



**Note d'orientation sur l'estimation de l'incertitude des
réductions d'émissions à l'aide de la simulation de
Monte-Carlo**

Version 1.0
Septembre 2021

Sommaire

Étape 1. Identifier les sources des valeurs utilisées dans les estimations de réduction d'émissions et déterminer si elles sont indépendantes ou partagées.	2
Étape 2. Identifier l'incertitude associée à chacune de ces variables.	5
Étape 3. Propager les incertitudes dans l'estimation des réductions d'émissions en utilisant la simulation de Monte Carlo	8
Étape 4. Évaluer la contribution de chaque source à l'incertitude globale.....	16
Annexe 1. Exemple de Monte Carlo dans Excel.....	21
Annexe 2. Exemple de Monte Carlo en R.....	28

Note d'orientation sur l'estimation de l'incertitude des réductions d'émissions à l'aide de la simulation de Monte-Carlo

Cette note d'orientation a été préparée par le projet QUERCA (Quantifying Uncertainty Estimates and Risk for Carbon Accounting) du SUNY College of Environmental Science and Forestry grâce à un financement du FCPF.

Le critère 9 du cadre méthodologique du FCPF exige que l'incertitude de l'estimation des réductions d'émissions soit quantifiée en utilisant des méthodes Monte Carlo. Les lignes directrices du FCPF sur l'analyse de l'incertitude des réductions d'émissions décrivent les sources d'incertitude à propager et fournissent des conseils pour effectuer des simulations de Monte Carlo.

Les Programmes ER diffèrent à la fois dans les activités qu'ils considèrent et dans les méthodes par lesquelles ils calculent les facteurs d'émission et les données d'activité. Les sources d'incertitude sont donc différentes, et les calculs pour les combiner correctement seront également différents. Ce document décrit l'approche générale et fournit un exemple simple pour illustrer l'approche. L'approche générale décrite ici comprend les étapes suivantes :

- Étape 1. Identifier les sources des valeurs utilisées dans les estimations de réduction des émissions et déterminer si elles sont indépendantes ou partagées.
- Étape 2. Identifier l'incertitude associée à chacune de ces variables
- Étape 3. Propager les incertitudes dans l'estimation des réductions d'émissions en utilisant la simulation de Monte Carlo
- Étape 4. Évaluer la contribution de chaque source à l'incertitude globale

L'exemple simple a été fourni en Excel et en R pour aider les utilisateurs à comprendre ces conseils et chacune des étapes. Chaque situation est unique, et les exemples donnés devront être adaptés aux programmes en question, mais les principes sous-jacents sont universels. L'annexe 1 contient plus de détails sur l'exemple.

Cette note d'orientation complète les directives du FCPF sur l'analyse des incertitudes.

Étape 1. Identifier les sources des valeurs utilisées dans les estimations de réduction d'émissions et déterminer si elles sont indépendantes ou partagées.

En suivant le processus par lequel les programmes d'ER ont estimé les réductions d'émissions, les programmes doivent identifier toutes les variables utilisées dans l'estimation des émissions et des absorptions. Le [Tableau 1](#) des directives du FCPF sur l'analyse des incertitudes fournit une liste des principales sources d'incertitude qui, au minimum, doivent être évaluées. Dans la simulation de Monte Carlo, les incertitudes de ces variables seront combinées par échantillonnage à partir des distributions probables de leurs valeurs, comme déterminé à l'étape 2 ci-dessous.

Pour combiner correctement les incertitudes de plusieurs variables, il faut comprendre quelles variables sont utilisées indépendamment et lesquelles sont partagées entre plusieurs calculs.

Les variables et leurs sources d'incertitude associées contribuent de manière indépendante à un calcul particulier si elles sont dérivées indépendamment et ne sont pas utilisées pour une autre variable. Par exemple, les données d'inventaire des arbres sont généralement collectées indépendamment pour chaque strate ou type de couverture terrestre et ne sont pas utilisées

dans les calculs pour les autres strates ou types de couverture terrestre (en bleu dans la Figure 1.1).

D'autres variables et les incertitudes qui leur sont associées sont partagées entre plusieurs calculs. Par exemple, la fraction de carbone (FC), le ratio racines:pousses (R:S) et les variables d'allométrie des arbres peuvent être utilisés pour plusieurs types de forêts (en rouge sur la Figure 1.1). Dans ce cas, les densités de carbone sont calculées par une combinaison de sources indépendantes et partagées et sont donc partiellement corrélées (en violet dans la Figure 1.1). Il est important de représenter correctement ces corrélations lorsqu'on les combine dans la propagation des erreurs.

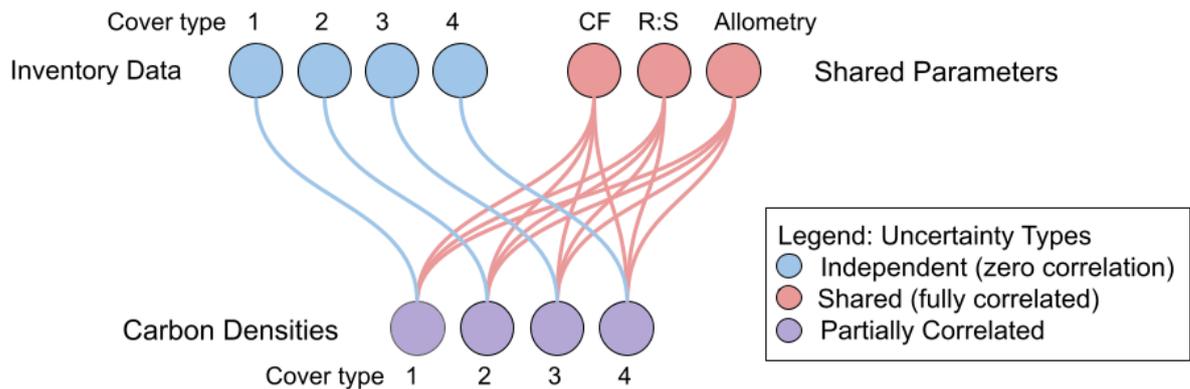
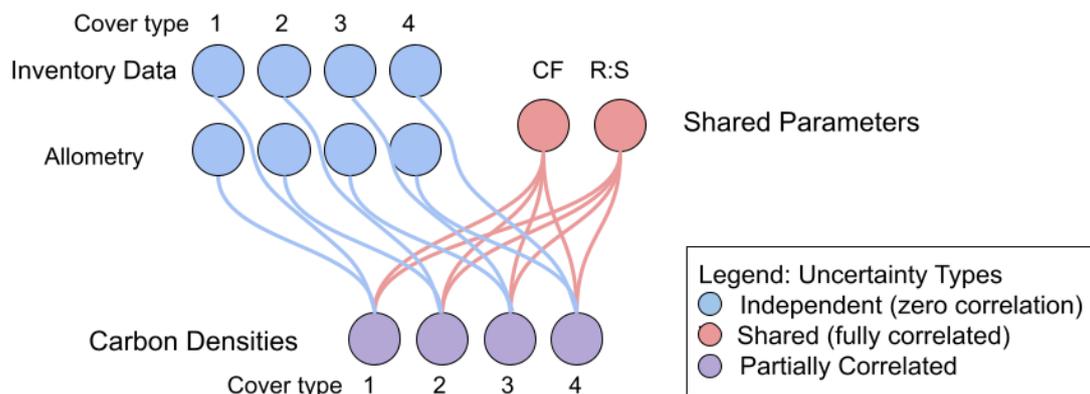


Figure 1.1 Dans le calcul des densités de carbone, les données d'inventaire recueillies dans quatre types de couverture terrestre sont indépendantes (en bleu). Si les mêmes valeurs de fraction de carbone (FC), de ratio racines:pousses (R:S) et d'allométrie des arbres sont utilisées dans plusieurs types de forêts, elles sont partagées (en rouge). Les incertitudes calculées à partir d'une combinaison de sources indépendantes et partagées seront partiellement corrélées (en violet), avec des coefficients de corrélation intermédiaires entre 0 (totalement indépendant) et 1 (totalement partagé).

