



WINROCK
INTERNATIONAL

GUÍA PARA APLICAR EL ENFOQUE MONTE CARLO AL ANÁLISIS DE INCERTIDUMBRE EN LA CONTABILIDAD FORESTAL Y DE GASES DE EFECTO INVERNADERO

Anna McMurray, Timothy Pearson y Felipe Casarim
2017

Contenido

1. Introducción	4
2. Enfoque de simulación de Monte Carlo	5
3. Pasos para llevar a cabo el análisis de incertidumbre utilizando Monte Carlo	7
3.1 Ajuste de las distribuciones	7
<i>Identificación de FDPs cuando todo el conjunto de datos está disponible</i>	<i>7</i>
<i>Identificación de FDPs cuando la distribución subyacente no está disponible.....</i>	<i>10</i>
3.2 Ejecución de simulaciones de Monte Carlo.....	11
<i>Selección de software.....</i>	<i>11</i>
<i>Número de simulaciones.....</i>	<i>13</i>
<i>Truncamiento de la distribución ajustada</i>	<i>13</i>
3.3 Combinación de simulaciones de Monte Carlo	14
3.4 Cálculo de intervalos de confianza.....	16
<i>Intervalos de confianza para distribuciones normales.....</i>	<i>16</i>
<i>Lo más probable es que el número de simulaciones sea muy alto (por ejemplo, 10 000) y, como resultado, el intervalo de confianza será pequeño.</i>	<i>16</i>
3.5 Cálculo del porcentaje de incertidumbre.....	18
4. Aplicación completa del enfoque Monte Carlo.....	19
5. Discusión sobre la aplicación de Monte Carlo a los análisis de incertidumbre	19
Anexo.....	21
Anexo 1: Ejemplo simplificado de aplicación del enfoque Monte Carlo	21
A. <i>Ajustar las distribuciones y ejecutar las simulaciones de Monte Carlo</i>	<i>21</i>
B. <i>Aplicación de simulaciones de Monte Carlo a ecuaciones para calcular la incertidumbre de las emisiones totales</i>	<i>24</i>
C. <i>Cálculo del intervalo de confianza.....</i>	<i>26</i>
D. <i>Cálculo de incertidumbre.....</i>	<i>27</i>

Figuras

Figura 1. Ejemplos de modelos de funciones de densidad de probabilidad que se utilizan normalmente (tomados de la Figura 3.5 del IPCC))	6
Figura 2. Ilustración del enfoque de Monte Carlo	6
Figura 3. Pasos para llevar a cabo el enfoque de Monte Carlo para calcular la incertidumbre	7
Figura 4. Ilustración de datos atípicos	8
Figura 5. Ejemplo de ajuste de FDP a un conjunto de datos	9
Figura 6. Simulación con distribución normal con y sin truncamiento.	14
Figura 7. Simulation using lognormal distribution with and without truncation.	14
Figura 8. Ejemplo de la aplicación de simulaciones de Monte Carlo al modelo que estima las emisiones totales	15
Figura 9. Los valores medios de una población se remuestran 1,000 veces, usando la técnica de bootstrapping.....	17
Figura 10. El intervalo de confianza calculado mediante el método de remuestreo (bootstrapping) es la diferencia entre el percentil 2.5 y el percentil 97.5 de la distribución inicial de las medianas	18
Figura 11. Ejemplos de uso de cuantiles de distribución de emisión final para calcular la incertidumbre.....	20

Este proyecto es parte de la Iniciativa climática internacional (IKI). El Ministerio federal alemán para el medio ambiente, la conservación de la naturaleza, la construcción y la seguridad nuclear (BMUB) apoya esta iniciativa sobre la base de una decisión adoptada por el Bundestag alemán.

Para comentarios o preguntas, por favor, póngase en contacto con el autor principal:
Anna McMurray: anna.mcmurray@winrock.org

1. Introducción

Al calcular las emisiones de gases de efecto invernadero, siempre es necesario evaluar y cuantificar las incertidumbres de las estimaciones. Los análisis de incertidumbre ayudan a los analistas y a los responsables de tomar decisiones a identificar cuán precisas son las estimaciones y el rango probable en el que cae el verdadero valor de las emisiones.

Hay tres pasos generales para realizar cualquier análisis de incertidumbre: 1) identificar las fuentes de incertidumbre en la estimación; 2) cuantificar las diferentes fuentes de incertidumbre, siempre que sea posible; y 3) combinar / agregar las diferentes incertidumbres para llegar a un valor de incertidumbre final. El capítulo 3 de las Directrices del IPCC de 2006 para los inventarios nacionales de gases de efecto invernadero Volumen 1¹ (en adelante denominada IPCC, por sus siglas en inglés) brinda información sobre métodos de análisis de incertidumbre.

Para realizar el tercer paso de los análisis de incertidumbre, la combinación de incertidumbres, el IPCC presenta dos enfoques:

- 1) Propagación de error, y
- 2) Simulación de Monte Carlo.

La propagación de error implica la combinación de estimaciones de incertidumbre en ecuaciones simples. Se considera un enfoque de nivel 1 y lo puede aplicar casi cualquier persona con experiencia en el uso de ecuaciones en hojas de cálculo.

El enfoque de simulación de Monte Carlo es significativamente más complejo ya que implica la generación repetida de valores aleatorios, basados en las distribuciones de los datos de entrada. Debido al mayor nivel de complejidad, los analistas sin antecedentes estadísticos significativos necesitarán una guía detallada sobre cómo llevar a cabo las simulaciones de Monte Carlo.

No obstante, para cualquier cosa más allá de los análisis de incertidumbre más básicos, se prefieren en gran medida las simulaciones de Monte Carlo. Un enfoque de propagación de errores no es apropiado en las siguientes circunstancias, tal como se señala en el IPCC:

- Las incertidumbres son grandes²;
- Las distribuciones no son normales;
- Las ecuaciones son complejas;
- Los datos están correlacionados;
- Diferentes incertidumbres en diferentes años de inventario

¹ Frey, C., Penman, J., Hanle, L., Monni, S., Ogle, S. (2006). Capítulo 3. Incertidumbres. En el Volumen 1, Orientación general y presentación de informes, Directrices del IPCC de 2006 para los inventarios nacionales de gases de efecto invernadero, Programa nacional de inventarios de gases de efecto invernadero (pp. 66). Kanagawa, Japón. Panel intergubernamental sobre cambio climático, Unidad de apoyo técnico. <https://www.ipcc-nggip.iges.or.jp/public/2006gl/vol1.html>.

² Según el IPCC, se considera que hay una gran incertidumbre cuando la desviación estándar dividida por la media es mayor a 0.3.

