representative (e.g. based on research plots). The mean and standard error of the mean of expert opinions should be used to define a normal distribution. Using the range would be sensitive to extreme values, and doubling the range (as currently recommended in the Guidelines) would inflate the uncertainty of the emission reductions. Alternatively, the values could be sampled by bootstrapping.

2.3 Using the distribution of the data (bootstrapping)

An alternative to representing the distribution of the inputs analytically is to resample the data, a procedure known as bootstrapping (Ephron and Tibshirani 1994) that is illustrated in figure 2.2. Values are randomly drawn from the data to create alternative possible data sets with similar distributions and the same number of observations. Each random sample is drawn from all possible samples (this is called "sampling with replacement") because sampling without replacement, if drawing the number of observations in the data set, would return the original data set every time. This approach requires no assumption of a distribution and is thus most true to the measured population. Bootstrapping is especially advantageous when the distribution is difficult to define. The drawback to this approach is that the representation of the population is only as good as the data, and if the data set is small, it may not accurately capture the range of potential values.
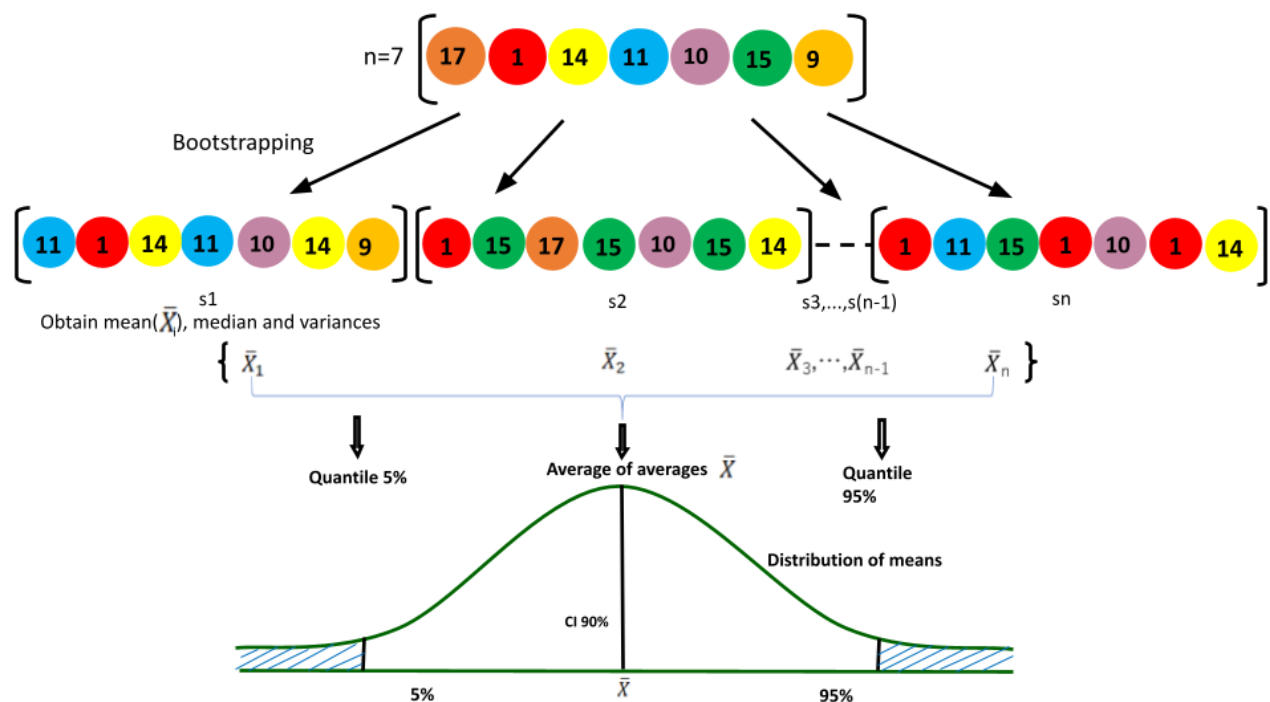
Figure 2.2. Random sampling of values from a data set. The data set has 7 values, which can be randomly sampled (with replacement) to create alternative possible data sets.

Potential Pitfalls

Bootstrap sampling must be configured to match the sampling design. For example, if the sampling design is stratified, it would be necessary to bootstrap for each stratum separately and produce the overall estimates by combining the stratum information. There could be similar issues for cluster sampling.

Often data are not representative, due to access to sampling locations (close to roads or forest edges), or ease of measurement (excavating roots of small trees). This is a potential pitfall both for characterizing the data with a PDF and for bootstrapping. In these cases, expert judgement should be used to correct for bias in the data.

Step 3. Propagate the uncertainties in the estimate of emission reductions using Monte Carlo simulation

The FCPF requires the use of Monte Carlo simulation to quantify the effects of uncertain inputs on the uncertainty of carbon emission reductions. Using this approach, the calculation of the emission reductions is iterated hundreds or thousands of times, with the inputs varying randomly to mimic the uncertainties in the values of all the variables that went into the calculation of both the Reference Level and the monitored emissions and removals. As described above in Step 2, random samples that mimic the likely distribution of a variable can be generated using a defined distribution (a PDF) or by randomly sampling values of the variable from a data set (bootstrapping). The distribution of the resulting hundreds or thousands of outputs reflects the net effects of the uncertainties in the inputs.

3.1 How to randomly sample from a distribution or a dataset

3.1.1 Sampling from a distribution

For simulating the uncertainty in an input using a defined distribution, the parameters describing the distribution are used to generate the random samples.

Distribution	Spreadsheet formulas	R code
Uniform	RAND() generates random numbers uniformly distributed between 0 and 1. Generate uniform random numbers between values in A1 and A2 using: = ($A$1 + RAND()*($A$2-$A$1))	Generate n uniform random numbers between a1 and a2 using: runif(n, min = a1, max = a2)
Normal	NORM.INV(probability, mean, sd) gives an inverse of the normal cumulative distribution, at a specified probability, mean, and standard deviation (sd). Combining RAND and NORM.INV gives a random sample from a distribution with a specified mean and standard deviation. If the mean and sd are in cells B1 and B2, use: =NORM.INV(RAND(), $B$1, $B$2)	Given the mean and standard deviation (sd) of the distribution, generate n random samples using: rnorm(n, mean, sd)

3.1.2 Sampling from a dataset (bootstrapping)

For simulating the uncertainty in an input using bootstrapping, each random sample is selected from the data set of possible inputs.