Le fait qu'une source d'incertitude doive être traitée comme partagée ou indépendante dépend de la façon dont elle a été collectée et de la façon dont elle est utilisée dans le calcul. Certains programmes collectent les données d'allométrie des arbres indépendamment pour chaque type de couverture terrestre. Dans ce cas, les incertitudes de l'allométrie des arbres seraient indépendantes pour chaque type de couverture (Figure 1.2). Théoriquement, la fraction de carbone et les ratios racines:pousses pourraient être déterminés indépendamment pour chaque type de couverture, ou une valeur unique pourrait être utilisée pour tous les types de forêts. La conversion du carbone en CO₂ est traitée comme une constante sans incertitude, car la variabilité des rapports isotopiques du carbone et de l'oxygène est négligeable.



Le degré de corrélation partielle (c.-à-d. la valeur du coefficient de corrélation) qui résulte de la combinaison d'un mélange de sources d'incertitude partagées et indépendantes peut être estimé analytiquement (non couvert dans ce guide) ou en examinant les résultats de la simulation de Monte Carlo. Il n'est pas important de connaître le coefficient de corrélation des variables intermédiaires dans le calcul. Cependant, si le calcul devait commencer avec des entrées partiellement corrélées, telles que les densités de carbone, les coefficients de corrélation devraient être estimés pour être propagés correctement (voir la Section 3.2.3).

Pièges potentiels

Traiter les sources d'incertitude comme indépendantes alors qu'elles sont partagées sous-estime la véritable incertitude combinée. Dans l'exemple simple fourni dans l'Annexe 1, l'incertitude de l'ER devrait être de 416 %. Si les facteurs d'émissions sont traités comme indépendants entre le niveau de référence et la période de surveillance, au lieu d'être partagés, on obtient une valeur de 403 %, ce qui est incorrect.

Étape 2. Identifier l'incertitude associée à chacune de ces variables.

L'utilisation de l'approche de Monte Carlo pour la propagation des erreurs nécessite de définir les distributions des variables utilisées dans le calcul. Il existe deux façons de générer des échantillons aléatoires qui imitent la distribution probable d'une variable. La première consiste à échantillonner à partir d'une distribution définie (une fonction de densité de probabilité ou PDF). La seconde consiste à échantillonner de manière aléatoire les valeurs de la variable à partir d'un ensemble de données (bootstrapping).

2.1 Choisir PDF ou bootstrap

Il est plus facile d'échantillonner à partir d'une distribution définie que d'échantillonner à partir d'un ensemble de données, surtout dans Excel. L'échantillonnage à partir d'un ensemble de données présente l'avantage de ne nécessiter aucune hypothèse sur la nature de la distribution. Si la distribution n'est pas normale, le bootstrap sera plus précis, à moins que les données ne soient pas représentatives. La figure ci-dessous fournit un arbre décisionnel simple pour choisir entre PDF et bootstrap pour générer des échantillons aléatoires qui imitent la distribution probable d'une variable.

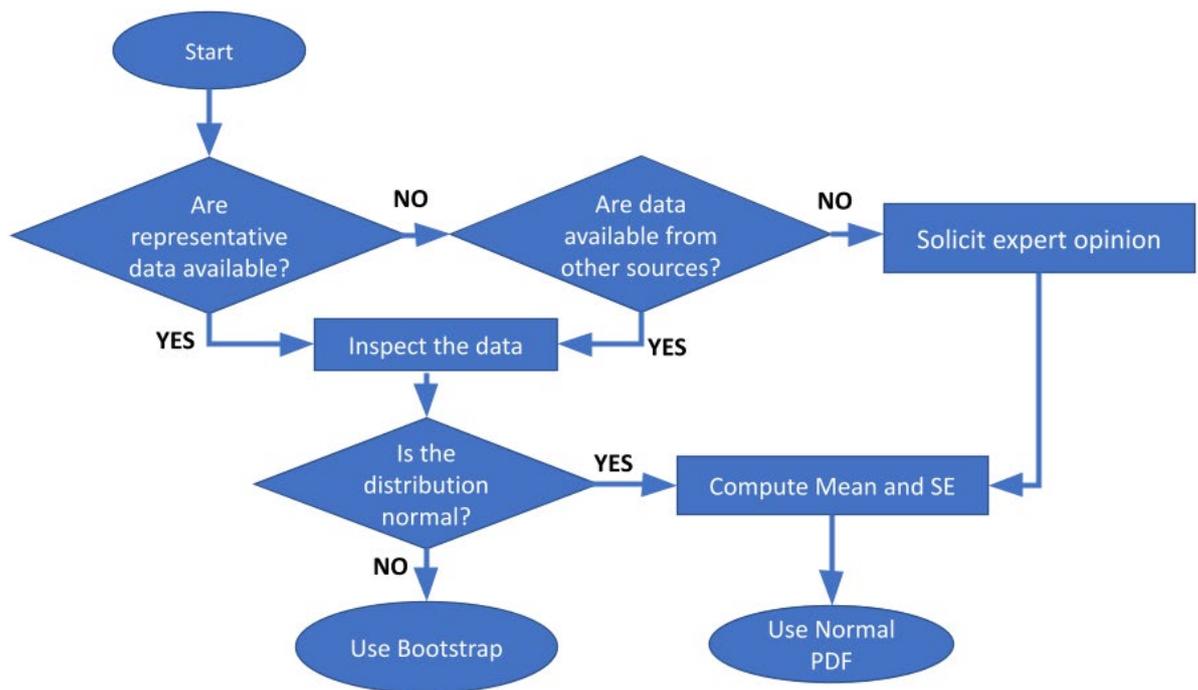


Figure 2.1. Arbre décisionnel pour choisir d'échantillonner à partir de données ou d'une PDF.

2.2 Décrire l'incertitude avec une fonction de densité de probabilité (PDF)

Une fonction mathématique peut être utilisée pour décrire la distribution des valeurs possibles d'une variable. En général, en l'absence d'informations contraires, on utilise des distributions normales. Une distribution autre que normale peut être choisie si les valeurs possibles ne sont pas distribuées normalement. Pour certaines applications, les distributions bêta, binomiale, gamma, weibull ou lognormale décrivent le mieux la distribution des observations. Cependant, s'il y a suffisamment de données pour déterminer que la distribution n'est pas normale, le bootstrapping est également une option.

Dans le cadre de la comptabilisation du carbone REDD+, des distributions uniformes ont parfois été utilisées pour décrire les valeurs possibles de la fraction de carbone ([Tableau 4.3 du GIEC 2006](#)) et du ratio racine:pousse ([Raffinement 2019 du GIEC 2006, Tableau 4.4](#)). Pour cette raison, nous fournissons des conseils sur la façon d'utiliser une PDF uniforme. Il semble néanmoins peu plausible qu'il existe une probabilité nulle d'une valeur en dehors de ces plages. Nous recommandons plutôt d'utiliser la moyenne et l'erreur standard des données disponibles pour définir une distribution normale. Alternativement, les valeurs pourraient être échantillonnées par bootstrapping.

En l'absence de données fiables, un jugement d'expert peut être utilisé pour définir une PDF. Les directives du FCPF sur l'analyse d'incertitude des réductions d'émissions recommandent de consulter indépendamment au moins trois experts lorsque l'estimation des paramètres n'est pas disponible ou n'est pas représentative (par exemple, basée sur des parcelles de recherche). La moyenne et l'erreur standard de la moyenne des opinions des experts devraient être utilisées pour définir une distribution normale. L'utilisation de la fourchette serait sensible aux valeurs extrêmes, et doubler la fourchette (comme le recommandent actuellement les lignes directrices) gonflerait l'incertitude des réductions d'émissions. Alternativement, les valeurs pourraient être échantillonnées par bootstrapping.