Por lo tanto, es importante ser consciente de que en la mayoría de los contextos de contabilidad forestal y de gases de efecto invernadero habrá una gran incertidumbre en los datos de entrada, muchas veces las distribuciones no serán normales, las ecuaciones pueden ser complejas, existen muchas correlaciones de conjuntos de datos y la variación anual es significativa en cualquier sistema natural. Por lo tanto, Monte Carlo es el enfoque correcto y el uso de los análisis de incertidumbre de Monte Carlo debe ser más frecuente.

El IPCC brinda información general sobre el enfoque de la simulación de Monte Carlo, pero información limitada sobre cómo implementarlo. Cualquier aplicación para ejecutar las simulaciones Monte Carlo y la aplicación de los resultados para estimar la incertidumbre plantea una serie de preguntas y problemas que no se abordan en las directrices del IPCC.

Esta guía tiene como objetivo ayudar a llenar la falta de información que existe actualmente y servir como una guía técnica para los analistas que desean aplicar el enfoque de Monte Carlo a los análisis de incertidumbre. Se supone que las personas que usan esta guía cuentan con lo siguiente:

- Una comprensión de las estadísticas descriptivas y algo de experiencia en la aplicación de estadísticas básicas (por ejemplo, la capacidad de llevar a cabo análisis de incertidumbre utilizando la propagación de errores), pero poca experiencia aplicando el enfoque de Monte Carlo en particular.
- Competencia básica en el uso de Excel (es decir, están familiarizadas con las funciones básicas de Excel y pueden crear ecuaciones simples). Se proporcionan diferentes recomendaciones de software basadas en Excel en base a la propia experiencia de los autores con las mismas. No obstante, el lector debe investigar las mejores opciones disponibles para él / ella, incluidos otros programas basados en Excel. Si el lector tiene competencia en otro software estadístico, como R, SAS y SPSS, él / ella debe considerar asimismo estas opciones.

La orientación se centrará en la aplicación de los análisis de incertidumbres de REDD +, pero se podría aplicar potencialmente a los análisis de incertidumbre de las emisiones de GEI de otros sectores o, de manera más general, a cualquier tipo de análisis de incertidumbre. Un ejemplo simple de implementación del enfoque de Monte Carlo para combinar las incertidumbres aparece en el Anexo 1.

2. Enfoque de simulación de Monte Carlo

El enfoque de Monte Carlo implica la simulación repetida de muestras dentro de las funciones de densidad de probabilidad de los datos de entrada (por ejemplo, los factores de emisión o eliminación y los datos de actividad).

***Las funciones de densidad de probabilidad (FDP)** explican el rango de valores potenciales de una variable dada y la probabilidad de que valores diferentes representen el valor verdadero.*

Las FDP se representan gráficamente como distribuciones. Los ejemplos comunes incluyen distribuciones normales (Gaussianas), lognormales, triangulares y uniformes, como se muestra en la Figura 1 (tomada del IPCC).

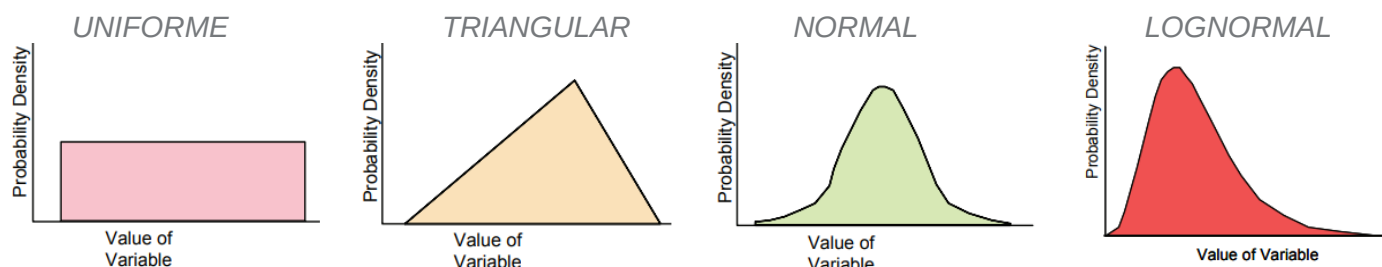


Figura 1. Ejemplos de modelos de funciones de densidad de probabilidad que se utilizan normalmente (tomados de la Figura 3.5 del IPCC)

Las simulaciones de Monte Carlo se ejecutan utilizando algoritmos que generan valores estocásticos (es decir, aleatorios) basados en la FDP de los datos. El objetivo de estas simulaciones repetidas es producir distribuciones que representen la probabilidad de diferentes estimaciones. Una vez que se han ejecutado las simulaciones, se aplican al modelo, que podría ser complejo o ser una ecuación simple, desarrollada para calcular la estimación final. Para calcular la incertidumbre, el intervalo de confianza se puede identificar para las distribuciones finales, tal como se muestra en la Figura 2.