Spreadsheet Formula	R code
If 10 data points are in cells A1:A10, you can randomly sample 1 of them using: =INDEX($A$1:$A$10,ROWS($A$1:$A$10)*RAND()+1,COLUMNS($A$1:$A$10)*RAND()+1) Since this formula draws a single value at random from the specified data set, you need to copy it many times to generate a bootstrap sample. Copy this formula to sample many observations and be copied for as many iterations as needed.	Given a sample of size n: (c1,c2,c3,...,cn), you can generate a vector: SampleC<-c(c1,c2,c3,...,cn) Defining a number of simulations (NS), you can generate bootstrap samples (BootstrapSam) by generating NS samples of size n sampled of vector "SampleC" with replacement:

	BootstrapSam <- replicate(NS, sample(SampleC, replace = T))

3.2 How to sample multiple sources of uncertainty

To account correctly for independent vs. shared sources of uncertainty (identified in Step 1, above) requires assigning the random samples of the input values independently, in the case of independent samples (3.2.1), or assigning the same random value in all calculations that share that value, in the case of shared sources (3.2.3), at each iteration of the Monte Carlo simulation.

3.2.1 How to assign independent random values

In the case of independent input variables, random values will be selected independently for each variable to represent uncertainty in those variables. For example, activity data are collected independently at each point in time.

Spreadsheet Formula	R code

For independent sources, the cells referenced for parameters (e.g. mean and sd in the case of a normally distributed source) are not the same:
 	
SourceA:
A3=NORM.INV(RANDARRAY(1,n),$A$1,$A$2)
SourceB:
  B3=NORM.INV(RANDARRAY(1, n),$B$1,$B$2)
 	
Cell A4 would have a different random number referenced by all the calculations in row 4, and so on for all the rows of the simulation.
	For independent sources (SourceA and SourceB), random numbers of a specific distribution can be generated independently.
	
	Considering two sources normally distributed:
	SourceA has mean=mSA and sd=sdSA and
	SourceB has mean=mSB and sd=sdSB.
	
	You can generate n random numbers of SourceA (SimNumSA) and SourceB (SimNumSB) independently as follows:
	
	  SimNumSA<-rnorm(n, mSA, sdSA)
	  SimNumSB<-rnorm(n, mSB, sdSB)

3.2.2 How to assign shared random values

There are cases where the same input variable is used multiple times in a calculation, and in these cases, that input variable should have only one random value for each iteration. For example, if a common root:shoot ratio is used across multiple forest types, a random sample of a possible R:S is selected at each iteration, and each forest type that requires a R:S value uses that same value at that iteration.

Spreadsheet Formula	R code

For shared sources, the random sample of an uncertainty input is shared:
 	
SourceA:
 	
  A3=NORM.INV(RANDARRAY(1, n),\$A\$1,\$A\$2)
 	
SourceB:
 	
  B3=\$A\$3
 	
In this example, iterations are in rows, and every column within a row that uses this source would reference cell A3 for a random value of this source.
	For sources (SourceA and SourceB) that are shared, the random sample for the input will be shared by both sources for each iteration of the Monte Carlo simulation.
	
	Considering a shared source normally distributed: SourceA has mean=mA and sd=sdA
	
	You can generate n random numbers of SourceA as follows:
	
	  SimNumA<- rnorm(n, mA, sdA)
	
	Each source that is shared with SourceA will reference the random number drawn for that iteration
	
	  SimNumB <- SimNumA

3.2.3 How to assign partially correlated random values

Input variables to a calculation may be neither shared nor independent, but rather partially correlated, if these variables are calculated from a combination of shared and independent sources. It is easier to conduct a Monte Carlo simulation beginning with the inputs that are fully independent and fully shared, and this is what we recommend. However, if beginning a calculation with partially correlated variables as inputs, such as emission factors based on independent forest inventory but shared root:shoot ratios, then the Monte Carlo will require partially correlated random samples for these variables.

Creating a matrix with more than two variables in Excel is possible but challenging (Zaiontz 2020).

Spreadsheet Formula	R code
1. For covariance between two vectors X (A1:A30) and Y (B1:B30), a 2 x 2 variance-covariance matrix will be needed.	Load the MASS package, which can develop multivariate normal random numbers:

Matrix A

	X	Y
X	=VAR.S(A1:A30)	=COVARIANCE.S(A1:A30,B1:B30)
Y	=COVARIANCE.S(B1:B30,A1:A30)	=VAR.S(B1:B30)

```
library(MASS)
```

First generate a 2 x 2 variance-covariance matrix between vectors X and Y

```
A<-cov(cbind(X,Y))
```

2. Calculate the mean for both variables with:

Matrix B

X	=AVERAGE(A1:A30)
Y	=AVERAGE(B1:B30)

Generate n pairs of random values of X and Y with the variance-covariance specified in A:

```
mvrnorm(n,c(mean(X),mean(Y)), A)
```

3. Calculate a Cholesky decomposition of the covariance matrix (Pistilli 2019). This will also be a 2 x 2 matrix.

Matrix C

	X	Y
X	=SQRT(A_{11})	=0
Y	=COVARIANCE.S(A_{21}/C_{11})	=SQRT(A_{22} - C_{21}^2)

4. Calculate a random bivariate normal vector. For n random number iterations, this would be a 2 x n matrix with the same formula in each cell. Highlight cells to form a 2 x n matrix, type the formula: =MMULT(**C** NORM.INV(RANDARRAY(2,n))+**B** After typing the formula, pressing ctrl + shift + enter on the keyboard will fill all of the cells in the matrix with the function. Each cell in the random bivariate normal vector is the mean of that variable plus a random residual with the desired correlation.	

3.3 How to iterate

Random sampling, whether from a PDF or from data, can be repeated many times to generate a distribution of estimates from which the uncertainty can be assessed.

Spreadsheet Formula	R code
Each calculation is repeated in rows (or columns). Excel 365 has a RANDARRAY function which facilitates this process. The RANDARRAY formula is used in combination with the formula used to sample from the data. RANDARRAY(R,C) where R is the specified number of rows and C is the specified number of columns. For example, if you are doing iterations in rows, R refers to the number of iterations, and C is 1. For example, if sampling from a normal distribution with mean in cell A1, sd in cell A2, and the number of iterations in cell A3, = NORM.INV(RANDARRAY($A$3,1), $A$1, $A$2) The advantage of RANDARRAY is that the formula is represented only once, instead of separately in each iteration. Having thousands of formulas makes the file enormous and the execution slow. You should probably invest in Office 365 if you want to do Monte Carlo in Excel.	In the case of a normal distribution with a specified mean (mean) and standard deviation (sd), a random number can be generated as follows: rnorm(1,mean,sd) R uses matrices and arrays to store data. Repeated calculations are handled by the number of elements in the array (n). To generate n random numbers, indicate n: rnorm(n,mean,sd)

| The # sign can be used in place of =NORM.INV(RANDARRAY) in simple formulas that do not sample from a distribution. The # copies the formula to fill the array. This is useful for calculations that do not require random sampling.