2.3 Utilisation de la distribution des données (bootstrapping)

Une alternative à la représentation analytique de la distribution des entrées est de ré-échantillonner les données, une procédure connue sous le nom de bootstrapping (Ephron et Tibshirani 1994) qui est illustrée à la Figure 2.2. Les valeurs sont tirées au hasard des données pour créer des ensembles de données alternatifs possibles avec des distributions similaires et le même nombre d'observations. Chaque échantillon aléatoire est tiré de tous les échantillons possibles (c'est ce qu'on appelle « l'échantillonnage avec remplacement ») car l'échantillonnage sans remplacement, s'il tire le nombre d'observations de l'ensemble de données, donnerait à chaque fois l'ensemble de données original. Cette approche ne nécessite aucune hypothèse de distribution et est donc plus fidèle à la population mesurée. Le bootstrap est particulièrement avantageux lorsque la distribution est difficile à définir. L'inconvénient de cette approche est que la qualité de la représentation de la population dépend entièrement de celle des données, et si l'ensemble de données est petit, il peut ne pas capturer avec précision la gamme des valeurs potentielles.

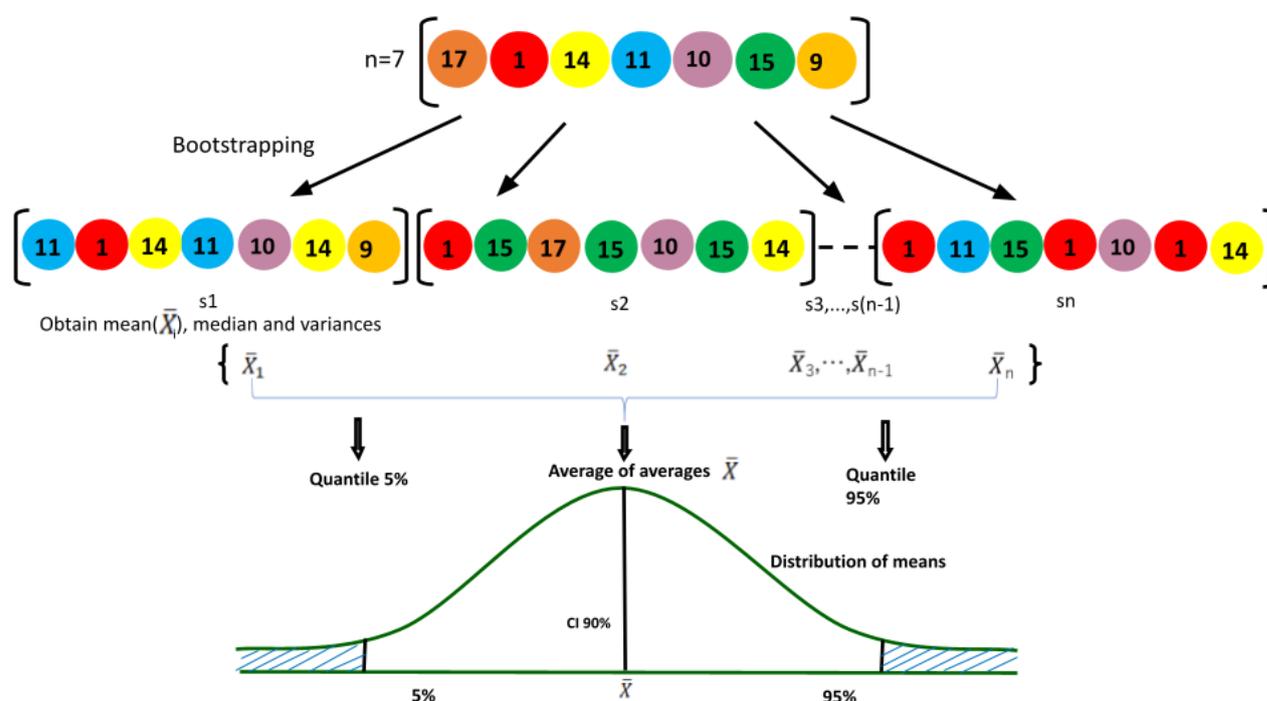


Figure 2.2. Échantillonnage aléatoire de valeurs à partir d'un ensemble de données. L'ensemble de données comporte 7 valeurs, qui peuvent être échantillonnées au hasard (avec remplacement) pour créer d'autres ensembles de données possibles.

Pièges potentiels

L'échantillonnage bootstrap doit être configuré pour correspondre au plan d'échantillonnage. Par exemple, si le plan d'échantillonnage est stratifié, il sera nécessaire d'effectuer un bootstrap pour chaque strate séparément et de produire les estimations globales en combinant les informations des strates. Des problèmes similaires peuvent se poser pour l'échantillonnage en grappes.

Souvent, les données ne sont pas représentatives, en raison de l'accès aux lieux d'échantillonnage (près des routes ou des lisières de forêt), ou de la facilité de mesure (excavation des racines de petits arbres). Il s'agit d'un piège potentiel tant pour la

caractérisation des données avec une PDF que pour le bootstrap. Dans ces situations, un jugement d'expert doit être utilisé pour corriger le biais dans les données.

Étape 3. Propager les incertitudes dans l'estimation des réductions d'émissions en utilisant la simulation de Monte Carlo

Le FCPF exige l'utilisation de la simulation de Monte Carlo pour quantifier les effets des entrées incertaines sur l'incertitude des réductions d'émissions de carbone. En utilisant cette approche, le calcul des réductions d'émissions est itéré des centaines ou des milliers de fois, avec des entrées variant de manière aléatoire pour imiter les incertitudes dans les valeurs de toutes les variables qui ont été utilisées dans le calcul du niveau de référence et des émissions et absorptions surveillées. Comme décrit ci-dessus à l'étape 2, les échantillons aléatoires qui imitent la distribution probable d'une variable peuvent être générés en utilisant une distribution définie (une PDF) ou en échantillonnant de manière aléatoire les valeurs de la variable à partir d'un ensemble de données (bootstrapping). La distribution des centaines ou des milliers de sorties qui en résultent reflète les effets nets des incertitudes des entrées.

3.1 Comment échantillonner de façon aléatoire à partir d'une distribution ou d'un ensemble de données ?

3.1.1 Échantillonnage à partir d'une distribution

Pour simuler l'incertitude d'une entrée en utilisant une distribution définie, les paramètres décrivant la distribution sont utilisés pour générer les échantillons aléatoires.

Distribution	Formules de tableur	Code R
Uniforme	<p>RAND() génère des nombres aléatoires uniformément distribués entre 0 et 1.</p> <p>Générer des nombres aléatoires uniformes entre les valeurs de A1 et A2 en utilisant :</p> <p>= (\$A\$1 + RAND()*(\$A\$2-\$A\$1))</p>	<p>Générer des nombres aléatoires uniformes entre a1 et a2 en utilisant :</p> <p>runif(n, min = a1, max = a2)</p>
Normal	<p>NORM.INV(probabilité, moyenne, sd) donne l'inverse de la distribution cumulative normale, à une probabilité, une moyenne et un écart-type (sd) spécifiés.</p> <p>En combinant RAND et NORM.INV, on obtient un échantillon aléatoire d'une distribution dont la moyenne et l'écart-type sont spécifiés. Si la moyenne et l'écart type sont dans les cellules B1 et B2, utiliser :</p> <p>=NORM.INV(RAND(), \$B\$1, \$B\$2)</p>	<p>Étant donné la moyenne et l'écart-type (sd) de la distribution, générer n échantillons aléatoires en utilisant :</p> <p>rnorm(n, moyenne, sd)</p>

3.1.2 Échantillonnage à partir d'un ensemble de données (bootstrapping)

Pour simuler l'incertitude d'une entrée à l'aide du bootstrap, chaque échantillon aléatoire est sélectionné dans l'ensemble de données des entrées possibles.