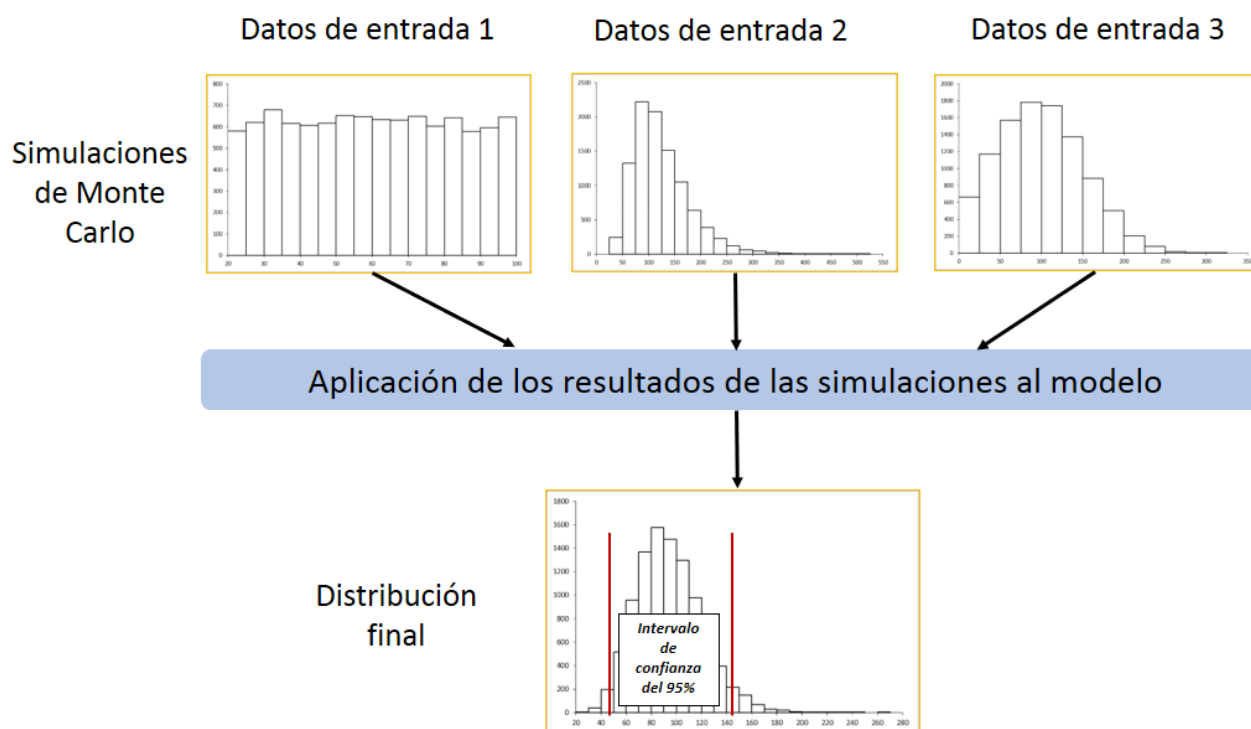


Figura 2. Ilustración del enfoque de Monte Carlo

3. Pasos para llevar a cabo el análisis de incertidumbre utilizando Monte Carlo

Una vez que se han identificado y se han cuantificado las diferentes fuentes de incertidumbre cuando sea posible, el enfoque de Monte Carlo se puede implementar a través de 5 pasos principales, tal como se muestra en la Figura 3 y se analiza en detalle en las siguientes secciones.

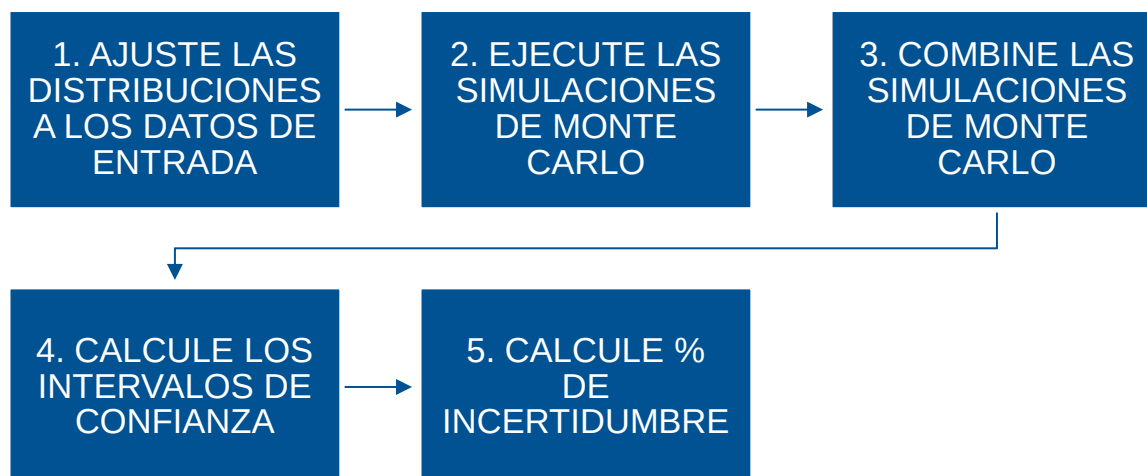


Figura 3. Pasos para llevar a cabo el enfoque de Monte Carlo para calcular la incertidumbre

3.1 Ajuste de las distribuciones

Antes de ejecutar las simulaciones de Monte Carlo, es necesario identificar las funciones de densidad de probabilidad (FDPs) que se ajustan a cada una de las fuentes de datos con fuentes de incertidumbre clave identificadas.

Identificación de FDPs cuando todo el conjunto de datos está disponible

Idealmente, todo el conjunto de datos está disponible para identificar su distribución y la base de datos se deriva de una muestra aleatoria que es representativa de la población subyacente. Cuando todo el conjunto de datos está disponible, los analistas deben ajustar los datos para corregir cualquier sesgo conocido en los datos o valores atípicos.

Tal como se define en el IPCC, un sesgo, también denominado error sistemático, es la falta de precisión, es decir, la falta de acuerdo entre el valor real y el promedio de las observaciones medidas repetidas o las estimaciones de la variable. La identificación y estimación de sesgos requerirá con frecuencia una buena comprensión del sistema que se analiza. Por ejemplo, si las mediciones de campo de biomasa solo se pueden realizar en bosques particularmente densos en comparación con el bosque promedio dentro de una jurisdicción o país determinados, el analista debe ajustar las estimaciones de los factores de emisión para la deforestación descendente para corregir este sesgo en base

a la opinión de peritos. Para los datos de actividad, se pueden realizar evaluaciones de precisión como el enfoque presentado por Olofsson y otros (2013)³ para identificar y corregir los principales sesgos.

Los valores atípicos son puntos de datos que se encuentran a una distancia anormal del resto del conjunto de datos (Figura 4). Estos valores pueden tener un impacto sustancial en la forma total de la distribución de datos y, por lo tanto, la distribución de probabilidad resultante. La decisión del analista de eliminar o no los datos atípicos depende de su conocimiento de los datos subyacentes. Los valores atípicos pueden ser componentes importantes del conjunto de datos y, por lo tanto, no se deben eliminar. No obstante, pueden representar asimismo errores de medición o de registro, en cuyo caso, se deben eliminar.