This is effectively the same as the advantage of RANDARRAY, the formula needs to be typed only in the first cell, and it will autofill for every iteration. | |

Potential Pitfalls

If converted to a Google Sheet, all of the formulas using the # will have an error message called ANCHOR ARRAY.

Using too few Monte Carlo iterations can provide imprecise uncertainty estimates. For your emission reduction calculation, you can determine the number of Monte Carlo iterations needed to achieve a desired confidence in your uncertainty estimates. In this example (based on the Simple Example in Annex 1), uncertainty estimates are accurate to only about 20% of the emission reduction even after 2000 iterations but approach 10% at 10,000 iterations (Figure 3.3).
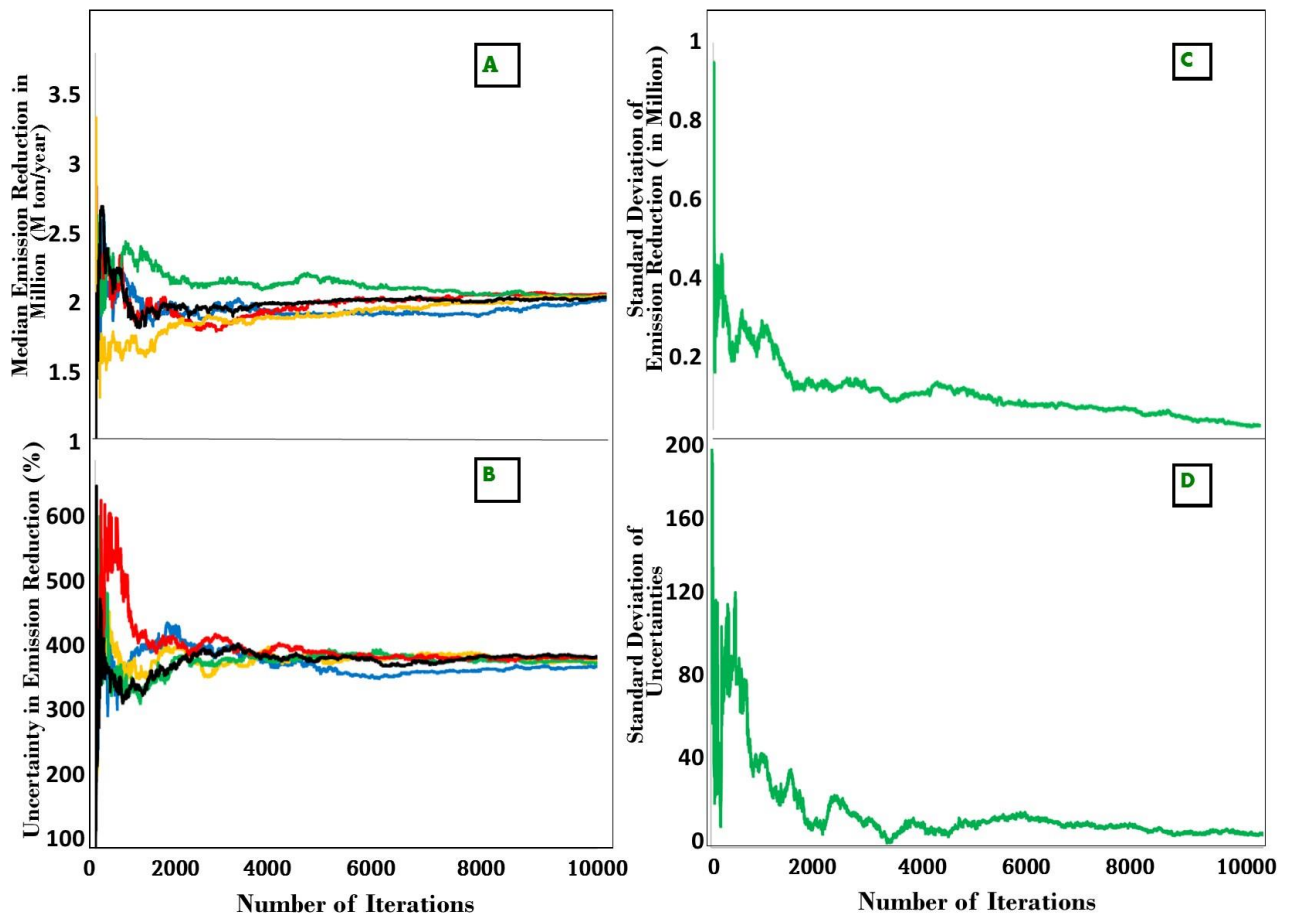
Figure 3.3. The similarity of Monte Carlo estimates depends on the number of iterations.

a) **Median of Monte Carlo estimates of the emission reductions, calculated for numbers of iterations up to 10,000 for 5 independent Monte Carlo runs.**

b) **(b) The half width of the 90% CI of the Monte Carlo iterations divided by the median ER for the same 5 Monte Carlo runs.**

c) **The standard deviation of the 5 Monte Carlo estimates of emission reductions.**

d) **The standard deviation of the 5 Monte Carlo estimates of uncertainty.**

3.4 How to interpret the output of the Monte Carlo

The Monte Carlo output can be analyzed to characterize the uncertainty in the results of the calculation.

1. Find the median (50th percentile) of the Monte Carlo outputs
2. Find the 5th percentile of the outputs
3. Find the 95th percentile of the outputs
4. Calculate the half-width of the 90% confidence interval.
5. Convert this to a percentage of the median.

Spreadsheet Formula	R Code
Where Monte Carlo output is in cells B1:B10000 A1 =MEDIAN(B1:B10000) A2 =PERCENTILE(B1:B10000, 0.05) A3 =PERCENTILE(B1:B10000, 0.95) A4 =(A3-A2)/2 A5 =A4/A1*100	Considering a vector of simulated numbers SimNumC (c1,c2,c3,...,cn): A1<- median(SimNumC) A2<- quantile(SimNumC , 0.05) A3<- quantile(SimNumC , 0.95) A4<- (A3-A2)/2 A5<- abs(A4/A1)*100

Potential Pitfalls

A common mistake in interpreting Monte Carlo output is to report the uncertainty in the mean or median of the distribution of the estimates. This mistake is a big one, typically underreporting uncertainty by a factor of 100, because calculating uncertainty in the central tendency (e.g., the standard error of the mean) involves dividing the standard deviation by the square root of the number of "observations," which is commonly 10,000 trials. The confidence in the mean could be made arbitrarily small by increasing the number of Monte Carlo iterations, but the 90% confidence interval of the increased number of estimates would remain just as wide. Increasing the number of iterations improves the precision of the uncertainty estimate, but, interpreted correctly, it does not make the uncertainty smaller.