Formule de tableur	Code R
<p>Si 10 points de données se trouvent dans les cellules A1:A10, on peut en échantillonner 1 au hasard en utilisant :</p> <p>=INDEX(\$A\$1:\$A\$10,LIGNES(\$A\$1:\$A\$10)*RAND()+1, COLONNES(\$A\$1:\$A\$10)*RAND()+1)</p> <p>Comme cette formule tire une seule valeur au hasard de l'ensemble de données spécifié, vous devez la copier plusieurs fois pour générer un échantillon bootstrap. Vous devez copier cette formule pour échantillonner de nombreuses observations et la copier pour autant d'itérations que nécessaire.</p>	<p>Étant donné un échantillon de taille n : (c1,c2,c3,...,cn), on peut générer un vecteur :</p> <p>ÉchantillonC<-c(c1,c2,c3,...,cn)</p> <p>En définissant un nombre de simulations (NS), vous pouvez générer des échantillons bootstrap (BootstrapSam) en générant des échantillons NS de taille n échantillonnés du vecteur « SampleC » avec remplacement :</p> <p>BootstrapSam <- dupliquer(NS, échantillon(ÉchantillonC, remplacer = T))</p>

3.2 Comment échantillonner les sources multiples d'incertitude

Pour tenir compte correctement des sources d'incertitude indépendantes ou partagées (identifiées à l'étape 1, ci-dessus), il faut attribuer les échantillons aléatoires des valeurs d'entrée de façon indépendante, dans le cas d'échantillons indépendants (3.2.1), ou attribuer la même valeur aléatoire dans tous les calculs qui partagent cette valeur, dans le cas de sources partagées (3.2.3), à chaque itération de la simulation de Monte Carlo.

3.2.1 Comment attribuer des valeurs aléatoires indépendantes ?

Dans le cas de variables d'entrée indépendantes, les valeurs aléatoires seront sélectionnées indépendamment pour chaque variable afin de représenter l'incertitude de ces variables. Par exemple, les données d'activité sont collectées indépendamment à chaque point dans le temps.

Formule de tableur	Code R

<p>Pour les sources indépendantes, les cellules référencées pour les paramètres (par exemple, moyenne et sd dans le cas d'une source normalement distribuée) ne sont pas les mêmes :</p> <p>SourceA : A3=NORM.INV(RANDARRAY(1,n),\$A\$1,\$A\$2)</p> <p>SourceB : B3=NORM.INV(RANDARRAY(1,n),\$B\$1,\$B\$2)</p> <p>La cellule A4 aurait un nombre aléatoire différent référencé par tous les calculs de la ligne 4, et ainsi de suite pour toutes les lignes de la simulation.</p>	<p>Pour les sources indépendantes (SourceA et SourceB), les nombres aléatoires d'une distribution spécifique peuvent être générés indépendamment.</p> <p>Considérons deux sources normalement distribuées :</p> <p>SourceA donne moyenne=mSA et sd=sdSA et SourceB donne moyenne=mSB et sd=sdSB.</p> <p>Vous pouvez générer n nombres aléatoires de SourceA (SimNumSA) et SourceB (SimNumSB) indépendamment comme suit :</p> <p>SimNumSA<-rnorm(n, mSA, sdSA) SimNumSB<-rnorm(n, mSB, sdSB)</p>
---	---

3.2.2 Comment attribuer des valeurs aléatoires partagées

Il existe des cas où la même variable d'entrée est utilisée plusieurs fois dans un calcul, et dans ces cas, cette variable d'entrée ne doit avoir qu'une seule valeur aléatoire pour chaque itération. Par exemple, si un rapport commun racine:pousse est utilisé dans plusieurs types de forêts, un échantillon aléatoire d'un R:S possible est sélectionné à chaque itération, et chaque type de forêt qui requiert une valeur R:S utilise cette même valeur à cette itération.

Formule de tableur	Code R
<p>Pour les sources partagées, l'échantillon aléatoire d'une entrée d'incertitude est partagé :</p> <p>SourceA : A3=NORM.INV(RANDARRAY(1,n),\$A\$1,\$A\$2)</p> <p>SourceB : B3=\$A\$3</p> <p>Dans cet exemple, les itérations se font par rangées, et chaque colonne d'une rangée qui utilise cette source fera référence à la cellule A3 pour une valeur aléatoire de cette source.</p>	<p>Pour les sources (SourceA et SourceB) qui sont partagées, l'échantillon aléatoire pour l'entrée sera partagé par les deux sources pour chaque itération de la simulation de Monte Carlo.</p> <p>En considérant une source partagée normalement distribuée : SourceA présente moyenne=mA et sd=sdA</p> <p>Vous pouvez générer n nombres aléatoires de SourceA comme suit :</p> <p>SimNumA<- rnorm(n, mA, sdA)</p> <p>Chaque source partagée avec SourceA fera référence au numéro aléatoire tiré pour cette itération.</p>

	<pre>SimNumB <- SimNumA</pre>
--	----------------------------------

3.2.3 Comment attribuer des valeurs aléatoires partiellement corrélées

Les variables d'entrée d'un calcul peuvent n'être ni partagées ni indépendantes, mais plutôt partiellement corrélées, si ces variables sont calculées à partir d'une combinaison de sources partagées et indépendantes. Il est plus facile d'effectuer une simulation de Monte Carlo en commençant par les entrées qui sont totalement indépendantes et totalement partagées, et c'est ce que nous recommandons. Cependant, si l'on commence un calcul avec des variables partiellement corrélées comme entrées, comme des facteurs d'émission basés sur un inventaire forestier indépendant mais des ratios racines:pousses partagés, alors la simulation de Monte Carlo nécessitera des échantillons aléatoires partiellement corrélés pour ces variables.

Créer une matrice avec plus de deux variables dans Excel est possible mais difficile (Zaiontz 2020).

Formule de tableur	Code R
---------------------------	---------------

1. Pour la covariance entre deux vecteurs X (A1:A30) et Y (B1:B30), une matrice de variance-covariance 2 x 2 sera nécessaire.

Matrice 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)

2. Calculer la moyenne pour les deux variables avec :

Matrice B

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

3. Calculer une décomposition de Cholesky de la matrice de covariance (Pistilli 2019). Il s'agira également d'une matrice 2 x 2.

Matrice C

	X	Y
X	=SQRT(A ₁₁)	=0
Y	=COVARIANCE.S(A ₂₁ /C ₁₁)	=SQRT(A ₂₂ - C ₂₁ ²)

4. Calculer un vecteur normal aléatoire à deux variables. Pour n itérations de nombres aléatoires, il

Charger le package MASS, qui peut développer des nombres aléatoires normaux à plusieurs variables :

bibliothèque(MASS)

Générer d'abord une matrice de variance-covariance 2 x 2 entre les vecteurs X et Y

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

Génère n paires de valeurs aléatoires de X et Y avec la variance-covariance spécifiée dans A :

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

<p>s'agira d'une matrice 2 x n avec la même formule dans chaque cellule. Mettre en surbrillance les cellules pour former une matrice 2 x n, taper la formule :</p> <p>=MMULT(C NORM.INV(RANDARRAY(2,n))+B</p> <p>Après avoir tapé la formule, appuyer sur les touches ctrl + maj + entrée du clavier pour remplir toutes les cellules de la matrice avec la fonction.</p> <p>Chaque cellule du vecteur normal bivarié aléatoire est la moyenne de cette variable plus un résidu aléatoire avec la corrélation souhaitée.</p>	
---	--

3.3 Comment itérer

L'échantillonnage aléatoire, que ce soit à partir d'une PDF ou de données, peut être répété de nombreuses fois pour générer une distribution d'estimations à partir de laquelle l'incertitude peut être évaluée.