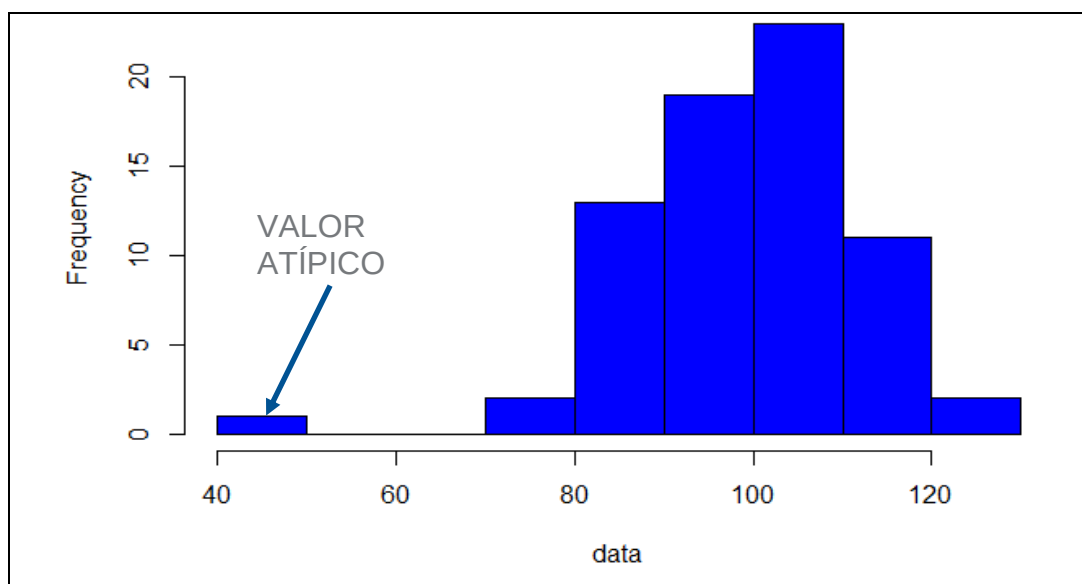


Figura 4. Ilustración de datos atípicos

Cuando este es el caso se puede aplicar una variedad de pruebas de bondad de ajuste para identificar la FDP que mejor se ajusta a los datos. La Figura 5 proporciona un ejemplo de cómo ajustar una FDP a los datos. En este caso, las barras azules ilustran la distribución de los datos reales y la línea violeta representa la FDP (Logística generalizada) que se identificó como un buen ajuste.

³ Olofsson, P., Foody, G. M., Stehman, S. V., & Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129, 122-131.

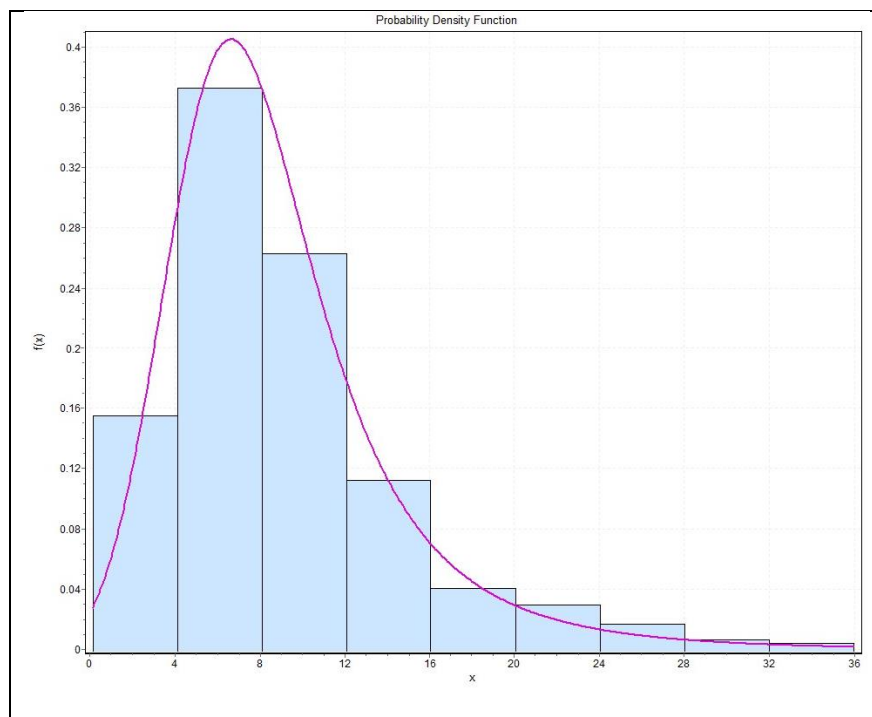


Figura 5. Ejemplo de ajuste de FDP a un conjunto de datos

Existen diferentes programas de software estadísticos que ejecutan múltiples pruebas de bondad de ajuste para diferentes FDPs en un conjunto de datos determinado. Estos programas le permiten al analista observar los resultados de las diferentes pruebas ejecutadas, ahorrando así un tiempo considerable en lugar de seleccionar manualmente las distribuciones para cada potencial FDP. Los ejemplos de este software⁴ incluyen:

- EasyFit (<http://www.mathwave.com/easyfit-distribution-fitting.html>): ejecuta 3 pruebas de bondad de ajuste (Anderson-Darling, Chi-cuadrado y Kolmogorov-Smirnov) simultáneamente en más de 55 distribuciones de probabilidad;
- XLStat (<https://www.xlstat.com/en/solutions/features/distribution-fitting>): le permite al usuario seleccionar uno de los 18 archivos FDP y ejecutar pruebas de bondad de ajuste de Chi-cuadrado y Kolmogorov.

Puede haber varias FDPs que se ajusten bien a los datos de acuerdo con los resultados de las pruebas estadísticas. Es muy probable que la selección de la FDP final coincida con la distribución que mejor se ajusta en el análisis estadístico a la lista de FDPs disponibles en el software que se utilizará para ejecutar las simulaciones de Monte Carlo.

Es importante que el analista tenga un conjunto de criterios establecidos que esté utilizando para seleccionar las FDPs. Al establecer estos criterios, se debe considerar lo siguiente:

- Las diferentes pruebas de bondad de ajuste brindan resultados, como valores p, que indicarán si una FDP determinada es estadísticamente diferente de la distribución del conjunto de datos. La FDP seleccionada no debe ser estadísticamente diferente del conjunto de datos. Como con cualquier prueba que involucre

⁴ Este software está patentado. Los costos variarán según el tipo de usuario y la cantidad de licencias necesarias.

significación estadística (es decir, pruebas de hipótesis), el nivel de significancia (por ejemplo, $P = 0.05$, $p = 0.1$, y demás) es subjetivo y depende de la opinión del analista. Idealmente, la FDP final seleccionada debería ser la que se considera que tiene el mejor ajuste al conjunto de datos de acuerdo con los resultados de las pruebas estadísticas.

- Dado que el software de Monte Carlo brinda opciones limitadas de FDP, la FDP final seleccionado debe ser una disponible en el software elegido para ejecutar Monte Carlo, que está clasificada como la que mejor se ajusta a las pruebas estadísticas y no es estadísticamente significativo a partir de los datos que utilizan el umbral de nivel de significación establecido por el analista.
- Si una distribución normal (gaussiana) está clasificada entre las distribuciones estadísticamente significativas que mejor se ajustan, entonces se debería preferir la distribución mencionada.

Las pruebas de bondad de ajuste⁵ deben brindar parámetros (por ejemplo, media y desviación estándar) de la FDP seleccionada, que el analista puede luego utilizar para ejecutar simulaciones de Monte Carlo. No obstante, antes de ejecutar las simulaciones, el analista debe asegurarse de que los parámetros sean los mismos que los requeridos para ejecutar las simulaciones. De lo contrario, se deben convertir a los parámetros necesarios para las simulaciones⁶.