Step 4. Evaluate the contribution of each source to the overall uncertainty

Understanding how much uncertainty is contributed by each source will help to identify opportunities for reducing uncertainties. Evaluating the overall uncertainty with different inputs assumed to be perfectly known is one way of assessing the sensitivity of the overall uncertainty to uncertainty in each input.

Conducting a sensitivity analysis is facilitated by switches for each input that turn the associated uncertainty on or off. The overall uncertainty can be calculated with different combinations of switches turned on, rather than changing the formulae in the Excel file or the code in R.

Spreadsheet Formula	R Code
This is accomplished with an "IF" statement referencing a cell used to switch a source "on" or "off" for the sensitivity analysis. For example, if cell A5 on the results page has the value "on" or "off", and cell B5 on the inputs page has IF('resultspage'A5="on", 1, 0.0000000001) Each iteration of a random draw for the source is multiplied by cell B5. If the uncertainty source is "off", 0.0000000001 is used instead of 0 to avoid errors in the NORM.INV function.	The code relevant to an uncertainty source can be activated using an "if" statement.

Comparing the importance of uncertainty in the various inputs can be accomplished by evaluating each one alone, with all other uncertainties turned off, or by removing each one, with all other sources turned on. From the point of view of evaluating the benefit of reducing a particular source in the context of all the others, it is more relevant to report how much uncertainty is reduced by eliminating that source than to report how much that source contributes alone, and this is the approach recommended in the FCPF Guidelines on the Application of the Methodological Framework. However, it is easier to understand the results of considering one source at a time. And if there is a need to revise one source,

The following table shows the results of a sensitivity analysis of the Simple Example provided in the Annex.

Sources included	Uncertainty (Megatons C/year)			Uncertainty (% of Emissions)		
	Reference Level	Crediting Period	ER	Reference Level	Crediting Period	ER

R:S	2.80	2.50	0.29	17	17	17
CF	0.58	0.52	0.07	4	4	4
Sampling uncertainty in EF	3.29	2.95	0.36	21	21	21
Emission factors (from the above 3 sources)	4.44	3.98	0.46	27	27	27
Activity data	4.10	7.20	8.07	25	49	472
All sources	6.09	8.54	8.46	37	58	492

In this example, uncertainties are high relative to the emission reduction. This is because in this example the reduction in emissions was small (1.7 megatons C/year, see Fig. 4.1 for a graphical explanation).
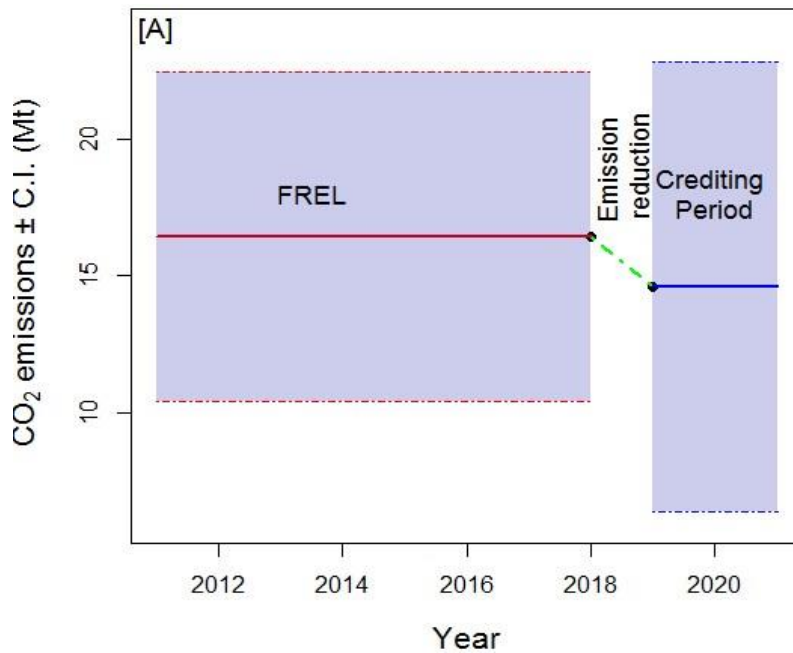
Figure 4.1. (a) Uncertainties are 37% of reference level emissions and 57% of the monitoring period emissions. (b) Because the emission reduction is small, the combined uncertainty is a large fraction of the emission reduction (>400%).

As shown in figure 4.2, uncertainties relative to the emission reduction will become smaller over time if the emission reduction increases over time, assuming that the contributing uncertainties are relatively constant over time (Neeff 2021).
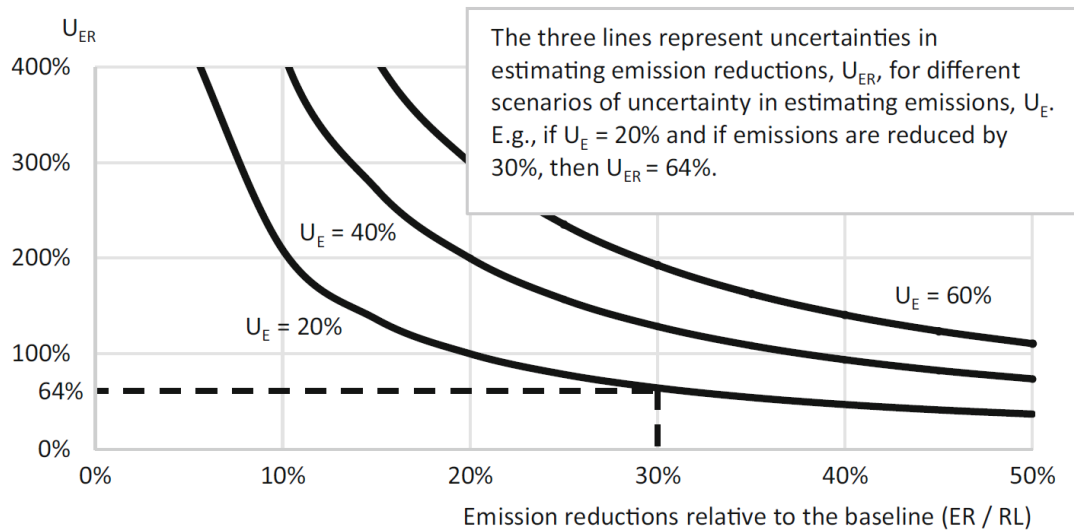
Figure 4.2. Uncertainty in estimating emission reductions by scenarios of effectiveness in reducing emissions below the reference level and by uncertainty in measuring emissions (from Neeff (2021), modified from FAO 2019)

Potential Pitfalls

If uncertainty estimates are not very accurate (based on a small number of Monte Carlo iterations), then by random variation, the uncertainty with a source turned off can be slightly higher than with the source turned on. Reporting each source turned on, rather than each source turned off, will avoid this problem. Increasing the number of Monte Carlo iterations makes the uncertainty estimates more precise (Figure 3.3).