Formule de tableur	Code R
<p>Chaque calcul est répété en lignes (ou en colonnes).</p> <p>Excel 365 dispose d'une fonction RANDARRAY qui facilite ce processus.</p> <p>La formule RANDARRAY est utilisée en combinaison avec la formule utilisée pour échantillonner les données.</p> <p style="text-align: center;">RANDARRAY(R,C)</p> <p>où R est le nombre spécifié de lignes et C le nombre spécifié de colonnes. Par exemple, si vous effectuez des itérations en lignes, R correspond au nombre d'itérations et C à 1.</p> <p>Par exemple, si vous échantillonnez à partir d'une distribution normale avec la moyenne dans la cellule A1, le sd dans la cellule A2 et le nombre d'itérations dans la cellule A3,</p> <p style="text-align: center;">= NORM.INV(RANDARRAY(\$A\$3,1), \$A\$1, \$A\$2)</p> <p>L'avantage de RANDARRAY est que la formule n'est représentée qu'une seule fois, au lieu de l'être</p>	<p>Dans le cas d'une distribution normale avec une moyenne (mean) et un écart type (sd) spécifiés, un nombre aléatoire peut être généré comme suit :</p> <p style="text-align: center;">rnorm(1,moyenne,sd)</p> <p>R utilise des matrices et des tableaux pour stocker les données. Les calculs répétés sont gérés par le nombre d'éléments du tableau (n). Pour générer n nombres aléatoires, indiquez n :</p> <p style="text-align: center;">rnorm(n,moyenne,sd)</p>

séparément à chaque itération. Avoir des milliers de formules rend le fichier énorme et l'exécution lente. Vous devriez probablement investir dans Office 365 si vous voulez utiliser Monte Carlo dans Excel.

Le signe # peut être utilisé à la place de =NORM.INV(RANDARRAY) dans les formules simples qui n'échantillonnent pas à partir d'une distribution. Le # copie la formule pour remplir le tableau. Ceci est utile pour les calculs qui ne nécessitent pas d'échantillonnage aléatoire.

C'est en fait le même avantage que celui de RANDARRAY, la formule ne doit être tapée que dans la première cellule, et elle se remplira automatiquement à chaque itération.

Pièges potentiels

Si elles sont converties en une feuille Google, toutes les formules utilisant le # auront un message d'erreur intitulé ANCHOR ARRAY.

L'utilisation de trop peu d'itérations de Monte Carlo peut fournir des estimations d'incertitude imprécises. Pour votre calcul de réduction des émissions, vous pouvez déterminer le nombre d'itérations de Monte Carlo nécessaires pour atteindre un niveau de confiance souhaité dans vos estimations d'incertitude. Dans cet exemple (basé sur l'exemple simple de l'annexe 1), les estimations d'incertitude ne sont précises qu'à environ 20 % de la réduction d'émissions même après 2000 itérations, mais approchent les 10 % à 10 000 itérations (Figure 3.3).

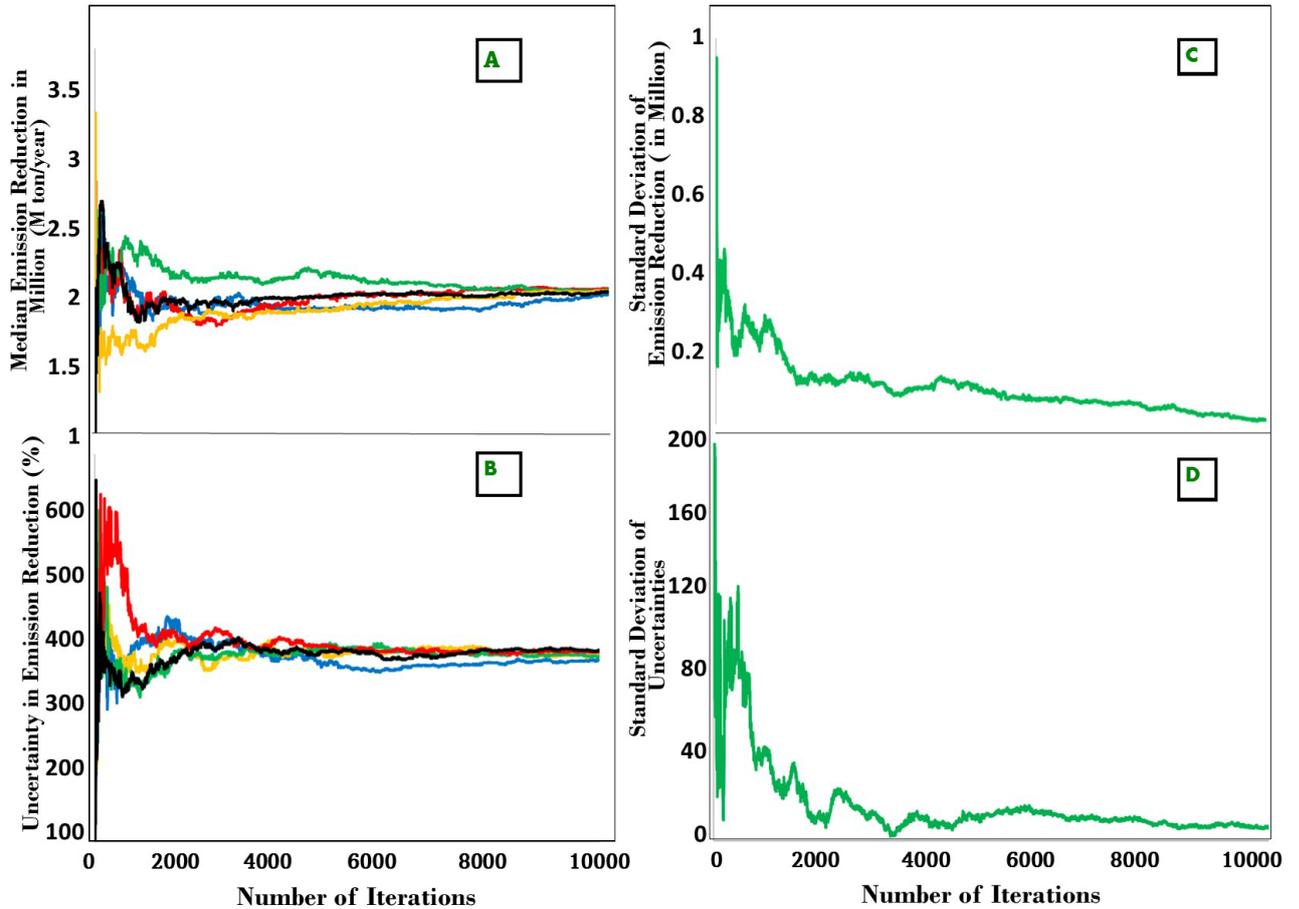


Figure 3.3. La similarité des estimations de Monte Carlo dépend du nombre d'itérations.

- Médiane des estimations de Monte Carlo des réductions d'émissions, calculées pour des nombres d'itérations allant jusqu'à 10 000 pour 5 exécutions de Monte Carlo indépendantes.
- (b) La demi-largeur du CI à 90 % des itérations de Monte Carlo divisée par la médiane des RE pour les mêmes 5 exécutions de Monte Carlo.
- L'écart-type des 5 estimations de Monte Carlo des réductions d'émissions.
- L'écart type des 5 estimations de Monte Carlo de l'incertitude.

3.4 Comment interpréter les résultats Monte Carlo

La sortie de Monte Carlo peut être analysée pour caractériser l'incertitude des résultats du calcul.

- Trouver la médiane (50ème percentile) des résultats Monte Carlo
- Trouver le 5ème percentile des sorties
- Trouver le 95ème percentile des sorties
- Calculer la demi-largeur de l'intervalle de confiance à 90 %.
- La convertir en pourcentage de la médiane.

Formule de tableur	Code R
<p>Lorsque la sortie de Monte Carlo se trouve dans les cellules B1:B10000</p> <p>A1 =MÉDIANE(B1:B10000) A2 =PERCENTILE(B1:B10000, 0.05) A3 =PERCENTILE(B1:B10000, 0.95) A4 =(A3-A2)/2 A5 =A4/A1*100</p>	<p>Considérant un vecteur de nombres simulés SimNumC (c1,c2,c3,...,cn) :</p> <p>A1<- médiane(SimNumC) A2<- quantile(SimNumC , 0.05) A3<- quantile(SimNumC , 0.95) A4<- (A3-A2)/2 A5<- abs(A4/A1)*100</p>

Pièges potentiels

Une erreur courante dans l'interprétation des résultats de Monte Carlo est de rapporter l'incertitude dans la moyenne ou la médiane de la distribution des estimations. Il s'agit d'une erreur importante, qui sous-estime généralement l'incertitude d'un facteur 100, car le calcul de l'incertitude de la tendance centrale (par exemple, l'erreur standard de la moyenne) implique de diviser l'écart standard par la racine carrée du nombre d'« observations », qui est généralement de 10 000 essais. La confiance dans la moyenne peut être rendue arbitrairement petite en augmentant le nombre d'itérations de Monte Carlo, mais l'intervalle de confiance à 90 % du nombre accru d'estimations resterait tout aussi large. L'augmentation du nombre d'itérations améliore la précision de l'estimation de l'incertitude, mais, interprétée correctement, elle ne rend pas l'incertitude plus petite.