Identificación de FDPs cuando la distribución subyacente no está disponible

Cuando todo el conjunto de datos no está disponible, el analista debe confiar en la comprensión de la fuente de los datos subyacentes, así como de cualquier métrica disponible (por ejemplo, desviación estándar, índice, desviación cuadrática media y demás) asociada con el valor estimado. En la Sección 3.2.2.4, el IPCC proporciona condiciones para el uso de varias FDPs comunes, incluidos normal, lognormal, uniforme, triangular y fractal (Cuadro 1).

Cuadro 1.

Ejemplos de FDPs comunes y las situaciones que representan (tomados de la Sección 3.2.2.4 del IPCC)

- La distribución normal es más apropiada cuando el rango de incertidumbre es pequeño y simétrico con relación a la media. La distribución normal surge en situaciones en las cuales muchos aportes individuales contribuyen a una incertidumbre global y en las cuales ninguna de las incertidumbres individuales domina la incertidumbre total.
- De manera similar, si un inventario es la suma de las incertidumbres de muchas categorías individuales, no obstante, ninguna de las cuales domina la incertidumbre total, entonces es probable que la incertidumbre general sea normal. Una suposición de normalidad es en general apropiada para muchas categorías para las cuales el rango relativo de incertidumbre es pequeño, por ejemplo,

⁵ Las pruebas de bondad de ajuste incluyen Shapiro-Wilks (solo para evaluar si las distribuciones son normales o no normales), Chi-squared, Kolmogorov-Smirnov y Anderson-Darling. Todas estas pruebas requieren la identificación de una FDP particular (por ejemplo, normal, lognormal, uniforme, etc.) para probar si la FDP se ajusta a los datos.

⁶ Por ejemplo, en EasyFit los parámetros proporcionados para las distribuciones son μ y σ . Para cierto software Monte Carlo, estos se deben convertir a la media lognormal y la desviación estándar. Las ecuaciones para hacer esto se pueden encontrar en <https://www.mathworks.com/help/stats/lognstat.html>.

factores de emisión de combustibles fósiles y datos de actividad.

- La distribución lognormal puede ser apropiada cuando las incertidumbres son grandes para una variable no negativa y se sabe que están sesgadas de manera positiva. El factor de emisión para el óxido nitroso de los fertilizantes aplicados al suelo proporciona un ejemplo de inventario típico. Si se multiplican muchas variables inciertas, el producto se aproxima asintóticamente a la lognormalidad. Debido a que las concentraciones son el resultado de procesos de mezcla, que a su vez son multiplicativos, los datos de concentración tienden a distribuirse de manera similar a lo normal. No obstante, los datos del mundo real pueden no ser tan estrictos como una distribución lognormal. Las distribuciones Weibull y Gamma tienen propiedades aproximadamente similares a las lognormales, pero son menos fuertes (tail heavy) y, por lo tanto, a veces se ajustan mejor a los datos que las lognormal.
- La distribución uniforme describe una probabilidad igual de obtener cualquier valor dentro de un rango. A veces, la distribución uniforme es útil para representar cantidades limitadas físicamente (por ejemplo, una fracción que debe variar entre 0 y 1) o para representar la opinión de un perito cuando un perito puede especificar un límite superior e inferior. La distribución uniforme constituye un caso especial de la distribución Beta.
- La distribución triangular es apropiada cuando los peritos proporcionan los límites superior e inferior y un valor preferido, pero no hay otra información sobre la FDP. La distribución triangular puede ser asimétrica.
- La distribución física es un tipo de distribución empírica en la cual se realizan juicios referentes a la probabilidad relativa de diferentes rangos de valores para una variable. Este tipo de distribución a veces es útil para representar la opinión del perito con respecto a la incertidumbre.

3.2 Ejecución de simulaciones de Monte Carlo

Selección de software

Una vez que se ha identificado la FDP con el mejor ajuste, la consideración más importante antes de ejecutar las simulaciones de Monte Carlo es qué software utilizar. Existe una amplia gama de software diferente para ejecutar simulaciones de Monte Carlo y las consideraciones para seleccionar un software incluyen:

- Las FDPs disponibles
 - El software debe incluir una amplia gama de FDPs para maximizar la capacidad para modelar la mejor opción seleccionada en el paso anterior.
- Facilidad de uso
 - Cierta software tal como los programas estadísticos como R o SAS pueden requerir cierto nivel de conocimiento del lenguaje de programación y la descarga de paquetes adicionales. Otro software puede ser más fácil de usar, más costoso o tener menos opciones de FDPs para elegir.
- Costo del software
- Relevancia al tema

- Cierta software se centra en ciertas aplicaciones de Monte Carlo, por ejemplo, las evaluaciones de riesgos financieros y por lo tanto es posible que no sea aplicable a los análisis de incertidumbre de las estimaciones de emisiones.

Ejemplos de software de Monte Carlo⁷:

- XLSTAT (<https://www.xlstat.com/en/solutions/features/simulation>) – proporciona más de 20 FDPs que se pueden utilizar para ejecutar simulaciones.
- SimVoi (<http://simvoipro.com/>) – proporciona 14 FDPs que se pueden utilizar para ejecutar simulaciones.

⁷ Este software está patentado. Los costos variarán según el tipo de usuario y la cantidad de licencias necesarias.

Número de simulaciones

Es asimismo importante especificar cuántas simulaciones ejecutar. El analista puede preestablecer el número de simulaciones o ejecutar simulaciones hasta que se establezca la medición de interés (mediana o media). En el primer caso, una regla general es usar 10,000 simulaciones, ya que estas muchas simulaciones llevan a resultados estables en la distribución de simulación (es decir, si se ejecutan 10,000 simulaciones varias veces, las distribuciones resultantes serán aproximadamente las mismas.)

Truncamiento de la distribución ajustada

Al ejecutar la simulación, los analistas deben revisar las simulaciones producidas para identificar si hay valores que no son realistas. Si hay valores poco realistas, puede ser necesario truncar la FDP ajustada, es decir, especificar los valores mínimos y / o máximos para la simulación de Monte Carlo (eliminando efectivamente los puntos de datos que se encuentran más allá del rango aceptable). Por ejemplo, para variables que tienen valores no negativos (por ejemplo, toneladas de carbono por hectárea de bosque o hectáreas de deforestación), es posible que el analista tenga que truncar las distribuciones para que solo haya valores mayores que cero (como en la Figura 6) Asimismo, es posible que sea necesario truncar ciertas FDPs a largo plazo, como las distribuciones lognormal y gamma, para evitar la simulación de valores pequeños o grandes que no son realistas (como en la Figura 7).