Literature Cited

Efron, B., Tibshirani, R. J. 1994. An introduction to the bootstrap. *CRC press.*

FCPF 2016. Carbon fund methodological framework. *Forest Carbon Partnership Fund Report.*

Available at: https://www.forestcarbonpartnership.org/system/files/documents/FCPF

%20Carbon%20Fund%20Methodological%20Framework%20revised%202016_1.pdf

FCPF 2020. Guidelines on the application of the Methodological Framework Number 4 On Uncertainty Analysis of Emission Reductions. Forest Carbon Partnership Fund. Forest Carbon Partnership Facility Report. Available at: https://www.forestcarbonpartnership.org/sites/fcp/files/FCPF%20Guidelines%20on%20 Uncertainty%20Analysis_2020_0.pdf

IPCC 2006. IPCC guidelines for national greenhouse gas inventories. vol. 4 agriculture, forestry and other land use. *Prepared by the National Greenhouse Gas Inventories Programme.* S Eggleston, L Buendia, K Miwa, T Ngaraand K Tanabe. Institute for Global Environmental Strategies, Hayama, Japan. Available at: https://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html

IPCC 2019. Refinement to the 2006 IPCC guidelines for national greenhouse gas inventories. *Prepared by the National Greenhouse Gas Inventories Programme. G.* Domke, A.

Brandon, R, Diaz-Lasco, S. Federici, E. Garcia-Apaza, G. Grassi, et al. Institute for Global Environmental Strategies*,* Hayama, Japan*.* Available at: https://www.ipcc-nggip.iges.or.jp/public/2019rf/vol4.html

Neeff, T. What is the risk of overestimating emission reductions from forests – and what can be done about it?. *Climatic Change* 166, 26 (2021). https://doi.org/10.1007/s10584-021-03079-z

Pistilli, Tony. "Behind the models: cholesky decomposition." Medium, Towards Data Science, 23 May 2019, towardsdatascience.com/behind-the-models-cholesky-

decomposition-b61ef17a65fb.

Zaiontz, C. 2020. "Cholesky Decomposition". Real Statistics Using Excel, www.real-statistics.com/linear-algebra-matrix-topics/cholesky-decomposition/.

Annex 1. Monte Carlo Example in Excel

The simple example in Excel has two versions, one with a 2-year crediting period, and one with a 4-year crediting period, which illustrates the high uncertainty associated with a short crediting period.

Two forest types are included in the simple example, and deforestation is the only land-use transition. The example illustrates a deforestation rate of 3% per year in both forest types.

Introduction

The first sheet of the Excel workbook describes the roles of the subsequent sheets.

This workbook provides an example of Monte Carlo error propagation for estimating the uncertainty in emission reductions for REDD+.
The data are not for a real country; the example is simple to make it easy to understand.

Input Variables sheet:
Ten years of Activity data are used for this example: The reference period is composed of two periods of 7 and 3 years. The crediting period is 2 years.
There are three land-cover types: two forest types (FT1 and FT2) and non-forest (NF).
The parameters for emission factors are the aboveground biomass per unit area, which are specific for each forest type, and the root:shoot ratio and carbon fraction, which are shared for the two forest types.
For each uncertainty source in the AD and EF sheet, the standard error (SE) is back-calculated as: $SE = mean * uncertainty\ (90\ \%\ CI)/100\ /1.96$

AD sheet: Monte Carlo iterations for Activity Data.
EF sheet: Monte Carlo iterations for Emission Factors
The number of simulations is 10,000. Of these, 9,990 rows are hidden for ease in navigating the spreadsheet.

ER and Sensitivity Analysis sheet:
The AD and EF iterations are combined to estimate the Emission Reduction.
Conduct a sensitivity analysis by turning uncertainty sources on or off using switches, updating the Monte Carlo sampling, and copying the results into the table.

Input Variables

The Input Variables sheet has all the data needed for the calculation of emission reductions. Table 1 is activity data, or the . area of land converted from forest type 1 to non forest (FT1-NF) and the area of land converted from forest type 2 to non forest (FT2-NF) for the total 10 year time period: seven years of reference period and three years of monitoring period. Table 2 has emission factors including carbon fraction (CF), root to shoot ratio (R:S), and above ground biomass (AGB).

Uncertainties are given as the half width of the 90% CI. The SE is calculated from the uncertainty and the value of the estimate, although we recognize that in reality, the uncertainty is calculated from the SE.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													
3		Table 1. Activity Data and associated uncertainties					Table 2. Carbon Densities, Carbon Fraction, R:S and associated uncertainties						
4		Period	Convertion	AD	U				Value	U				
5		1	FT1-NF	739,643	34		CF	0.47	4				
6		(7 years)	FT2-NF	248,302	46		R:S	0.321	67				
7		2	FT1-NF	232,293	49		AGB_FT1	263	20				
8		(3 years)	FT2-NF	70,935	60		AGB_FT2	150	38				
9		3	FT1-NF	174,937	55		AGB_NF	26	29				
10		(2 years)	FT2-NF	54757	67								
11													
12													
13													
14													
15													
16													
17													
18													
19													

Introduction | **Input Variables** | AD | EF | ER and Sensitivity Analysis | ⊕

Activity Data (AD)

In the Activity Data sheet, forest transition data for each time period are simulated using Monte Carlo Simulation as: =NORM.INV(RAND(), B$5, B$7*B$1)

NORM.INV(probability, mean, sd) gives an inverse of the normal cumulative distribution, at the specified probability, mean, and standard deviation (sd).

RAND() generates uniform random normal values between 0-1.

B$5 is the mean of the value.

B$7 is the standard error.

B$1 is a multiplier controlled by a switch in the ER and Sensitivity Analysis sheet.