Étape 4. Évaluer la contribution de chaque source à l'incertitude globale

Comprendre la contribution de chaque source à l'incertitude aidera à identifier les possibilités de réduire les incertitudes. L'évaluation de l'incertitude globale avec différentes entrées supposées parfaitement connues est une façon d'évaluer la sensibilité de l'incertitude globale à l'incertitude de chaque entrée.

La réalisation d'une analyse de sensibilité est facilitée par des commutateurs pour chaque entrée qui activent ou désactivent l'incertitude associée. L'incertitude globale peut être calculée avec différentes combinaisons de commutateurs activés, plutôt que de changer les formules dans le fichier Excel ou le code dans R.

Formule de tableur	Code R
<p>Cela se fait à l'aide d'une instruction « IF » faisant référence à une cellule utilisée pour activer ou désactiver une source pour l'analyse de sensibilité.</p> <p>Par exemple, si la cellule A5 de la page des résultats a la valeur « on » ou « off », et que la cellule B5 de la page des entrées donne</p> <p>IF('resultspage'A5="on", 1, 0.0000000001)</p>	<p>Le code relatif à une source d'incertitude peut être activé à l'aide d'une instruction « if ».</p>

Chaque itération d'un tirage aléatoire pour la source est multipliée par la cellule B5. Si la source d'incertitude est « off », 0.0000000001 est utilisé au lieu de 0 pour éviter les erreurs dans la fonction NORM.INV.	
--	--

La comparaison de l'importance de l'incertitude dans les différentes entrées peut être réalisée en évaluant chacune d'elles seule, toutes les autres incertitudes étant désactivées, ou en supprimant chacune d'elles, toutes les autres sources étant activées. Du point de vue de l'évaluation du bénéfice de la réduction d'une source particulière dans le contexte de toutes les autres, il est plus pertinent de rapporter combien l'incertitude est réduite en éliminant cette source que de rapporter combien cette source contribue seule, et c'est l'approche recommandée dans les Directives du FCPF sur l'application du cadre méthodologique. Cependant, il est plus facile de comprendre les résultats en considérant une seule source à la fois. Et s'il y a un besoin de réviser une source,

Le tableau suivant montre les résultats d'une analyse de sensibilité de l'exemple simple fourni en annexe.

Sources incluses	Incertitudes (Mégatonnes C/an)			Incertitudes (% des émissions)		
	Niveau de Référence	Crédité Période	RE	Niveau de Référence	Crédité Période	RE
R:S	2,80	2,50	0,29	17	17	17
FC	0,58	0,52	0,07	4	4	4
Incertitude d'échantillonnage dans EF	3,29	2,95	0,36	21	21	21
Facteurs d'émission (à partir des 3 sources ci-dessus)	4,44	3,98	0,46	27	27	27
Données d'activité	4,10	7,20	8,07	25	49	472

Toutes les sources	6,09	8,54	8,46	37	58	492
--------------------	------	------	------	----	----	-----

Dans cet exemple, les incertitudes sont élevées par rapport à la réduction des émissions. Ceci est dû au fait que dans cet exemple, la réduction des émissions était faible (1,7 mégatonnes C/an, voir Fig. 4.1 pour une explication graphique).

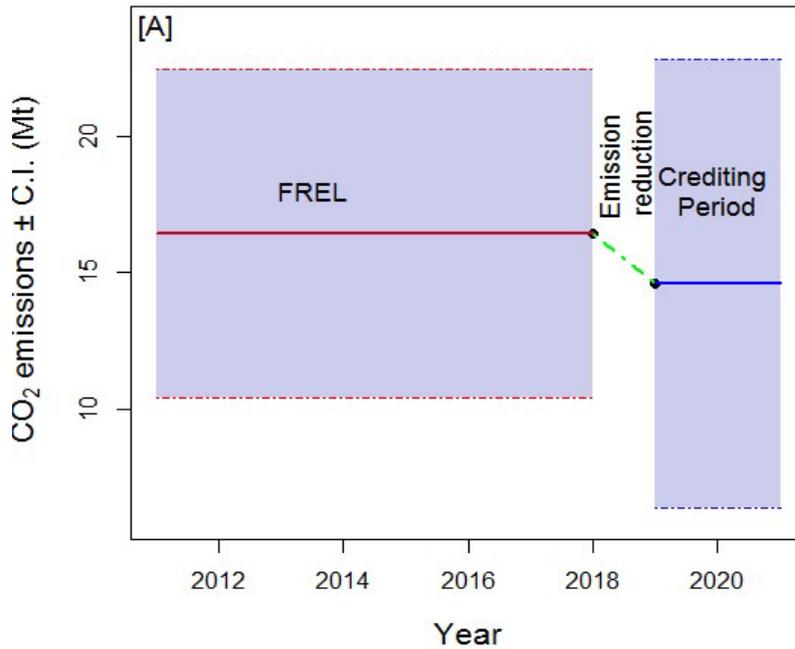


Figure 4.1. (a) Les incertitudes représentent 37 % des émissions du niveau de référence et 57 % des émissions de la période de surveillance. (b) Comme la réduction des émissions est faible, l'incertitude combinée représente une fraction importante de la réduction des émissions (>400 %).

Comme le montre la figure 4.2, les incertitudes relatives à la réduction des émissions deviendront plus petites avec le temps si la réduction des émissions augmente avec le temps, en supposant que les incertitudes contributives sont relativement constantes dans le temps (Neeff 2021).

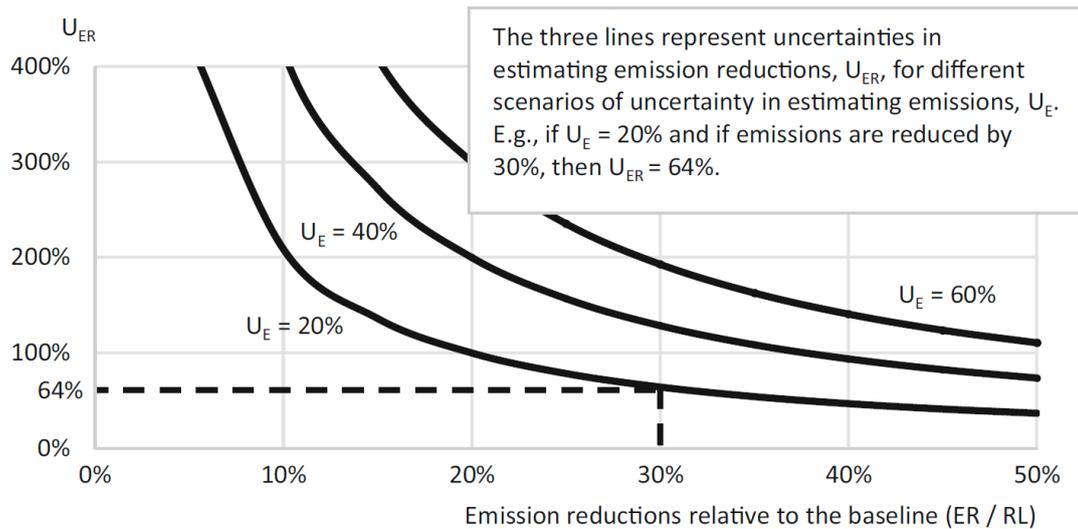


Figure 4.2. Incertitude dans l'estimation des réductions d'émissions par des scénarios d'efficacité dans la réduction des émissions sous le niveau de référence et par l'incertitude dans la mesure des émissions (de Neff (2021), modifié de FAO 2019).