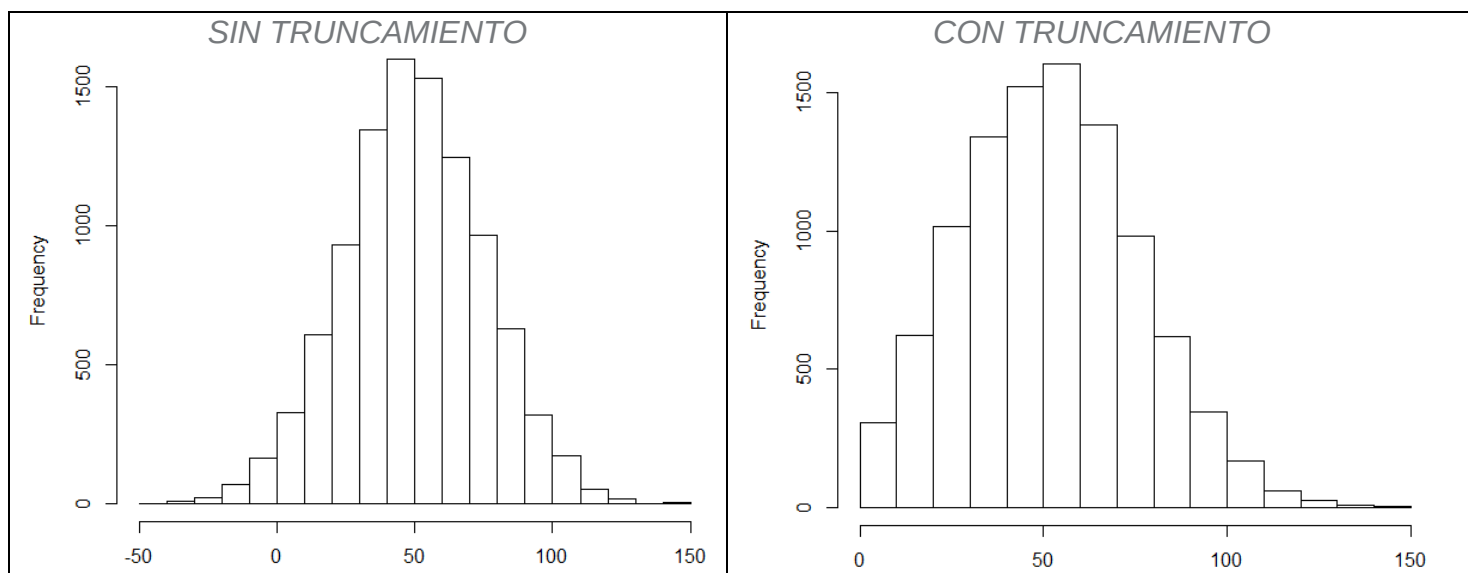


Figura 6. Simulación con distribución normal con y sin truncamiento.

En la distribución truncada, el valor mínimo se establece en cero (0).

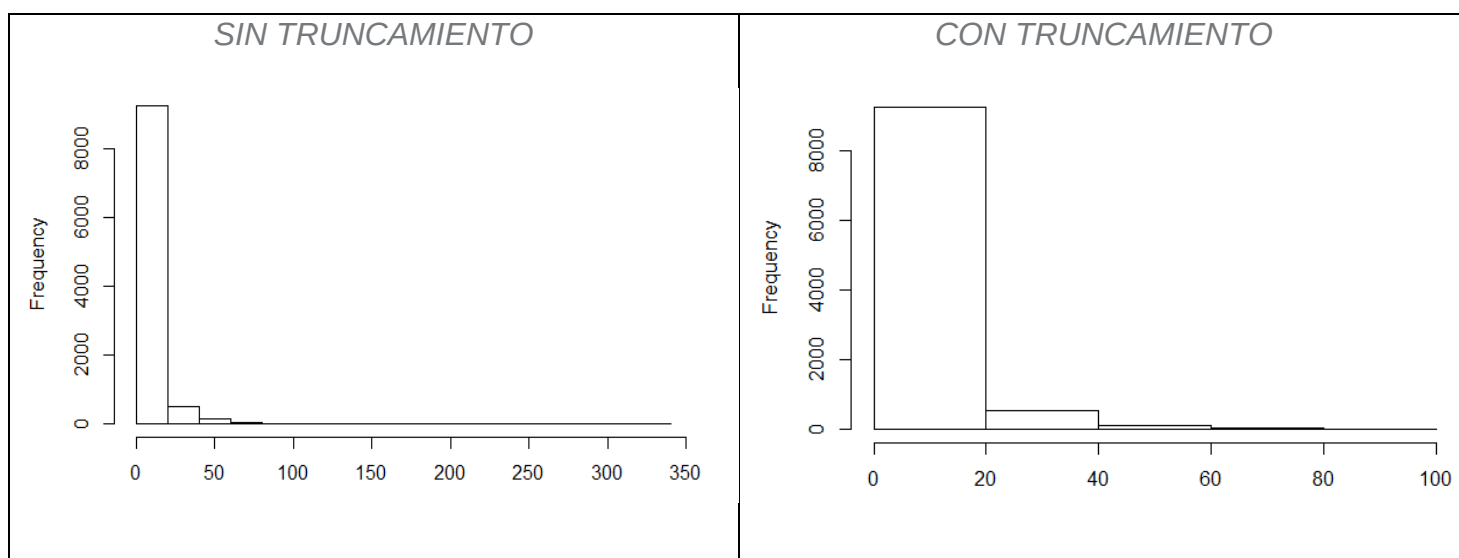


Figura 7. Simulation using lognormal distribution with and without truncation.

En la distribución truncada, el valor máximo establecido en sesenta (60).

3.3 Combinación de simulaciones de Monte Carlo

Una vez que se han ejecutado las simulaciones de las FDP de las diferentes entradas de datos, es necesario aplicar cada uno de los resultados de la simulación en las ecuaciones (por ejemplo, emisiones totales = factor de emisión * datos de la actividad) para identificar la distribución final de la estimación de emisiones (como en la Figura 8) o cualquier número final que sea de interés. El software de Monte Carlo utilizado para ejecutar las simulaciones asimismo debería poder incluir estas simulaciones automáticamente en las ecuaciones.

$$\text{Factor de emisión} * \text{Datos de actividad} = \text{Emisiones totales}$$