B9 f_x =NORM.INV(RAND(),B$5, B$7*B$1)

	A	B	C	D	E	F	G	H
1	Multiplier	1	1	1	1	1	1	
2		Period 1 (7 years)		Period 2 (3 years)		Period 3 (2 years)		
3		Deforestation		Deforestation		Deforestation		
4		FT1-NF	FT2-NF	FT1-NF	FT2-NF	FT1-NF	FT2-NF	
5	AD (ha)	739,643	248,302	232,293	70,935	174,937	54757	
6	U (%)	34	46	49	60	55	67	
7	SE	128305	58275	58073	21715	49090	18718	
8	Iterations							
9	1	839422	270794	222858	91907	102908	37920	
10	2	717687	306853	183361	61983	72253	52898	
11	3	548353	204995	243763	56959	126084	71099	
12	4	697787	196958	264925	80956	181057	91994	
13	5	663700	233860	134214	55382	122318	37432	
14	6	673559	229470	151467	98678	150497	46524	
10005	9997	713932	293956	205358	79225	156485	61649	
10006	9998	686580	157331	211198	44622	93772	66896	
10007	9999	718346	267479	271383	110406	212800	48783	
10008	10000	633855	227603	184761	62645	157602	56779	
10009								

Introduction | Input Variables | **AD** | EF | ER and Sensitivity Analysis

Emission Factors (EF)

In the Emission Factor Sheet, R:S, CF and AGB for each forest type are simulated using Monte Carlo simulation.

Belowground biomass (BGB) is calculated by multiplying each simulated AGB value with the simulated R:S value.

Total carbon for each land cover type is calculated as the sum of AGB and BGB.

Emission Factors (EF) are calculated as the difference in total carbon between land cover types. Column M depicts the transition from forest type 1 (FT1) to non forest (NF). The process is repeated for FT2-NF in column N.

The cells below row 10 are simulated values for each input variable. Multipliers in row 8 are controlled by switches on the ER and Sensitivity Analysis sheet.

Formula bar (B11): `=NORM.INV(RAND(),$B$3,$D$3*$B$8)`

	Value	U (%)	SE										
CF	0.47	4	0.01										
R:S	0.321	67	0.11										
AGB_FT1 (ton/ha	263	20	26.84										
AGB_FT2 (ton/ha	150	38	29.08										
AGB_NF (ton/ha	26	29	3.85										
	R:S Uncertainty	CF Uncertainty	Sampling Uncertainty			Sampling Uncertainty			Sampling Uncertainty				
Multiplier	1	1	1			1			1				
	Parameters		Simulated Carbon Densities: Forest			Simulated Carbon Densities:			Simulated Carbon Densities:			FE: FT1-NF	FE: FT2-NF
Iterations	R:S	CF	AGB	BGB	Total C (AGB+BGB)	AGB	BGB	Total C (AGB+BGB)	AGB	BGB	Total C (AGB+BGB)	C (AGB+BGB)	C (AGB+BGB)
1	0.32	0.48	256	82	162	92	30	58	18	6	11	151	47
2	0.41	0.46	261	106	170	138	56	90	28	11	18	151	72
3	0.18	0.46	167	30	90	128	23	69	27	5	15	75	54
4	0.32	0.48	295	94	186	198	63	125	28	9	18	168	107
5	0.14	0.48	290	42	159	155	22	85	20	3	11	148	74

Sheets: Introduction | Input Variables | AD | **EF** | ER and Sensitivity Analysis

ER and Simulation

Emissions for each forest transition are calculated by multiplying Activity Data (AD) from the AD sheet by Emission Factors (EF) in the EF sheet.

Formula bar (B4): `=AD!B8*EF!M11`

	A	B	C	
1	E			
2		Period 1		
3	Id Sim	FT1-NF	FT2-NF	
4	1	16904942	2507280	
5	2	9276812	1012690	
6	3	18355428	1141584	
7	4	14278623	1788756	
8	5	15418026	2178583	
9	6	14756366	2966326	
10000	9997	10967077	1086876	
10001	9998	15375514	2164624	
10002	9999	11635320	1090530	
10003	10000	15471274	1333783	
10004				
10005				
10006	Median	13463090.5	1716809.5	
10007	Percentile 5%	9143658.75	968009.35	1
10008	Percentile 95%	18660940.4	2709561.9	2
10009	Half of confidence interval	4758640.825	870776.275	6(

Sheets: Instruction | Data | AD | EF | **ER and Sensitivity Analysis**

1. In row 10006, the median of all simulations for each forest type is calculated as

    =MEDIAN(B4:B10003)

2. In row 10007, the 5% percentile of all simulations for each forest type is calculated as

=PERCENTILE(B4:B10003, 0.05)

3. In row 10008, the 95% percentile of all simulations for each forest type is calculated as

   =PERCENTILE(B4:B10003, 0.95)

4. In row 10009, the half width of the 90% confidence interval is calculated as: (95 percentile-5 percentile)/2

   i.e. (B10008-B10007)/2

5. Finally in row 10010, uncertainty is calculated as: (half of 90% confidence interval/median)*100

   i.e. (B10009/B10006)*100

B10006	fx	=MEDIAN(B4:B10003)				
	A	B	C	D	E	F
1	E					
2			Period 1		Period 2		Pe
3	Id Sim	FT1-NF	FT2-NF	FT1-NF	FT2-NF	FT1-NF
4	1	12579300	2145060	19711800	5966694	718
5	2	11624181	2053586	18192807	5124254	708
6	3	12199296	799607	14862848	2329076	493
7	4	15251342	1491441	14820434	4445804	845
8	5	12930120	1401435	16500603	3238515	951
9	6	11422763	1956069	15082771	4544586	821
10000	9997	14190176	913164	20319442	1728174	745
10001	9998	17154368	1587364	22550704	3345243	744
10002	9999	12808888	1970360	17565880	3586100	959
10003	10000	11801514	1840880	15995044	3159120	773
10004						
10005						
10006	Median	13427966	1717174.5	17394955	4132576.5	823
10007	Percentile 5%	9187692.6	962025.2	11866981.5	2331686.6	56404
10008	Percentile 95%	18695230.75	2662617.65	23989526.25	6419677.6	114040
10009	Half of confidence interval	4753769.075	850296.225	6061272.375	2043995.5	288179
10010	Uncertainty	35.4	49.52	34.85	49.46	3
10011						

Switches

In the image below, the final uncertainty of Emission Reductions (cell X10009) is highlighted in blue. When the switches in cells D10012:D10017 are 'on' the related multipliers are set to 1. When the switches are 'off', the multipliers are set to 0 (or 1E-10, which is close to 0, and avoids errors in the excel formula). The uncertainties change slightly every time the workbook updates and the random values are resampled.