Pièges potentiels

Si les estimations de l'incertitude ne sont pas très précises (basées sur un petit nombre d'itérations de Monte Carlo), par variation aléatoire, l'incertitude avec une source désactivée peut être légèrement supérieure à celle avec la source activée. Le fait de déclarer chaque source activée, plutôt que chaque source désactivée, permet d'éviter ce problème. L'augmentation du nombre d'itérations de Monte Carlo rend les estimations de l'incertitude plus précises (Figure 3.3).

Documents cités

- Efron, B., Tibshirani, R. J. 1994. An introduction to the bootstrap. *CRC press*.
- FCPF 2016. Carbon fund methodological framework. *Rapport du Fonds de Partenariat pour le Carbone Forestier*.
- 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. Fonds de Partenariat pour le Carbone Forestier. *Rapport du Fonds de Partenariat pour le Carbone Forestier*. Disponible sur : https://www.forestcarbonpartnership.org/sites/fcp/files/FCPF%20Guidelines%20on%20Uncertainty%20Analysis_2020_0.pdf
- GIEC 2006. IPCC guidelines for national greenhouse gas inventories. vol. 4 agriculture, forestry and other land use. *Préparées par le National Greenhouse Gas Inventories Programme*. S Eggleston, L Buendia, K Miwa, T Ngaraand K Tanabe. Institute for Global Environmental Strategies, Hayama, Japon. Disponible sur : <https://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html>
- GIEC 2019. Refinement to the 2006 IPCC guidelines for national greenhouse gas inventories. *Préparé par le 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, Japon. Disponible sur : <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 mai 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/.

Annexe 1. Exemple de Monte Carlo dans Excel

L'exemple simple dans Excel a deux versions, l'une avec une période de crédit de 2 ans, et l'autre avec une période de crédit de 4 ans, ce qui illustre la forte incertitude associée à une courte période de crédit.

Deux types de forêts sont inclus dans l'exemple simple, et la déforestation est la seule transition d'utilisation des terres. L'exemple illustre un taux de déforestation de 3 % par an dans les deux types de forêt.

Introduction

La première feuille du classeur Excel décrit les rôles des feuilles suivantes.

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: $SE = \text{mean} * \text{uncertainty} (90 \% CI) / 100 / 1.96$

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.

Variables d'entrée

La feuille Variables d'entrée contient toutes les données nécessaires au calcul des réductions d'émissions. Le Tableau 1 est constitué des données d'activité, soit la surface de terres convertie du type de forêt 1 en non forêt (FT1-NF) et la surface de terres convertie du type de forêt 2 en non forêt (FT2-NF) pour la période totale de 10 ans : sept ans de période de référence et trois ans de période de surveillance. Le tableau 2 présente les facteurs d'émission, notamment la fraction de carbone (CF), le ratio racines:pousses (R:S) et la biomasse aérienne (AGB).

Les incertitudes sont données comme la demi-largeur CI à 90 %. Le SE est calculé à partir de l'incertitude et de la valeur de l'estimation, bien que nous reconnaissons qu'en réalité, l'incertitude est calculée à partir du 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	Conversion	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													

Données d'activité (AD)

Dans la feuille de données d'activité, les données de transition forestière pour chaque période de temps sont simulées à l'aide de la simulation de Monte Carlo comme : =NORM.INV(RAND(), B\$5, B\$7*B\$1)

NORM.INV(probabilité, moyenne, sd) donne un inverse de la distribution cumulative normale, à la probabilité, la moyenne et l'écart type (sd) spécifiés.

RAND() génère des valeurs normales aléatoires uniformes entre 0 et 1.

B\$5 est la moyenne de la valeur.

B\$7 est l'erreur standard.

B\$1 est un multiplicateur contrôlé par un commutateur dans la feuille ER et analyse de sensibilité.

		A	B	C	D	E	F	G	H
B9		=NORM.INV(RAND(),B\$5, B\$7*B\$1)							
1	Multiplier	1	1	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									

Facteurs d'émissions (EF)

Dans la feuille des facteurs d'émission, R:S, CF et AGB pour chaque type de forêt sont simulés à l'aide de la simulation de Monte Carlo.

La biomasse souterraine (BGB) est calculée en multipliant chaque valeur AGB simulée par la valeur R:S simulée.

Le carbone total pour chaque type de couverture terrestre est calculé comme la somme de l'AGB et de la BGB.

Les facteurs d'émissions (EF) sont calculés comme la différence de carbone total entre les types de couverture terrestre. La colonne M représente la transition de la forêt de type 1 (FT1) à la non-forêt (NF). Le processus est répété pour FT2-NF dans la colonne N.

Les cellules sous la ligne 10 sont des valeurs simulées pour chaque variable d'entrée. Les multiplicateurs de la ligne 8 sont contrôlés par des commutateurs sur la feuille ER et analyse de sensibilité.

B11 : X ✓ fx =NORM.INV(RAND(),\$B\$3,\$D\$3*\$B\$8)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Value	U (%)	SE										
2	CF	0.47	4	0.01										
3	R:S	0.321	67	0.11										
4	AGB_FT1 (ton/ha)	263	20	26.84										
5	AGB_FT2 (ton/ha)	150	38	29.08										
6	AGB_NF (ton/ha)	26	29	3.85										
7		R:S Uncertainty	CF Uncertainty	Sampling Uncertainty			Sampling Uncertainty			Sampling Uncertainty				
8	Multiplier	1	1	1			1			1				
9		Parameters		Simulated Carbon Densities: Forest			Simulated Carbon Densities:			Simulated Carbon Densities:			FE: FT1-NF	FE: FT2-NF
10	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)
11	1	0.32	0.48	256	82	162	92	30	58	18	6	11	151	47
12	2	0.41	0.46	261	106	170	138	56	90	28	11	18	151	72
13	3	0.18	0.46	167	30	90	128	23	69	27	5	15	75	54
14	4	0.32	0.48	295	94	186	198	63	125	28	9	18	168	107
15	5	0.14	0.48	290	42	159	155	22	85	20	3	11	148	74

ER et simulation

Les émissions pour chaque transition forestière sont calculées en multipliant les données d'activité (AD) de la feuille AD par les facteurs d'émission (EF) de la feuille EF.

B4 : X ✓ fx =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
10008	Percentile 95%	18660940.4	2709561.9
10009	Half of confidence interval	4758640.825	870776.275

1. Dans la rangée 10006, la médiane de toutes les simulations pour chaque type de forêt est calculée comme suit

$$=MÉDIANE(B4:B10003)$$

- Dans la ligne 10007, le percentile de 5 % de toutes les simulations pour chaque type de forêt est calculé comme suit

=PERCENTILE(B4:B10003, 0.05)

- Dans la ligne 10008, le percentile de 95 % de toutes les simulations pour chaque type de forêt est calculé comme suit

=PERCENTILE(B4:B10003, 0.95)

- Dans la ligne 10009, la demi-largeur de l'intervalle de confiance à 90 % est calculée comme suit : (percentile 95-percentile 5)/2

c'est-à-dire (B10008-B10007)/2

- Enfin, dans la ligne 10010, l'incertitude est calculée comme suit : (moitié de l'intervalle de confiance à 90 %/médiane)*100

c'est-à-dire (B10009/B10006)*100

B10006							
=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							

Commutateurs

Dans l'image ci-dessous, l'incertitude finale des réductions d'émissions (cellule X10009) est surlignée en bleu. Lorsque les commutateurs des cellules D10012:D10017 sont activés, les multiplicateurs correspondants sont fixés à 1. Lorsque les commutateurs sont « off », les multiplicateurs sont fixés à 0 (ou 1E-10, ce qui est proche de 0 et évite les erreurs dans la formule Excel). Les incertitudes changent légèrement chaque fois que le classeur est mis à jour et que les valeurs aléatoires sont rééchantillonnées.