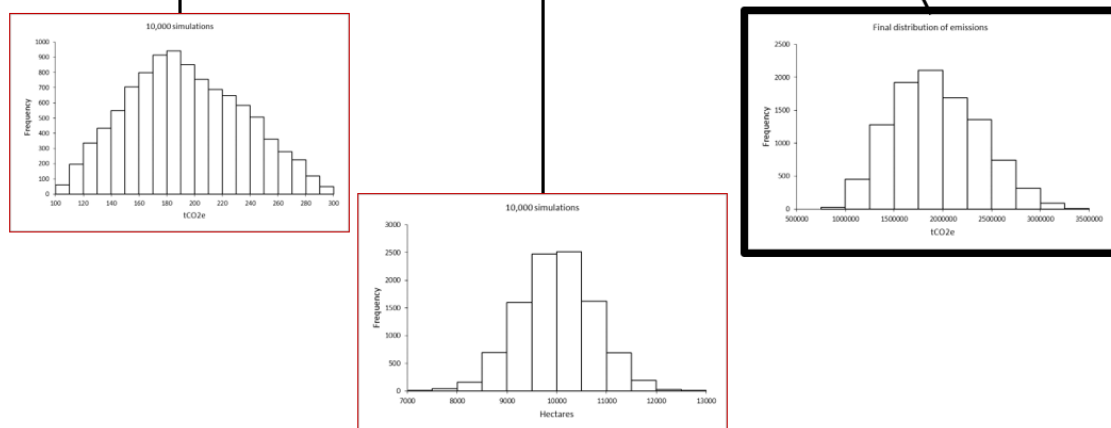


Figura 8. Ejemplo de la aplicación de simulaciones de Monte Carlo al modelo que estima las emisiones totales

Lo que está haciendo el software es insertar cada valor aleatorio producido por las simulaciones de Monte Carlo en el modelo de interés. Por ejemplo, la estimación de las emisiones totales de gases de efecto invernadero se calcula para cada ronda de simulaciones, como se muestra en la Tabla 1. La primera simulación del factor de emisión se multiplica por la primera simulación de los datos de actividad para identificar una simulación de emisiones totales; la segunda simulación del factor de emisión se multiplica por la segunda simulación de la actividad para identificar otra simulación de emisiones totales. Estos cálculos continúan para cada ronda de simulación hasta la diezmilésima simulación. La distribución final de las emisiones totales, como se muestra en la Figura 8, representa los cálculos de todas las diferentes rondas de simulaciones.

Tabla 1. Ejemplo del proceso de cálculo de emisiones totales utilizando los valores aleatorios producidos por las simulaciones de Monte Carlo

# Simulación de Monte Carlo	Factor de emisión (tCO ₂ e)	Datos de actividad (Hectares)	Emisiones totales (factor de emisión * datos de actividad)
1	163.13	9,951.46	1,623,412.40
2	242.66	10,836.08	2,629,464.37
3	259.21	9,555.77	2,476,964.17
4	250.98	10,682.08	2,681,040.23
...	...	...	...
9,999	215.29	9,480.64	2,041,065.83
10,000	240.95	8,280.32	1,995,104.00

En algunos casos, puede haber correlaciones entre las diferentes variables y, por lo tanto, entre las distribuciones resultantes. En los casos mencionados, se debe seleccionar un software que integre esta correlación entre variables en el análisis. XLSTAT brinda esta capacidad.

3.4 Cálculo de intervalos de confianza

El método para calcular los intervalos de confianza de la medida de la tendencia central de interés (es decir, la media o la mediana) dependerá de si la distribución es normal o no. Las pruebas de bondad de ajuste que se analizan previamente en la Sección 3.1 pueden identificar la normalidad de los datos.

Intervalos de confianza para distribuciones normales

Al igual que con el método de propagación del error, si la distribución final es normal, se puede calcular el intervalo de confianza utilizando la siguiente ecuación:

$$\bar{x} \pm z * \frac{\sigma}{\sqrt{n}}$$

En la cual:

$\bar{x}$ = la media de muestra de la distribución

z = z-valor para un nivel de confianza dado

σ = desviación estándar de la media

n = número de simulaciones

Lo más probable es que el número de simulaciones sea muy alto (por ejemplo, 10 000) y, como resultado, el intervalo de confianza será pequeño.

Intervalos de confianza para distribuciones no normales

Cuando la distribución final no es normal, existen diferentes métodos para calcular los intervalos de confianza para las medidas de tendencia central. (Para las distribuciones no normales, la mediana generalmente se considera una medida más representativa que la media). Describimos un método común conocido como “*bootstrapping*”⁸.

En el bootstrapping, la población (en este caso, todos los resultados de la simulación) se remuestrea un cierto número de veces con reemplazo para estimar el valor del parámetro de interés (por ejemplo, la mediana o la media) de la población. El muestreo con reemplazo significa que una vez que se ha seleccionado una unidad, se devuelve a la población antes de que se seleccione la unidad siguiente.

En cada evento de remuestreo, se vuelve a calcular la mediana o la media. Esto produce una distribución de medias o medianas o cualquier otro parámetro de interés. Por ejemplo, la Figura 9 muestra la distribución final producida a través del enfoque Monte Carlo, junto con la mediana de la distribución. Mediante el bootstrapping, las medianas se vuelven a muestrear de la distribución final de emisiones mil veces.

⁸ Dado que el bootstrapping no depende de la distribución de los datos, se puede aplicar asimismo a los datos normales.