	A	B	C	D	E	F	G	H	I	J
10009	Uncertainty (%)	38	58	48	67	53	74	31	47	414.55
10010										
10011			Parameter	Switches						
10012			R:S Uncertainty	on						
10013			CF Uncertainty	on						
10014			Sampling Uncertainty (AGB)							
10015			AGB of Forests	on						
10016			Activity Data Uncertainty							
10017			Activity Data	on						
10018										
10019	Sensitivity analysis									

		Uncertainty (%)	Parameter	Uncertainty with one turned off (%)	Change from all Parameters	Uncertainty with one turned on (%)
10021	All on	417.94	R:S Uncertainty	413.74	4.2	13.58
10022	All off	0	CF Uncertainty	414.59	3.35	3.49
10023			Sampling uncertainty	393.27	24.67	17.11
10024			Emission Factor uncertainty	395.47	22.47	22.38
10025			Activity Data	22.38	395.56	395.47

Introduction | Input Variables | AD | EF | **ER and Sensitivity Analysis** | (+)

This image depicts what happens when R:S uncertainty is turned off in the EF tab.

B11   fx   =ROUND(NORM.INV(RAND(),$B$3,$C$3*$B$8),2)

	A	B	C	D	E	F	G
6	AGB_NF (	26	3.8	29		BGB_NF (ton/ha)	8.346
7		R:S Uncertainty	CF Uncertair	Sampling Uncertainty			Sampling
8	Switch	1E-10	1	1			1
9		Parameters		Simulated Carbon Densities: Forest Type 1			Simulate
10	Iterations	R:S	CF	AGB	BGB	Total C (AGB+BGB)	AGB
11	1	0.32	0.47	272	87	169	140
12	2	0.32	0.47	269	86	167	135
13	3	0.32	0.47	278	89	172	144
14	4	0.32	0.49	253	81	164	168
15	5	0.32	0.48	270	86	171	137
16	6	0.32	0.46	274	88	167	148
10007	9997	0.32	0.47	241	77	149	122
10008	9998	0.32	0.47	250	80	155	109
10009	9999	0.32	0.48	257	82	163	202
10010	10000	0.32	0.46	201	64	122	178
10011							

Instruction | Data | AD | **EF** | ER and Sensitivity Analysis | (+)

The contribution of each source to the overall uncertainty can be determined by starting with all sources on and turning off each source one by one, or by starting with all sources off and turning them on one by one. The sensitivity analysis table is populated with the values copied from cell J1009 with different combinations of switches on or off.

	A	B	C	D	E	F
10011			Parameter	Switches		
10012			R:S Uncertainty	on		
10013			CF Uncertainty	on / off		
10014			**Sampling Uncertainty (AGB)**			
10015			AGB of Forests	on		
10016			**Activity Data Uncertainty**			
10017			Activity Data	on		
10018						
10019	Sensitivity analysis					
10020		Uncertainty (%)	Parameter	Uncertainty with one turned off (%)	Change from all Parameters	Uncertainty with one turned on (%)
10021	All on	417.94	R:S Uncertainty	413.74	4.2	13.58
10022	All off	0	CF Uncertainty	414.59	3.35	3.49
10023			Sampling uncertainty	393.27	24.67	17.11
10024			Emission Factor uncertainty	395.47	22.47	22.38
10025			Activity Data	22.38	395.56	395.47

Introduction | Input Variables | AD | EF | **ER and Sensitivity Analysis** | ⊕

Annex 2. Monte Carlo Example in R

https://github.com/mark78b/MCS_in_R/

The Simple Example in R is the same as that in Excel, showing deforestation rates of two forest types.

---

1. **Library loading**

   To run the current R code, load the following libraries:

   ```
   library(matrixStats)
   library(gridExtra)
   library(reshape)
   ```

2. **Reading of inputs by Forest Type and Period**

   Read the specific values of AD and EF (and associated uncertainties) by Forest Type (FT) and Period (P). Read csv Tables of AD and EF as follows:

   ```
   rm(list=ls(all=TRUE))

   ### Address to read inputs
   setwd("C:/Desktop/Inputs")

   ### Reading of inputs
   # AD inputs
   BaseAD <- read.csv("1_Activity_Data.csv")
   dim(BaseAD)

   BaseAD_4yr <- read.csv("1_Activity_Data_4yr.csv")
   dim(BaseAD_4yr)

   # EF inputs
   BaseEF <- read.csv("2_Emission_Factors.csv")
   dim(BaseEF)
   ```

3. **Simulations of random numbers of AD and EF**

   A key step of MCS is the generation of random numbers of AD and EF. To generate random numbers it is necessary: (i) the estimator of AD and EF and, (ii) the standard error (SE) of AD

---

and EF estimators. Nevertheless, in most cases, only uncertainties of AD and EF are reported. So it is necessary to compute SE of AD and EF as following:

$$U = \frac{1/2\ IC}{\hat{\theta}}\ x\ 100 \quad \Rightarrow \quad U = \frac{1.96\ x\ \sigma}{\hat{\theta}}\ x\ 100 \quad \Rightarrow \quad \sigma = \frac{U\ x\ \hat{\theta}}{1.96\ x\ 100}$$

$$U = \frac{1/2\ IC}{\hat{\theta}}\ x100 = \frac{1.96\ x\ \sigma}{\hat{\theta}}\ x100 \Rightarrow \sigma = \frac{U\ x\ \hat{\theta}}{100}$$

U = uncertainty

IC= confidence Interval

θ= estimator of AD or EF

σ= standard deviation

Once SE of AD and EF have been computed, it is possible (i) to generate vectors of random numbers of AD and EF per FT-P, (ii) to estimate emissions per FT, and (iii) to save vectors of simulated emissions by FT in matrix format.

Following is show how to simulate random numbers (of normal distribution, using mean and standard error) of independient AD and partially correlated EF:

#################### 1. Simulation of Activity Data   ###################

Computing of SD of AD
BaseAD$DesEstDA<-abs((BaseAD$U_AD_per*BaseAD$AD_ha)/(1.65*100))

Number of simulations
n<-1000000
MatrizDef<-seq(1:n)

Switch for AD if it is 1 incorporates MC and if it is 0 then use simple value
SWITCH_AD =1

Simulations of AD per period and conversions
for (i in 1:length(BaseAD$Code))
{
  DAsim<-rnorm(n,mean=BaseAD$AD_ha[i], sd=BaseAD$DesEstDA[i]*SWITCH_AD)
  MatrizDef<-cbind(MatrizDef,DAsim)
}