	A	B	C	D	E	F	G	H	I	J
10009	Uncertainty (%)	38	58	48	67	53	74	31	47	414.55
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						
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				

Cette image montre ce qui se passe lorsque l'incertitude R:S est désactivée dans l'onglet EF.

B11: `=ROUND(NORM.INV(RAND(),B3,C3*B8),2)`

	A	B	C	D	E	F	G
6	AGB_NF (t/ha)	26	3.8	29		BGB_NF (ton/ha)	8.346
7		R:S Uncertainty	CF Uncertainty	Sampling Uncertainty			Sampling
8	Switch	1E-10	1	1			1
9		Parameters			Simulated Carbon Densities: Forest Type 1		Simulat
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	105
10009	9999	0.32	0.48	257	82	163	202
10010	10000	0.32	0.46	201	64	122	178
10011							

La contribution de chaque source à l'incertitude globale peut être déterminée en commençant avec toutes les sources activées et en désactivant chaque source une par une, ou en commençant avec toutes les sources désactivées et en les activant une par une. Le tableau d'analyse de sensibilité est rempli avec les valeurs copiées de la cellule J1009 avec différentes combinaisons de commutateurs activés ou désactivés.

	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

Annexe 2. Exemple de Monte Carlo en R

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

L'exemple simple dans R est le même que celui dans Excel, montrant les taux de déforestation de deux types de forêts.

1. Chargement de bibliothèque

Pour exécuter le code R actuel, chargez les bibliothèques suivantes :

```
bibliothèque(matrixStats)
bibliothèque(gridExtra)
bibliothèque(reshape)
```

2. Lecture des entrées par type de forêt et période

Lire les valeurs spécifiques de AD et EF (et les incertitudes associées) par type de forêt (FT) et période (P). Lire les tableaux csv de AD et EF comme suit :

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

```
### Adresse pour lire les entrées
setwd("C:/Desktop/Inputs")
```

```
### Lecture des entrées
```

```
# Entrées AD
```

```
BaseAD <- read.csv("1_Activity_Data.csv")
dim(BaseAD)
```

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

```
# Entrées EF
```

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

3. Simulations de nombres aléatoires de AD et EF

Une étape clé du MCS est la génération de nombres aléatoires de AD et EF. Pour générer des nombres aléatoires, il est nécessaire : (i) l'estimateur de AD et EF et, (ii) l'erreur standard (SE)

des estimateurs de AD et EF. Néanmoins, dans la plupart des cas, seules les incertitudes de AD et EF sont rapportées. Il est donc nécessaire de calculer la SE de AD et EF comme suit :

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

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

U = incertitude

IC= intervalle de confiance

θ = estimateur de AD ou EF

σ = écart-type

Une fois que les SE de AD et EF ont été calculés, il est possible (i) de générer des vecteurs de nombres aléatoires de AD et EF par FT-P, (ii) d'estimer les émissions par FT, et (iii) d'enregistrer les vecteurs d'émissions simulées par FT sous forme de matrice.

Nous montrons ci-dessous comment simuler des nombres aléatoires (de distribution normale, en utilisant la moyenne et l'erreur standard) de AD indépendants et de EF partiellement corrélés :

1. Simulation des données d'activité

Calcul de SD d'AD

```
BaseAD$DesEstDA<-abs((BaseAD$U_AD_per*BaseAD$AD_ha)/(1.65*100))
```

Nombre de simulations

```
n<-1000000
```

```
MatrizDef<-seq(1:n)
```

Commutateur pour AD si c'est 1 incorpore MC et si c'est 0 alors utilise la valeur simple

```
SWITCH_AD =1
```

Simulations de AD par période et conversions

```
pour (i dans 1:length(BaseAD$Code))
```

```
{
```

```
  DAsim<-rnorm(n,mean=BaseAD$AD_ha[i], sd=BaseAD$DesEstDA[i]*SWITCH_AD)
```

```
  MatrizDef<-cbind(MatrizDef,DAsim)
```

```
}
```

Conversion de matrice AD en DataFrame

```

Matrix_AD<-as.data.frame(MatrizDef)

### Colnames corrects de AD-Dataframe par « Période » et « 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 des facteurs d'émissions #####

### Simulations et CF, R:S et AGB par type de forêt et transition
### Matrice pour enregistrement de CF, R:S et AGB simulés
MatrizEF_Def <-seq(1:n)
### Commutateurs pour EF
SWITCH_EF =1

### Simulations de CF, R:S et AGB
pour (i dans 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)
}

### Conversion de la matrice "CF, R:S and AGB" en DataFrame
Matrix_EF<-as.data.frame(MatrizEF_Def)

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

#####

### Simulation de BGB, densités carbone et EF par Transition

### Simulation de BGB par strate
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 des densités carbone par strate
Matrix_EF$C_FT1 <- (Matrix_EF$AGB_FT1 + Matrix_EF$BGB_FT1 ) * Matrix_EF$CF

```

```
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 d'EF par 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
```

```
### Filtrate de EF simulé par 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 des émissions simulées

En utilisant les AD et EF simulés, les émissions par transition d'utilisation des terres et par période sont estimées :

```
### Fusion de "AD-Dataframe" et "EF-Dataframe"
```

```
Table_Emi<- merge(Matrix_AD, Matrix_EF1, by.x = "Id_sim_AD", by.y = "Id_sim_EF",all=T)
```

```
### Estimation des émissions par période et transition annualisées
```

```
anP1=7
```

```
anP2=3
```

```
anP3=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 des émissions de la ligne de base (2 périodes, 7 et 3 ans) #####
```

```
### Sélection des émissions pour la période de la ligne de base.
```

```
Table_Emi_FREL<- Table_Emi[, c(1,10:13)]
```

```
Table_Emi_FREL$FREL<- ( rowSums(Table_Emi_FREL[,c(2:5)]) ) / (yearP1+yearP2)
```

5. Estimation des quantiles et des incertitudes

En utilisant les vecteurs d'émissions simulées, il est possible d'estimer les incertitudes inférieures/supérieures associées comme suit :

$$U_{\text{inf}} = \frac{|Q(0.025) - \hat{E}|}{\hat{E}} \times 100 \quad \text{and} \quad U_{\text{sup}} = \frac{|Q(0.975) - \hat{E}|}{\hat{E}} \times 100$$

U_{inf} = Incertitude sur la partie gauche des émissions simulées

U_{sup} = Incertitude sur le côté droit des émissions simulées

$Q(0.025)$ = quantile 0.025 des émissions simulées

$Q(0.975)$ = quantile 0.975 des émissions simulées

E = émissions simulées par la méthode de Monte Carlo

(i) Ainsi, les quantiles inférieur/supérieur des émissions de la ligne de base peuvent être estimés comme suit :

```
Q_05_FREL <-quantile(Table_Emi_FREL$FREL,0.05)[[1]]
```

```
Q_95_FREL <-quantile(Table_Emi_FREL$FREL,0.95)[[1]]
```

(ii) de même, les incertitudes inférieures/supérieures des émissions de la ligne de base peuvent être estimées comme suit

```
half_CI <- (Q_95_FREL - Q_05_FREL)/2
```

```
U_FREL<-abs( half_CI / median(Table_Emi_FREL$FREL))*100
```

6. Sauvegarde des quantiles et des incertitudes

Les quantiles et les incertitudes des émissions par période et l'émission médiane sont sauvegardés:

```
### Sauvegarde des émissions de la ligne de base et des quantiles et incertitudes associés
```

```
Table_FREL<-data.frame(Period = "FREL",
```

```
Émission = 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. Fonction de densité de probabilité des émissions par période et ligne de base

Enfin, les fonctions de densité de probabilité (PDF) des émissions par période et pour les émissions médianes sont tracées :

```
### Enregistrement de la PDF des émissions de la ligne de base
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()
```

Une version préliminaire du code R présenté ci-dessus est disponible au lien suivant :
https://github.com/mark78b/MCS_in_R/

Informations sur le document

Version	Date	Description
1,0	21 septembre 2021	Version initiale