Conversion of AD Matrix to DataFrame

```
Matrix_AD<-as.data.frame(MatrizDef)


### Correct colnames of AD-Dataframe per "Period" and "Transition"
colnames(Matrix_AD) = c("Id_sim_AD",
                "AD_P1_FT1_NF","AD_P1_FT2_NF",
                "AD_P2_FT1_NF","AD_P2_FT2_NF",
                "AD_P3_FT1_NF","AD_P3_FT2_NF")


################## 2. Simulation of Emissions Factors ####################


### Simulations and CF, R:S and AGB  per Forest Type and Transition
### Matrix for saving simulated CF, R:S and AGB
MatrizEF_Def <-seq(1:n)
### Switches for EF
SWITCH_EF =1


### Simulations of CF, R:S and AGB
for (i in 1:length(BaseEF$Value))
{
  EF_Sim<-rnorm(n,mean=BaseEF$Value[i], sd=BaseEF$SE[i]*SWITCH_EF)
  MatrizEF_Def<-cbind(MatrizEF_Def,EF_Sim)
}


### Converting the "CF, R:S and AGB" Matrix to DataFrame
Matrix_EF<-as.data.frame(MatrizEF_Def)


### Correct colnames of "CF, R:S and AGB" Dataframe
colnames(Matrix_EF) = c("Id_sim_EF","CF","Root_S","AGB_FT1","AGB_FT2","AGB_NF")


#######################################################################
### Simulation of BGB, Carbon Densities and EF per Transition


### Simulation of BGB per stratum
Matrix_EF$BGB_FT1 <-  Matrix_EF$AGB_FT1 * Matrix_EF$Root_S
Matrix_EF$BGB_FT2 <-  Matrix_EF$AGB_FT2 * Matrix_EF$Root_S
Matrix_EF$BGB_NF  <-  Matrix_EF$AGB_NF  * Matrix_EF$Root_S


### Simulation of Carbon Densities per stratum
Matrix_EF$C_FT1  <- (Matrix_EF$AGB_FT1 + Matrix_EF$BGB_FT1 ) * Matrix_EF$CF
```

```r
Matrix_EF$C_FT2  <- (Matrix_EF$AGB_FT2 + Matrix_EF$BGB_FT2 ) * Matrix_EF$CF
Matrix_EF$C_NF   <- (Matrix_EF$AGB_NF  + Matrix_EF$BGB_NF  ) * Matrix_EF$CF


### Simulation of EF per Transition
Matrix_EF$EF_FT1_NF <- Matrix_EF$C_FT1 - Matrix_EF$C_NF
Matrix_EF$EF_FT2_NF <- Matrix_EF$C_FT2 - Matrix_EF$C_NF


### Filtering of simulated EF per Transition
Matrix_EF1<-data.frame(Id_sim_EF =Matrix_EF$Id_sim_EF,
              EF_FT1_NF =Matrix_EF$EF_FT1_NF,
              EF_FT2_NF =Matrix_EF$EF_FT2_NF)
length(Matrix_EF1$Id_sim_EF)
```

4. **Estimation of simulated emissions**

   Using the AD and EF simulated, emissions per Land Use transition and period are estimated:

```r
### Merge of "AD-Dataframe" and "EF-Dataframe"
Table_Emi<- merge(Matrix_AD, Matrix_EF1, by.x = "Id_sim_AD", by.y = "Id_sim_EF",all=T)


### Estimation of Emission per Period and Transition annualized
yearP1=7
yearP2=3
yearP3=2


Table_Emi$Emi_P1_FT1_NF  <- Table_Emi$AD_P1_FT1_NF * Table_Emi$EF_FT1_NF
Table_Emi$Emi_P1_FT2_NF  <- Table_Emi$AD_P1_FT2_NF * Table_Emi$EF_FT2_NF
Table_Emi$Emi_P2_FT1_NF  <- Table_Emi$AD_P2_FT1_NF * Table_Emi$EF_FT1_NF
Table_Emi$Emi_P2_FT2_NF  <- Table_Emi$AD_P2_FT2_NF * Table_Emi$EF_FT2_NF
Table_Emi$Emi_P3_FT1_NF  <- Table_Emi$AD_P3_FT1_NF * Table_Emi$EF_FT1_NF
Table_Emi$Emi_P3_FT2_NF  <- Table_Emi$AD_P3_FT2_NF * Table_Emi$EF_FT2_NF
dim(Table_Emi)


####################################################################
########## 3.1 Simulation of Emissions of Base Line  (2 Periods,7 and 3 years)  ####


### Selecting of emissions for Base Line period
Table_Emi_FREL<- Table_Emi[, c(1,10:13)]
Table_Emi_FREL$FREL<- ( rowSums(Table_Emi_FREL[,c(2:5)]) ) / (yearP1+yearP2)
```

5. Estimation of quantiles and uncertainties

Using vectors of simulated emissions it is possible to estimate associated lower/upper uncertainties as follows:

$$U\_\text{inf} = \frac{\left|\, Q(0.025)\text{-}\hat{E}\,\right|}{\hat{E}} \text{ x } 100 \qquad \text{and} \qquad U\_\text{sup} = \frac{\left|\, Q(0.975)\text{-}\hat{E}\,\right|}{\hat{E}} \text{ x } 100$$

U_inf= Uncertainty on the left side of the simulated emissions

U_sup= Uncertainty on the right side of the simulated emissions

Q (0.025)= quantile 0.025 of simulated emissions

Q (0.975)= quantile 0.975 of simulated emissions

E = Monte Carlo simulated emissions

(i) so, lower/upper quantiles of Base Line emissions can be estimated as:
Q_05_FREL <-quantile(Table_Emi_FREL$FREL,0.05)[[1]]
Q_95_FREL <-quantile(Table_Emi_FREL$FREL,0.95)[[1]]

(ii) also, lower/upper uncertainties of emissions Base Line can be estimated as:
half_CI <- (Q_95_FREL - Q_05_FREL)/2
U_FREL<-abs(half_CI / median(Table_Emi_FREL$FREL))*100

6. Saving of quantiles and uncertainties

Quantiles and uncertainties of emissions by period and median emission are saved:

Saving of Base Line emissions and associated quantiles and Uncertainties
Table_FREL<-data.frame(Period = "FREL",
                       Emission = median(Table_Emi_FREL$FREL),
                       Q_05 = round(Q_05_FREL , digits = 3),
                       Q_95 = round(Q_95_FREL , digits = 3),
                       Uncertainty = round(U_FREL, digits = 2))

7. Probability density function of emissions by period and Base Line

```
### Saving of PDF of Base Line emissions
setwd("C:/Users/Invited/Outputs")


pdf("1_Emissions_Uncertainty_FREL.pdf")
par(mfrow=c(2,1))
hist(Table_Emi_FREL$FREL,
    main="Histogram of FREL",
    xlab="Average annual emissions from deforestation (Ton of CO2e)",
    cex.lab=1, cex.axis=0.8, cex.main=1,
     #border="blue",
    col="green",
    las=1,
    breaks = 200,
    prob = TRUE)
lines(density(Table_Emi_FREL$FREL ))
dev.off()
```

A draft version of the R code shown above is available at the following link:
https://github.com/mark78b/MCS_in_R/

Document information

Version	Date	Description
1.0	21 September 2021	Initial version