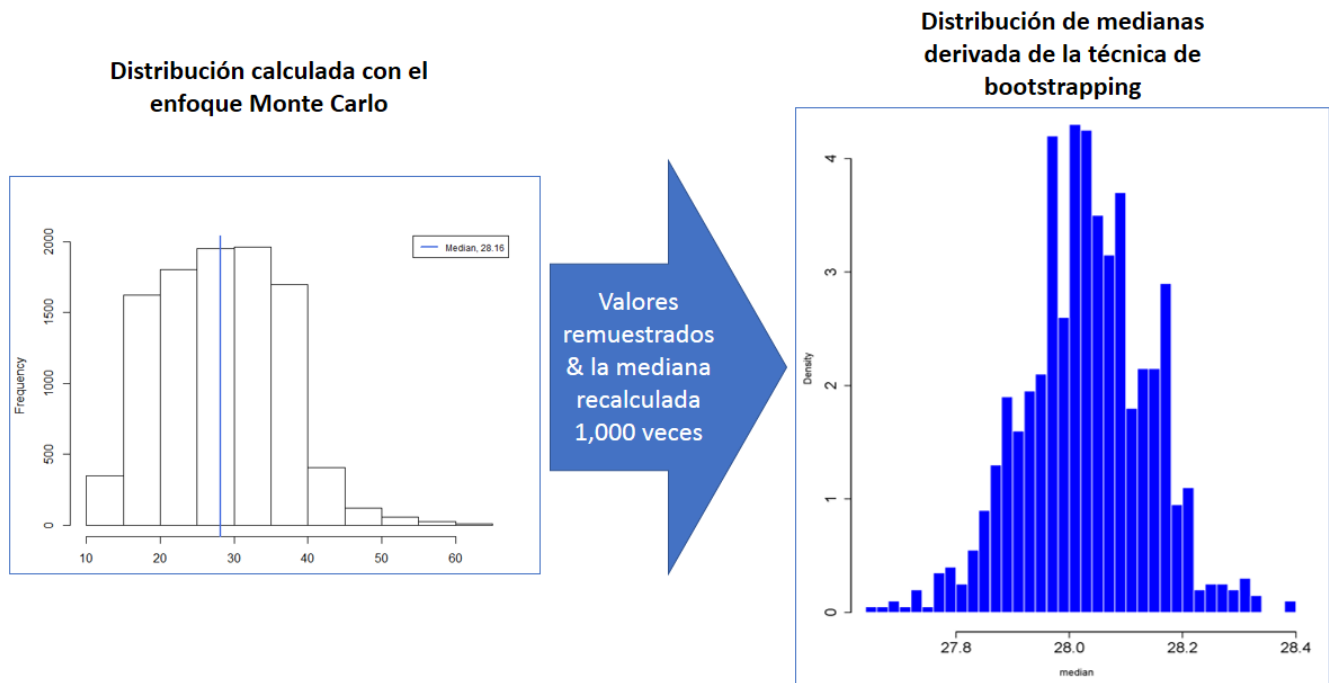


Figura 9. Los valores medios de una población se remuestran 1,000 veces, usando la técnica de bootstrapping

Existe una serie de formas para calcular los intervalos de confianza a partir de la distribución de remuestreo (bootstrapped)⁹. En el **método percentil**, para cualquier intervalo de confianza dado (por ejemplo, 95% o 90%), se supone que el valor verdadero de la estadística (es decir, mediana o media) caerá dentro de los percentiles asociados de la distribución bootstrapped de esa estadística. En el caso de un intervalo de confianza del 95%, la amplitud del intervalo sería la diferencia entre el percentil 2.5 y el percentil 97.5, como se muestra en la Figura 10. El beneficio de usar el método percentil para otros métodos es que se puede aplicar a cualquier tipo de distribución bootstrapped.

⁹ Estos incluyen, pero no se limitan a, el método percentil, el método normal, el método studentized-t, y el método corregido por sesgo y acelerado.
<https://www.unc.edu/courses/2007spring/biol/145/001/docs/lectures/Sep17.html>.

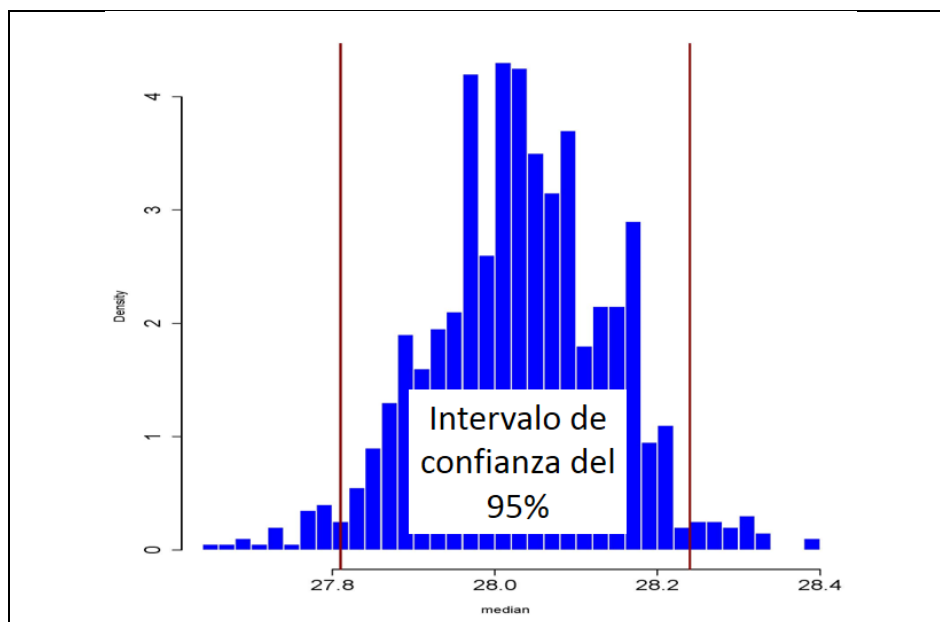


Figura 10. El intervalo de confianza calculado mediante el método de remuestreo (bootstrapping) es la diferencia entre el percentil 2.5 y el percentil 97.5 de la distribución inicial de las medianas

Como otro ejemplo, en el **método normal**, el error estándar de la distribución de remuestreo (bootstrapped) (la misma que la desviación estándar) se aplica para identificar el intervalo de confianza del 95% (media + 1.96* error estándar) de la distribución de remuestreo (bootstrapped). Este método, no obstante, solo es aplicable cuando la distribución de remuestreo (bootstrapped) es normal.

El “bootstrapping” se puede completar en software estadístico incluyendo Excel add-on XLSTAT.

Como en el caso de la distribución normal, debido al alto número de simulaciones, el intervalo de confianza será muy pequeño.

3.5 Cálculo del porcentaje de incertidumbre

Una vez que se ha identificado el intervalo de confianza, el analista calcula el% de incertidumbre de la misma manera que si utilizara el método de propagación de error, usando la siguiente ecuación.

$$\% \text{ de incertidumbre} = \frac{\frac{1}{2} \times (\text{Amplitud del intervalo de confianza})}{\text{Estimación de la emisiones}} \times 100$$

4. Aplicación completa del enfoque Monte Carlo

El IPCC presenta Monte Carlo como un enfoque solo para calcular la incertidumbre. No obstante, no menciona que si se aplica Monte Carlo para calcular la incertidumbre, constituye asimismo una buena práctica aplicarlo al calcular las emisiones y eliminaciones totales (o cualquier otro valor final) por dos razones principales:

1. Aplicar simulaciones Monte Carlo a las ecuaciones de emisión lleva a estimaciones más precisas de las emisiones finales. Esto se debe a que Monte Carlo tiene en cuenta todo el rango y la forma de la distribución de los datos de entrada, en contraste con las estimaciones únicas de los datos de entrada que se utilizan normalmente, como las medias o las medianas.
2. Cuando se aplica Monte Carlo solo al análisis de incertidumbre y no a todo el análisis de emisiones, los intervalos de confianza calculados a través del método de Monte Carlo pueden ser para una estimación completamente diferente de la calculada sin Monte Carlo (es decir, con la aplicación de estimaciones simples y no-simuladas de datos de entrada, como medias o medianas, para calcular las emisiones finales).

5. Discusión sobre la aplicación de Monte Carlo a los análisis de incertidumbre

Las simulaciones de Monte Carlo permiten estimar la incertidumbre en condiciones más flexibles (incluyendo los datos no normales o las correlaciones entre los datos de entrada) que las que se requieren para la propagación del error. Además de identificar la incertidumbre, las simulaciones de Monte Carlo producen asimismo estimaciones de emisiones que son más firmes.

Como se menciona en la sección 3.4, la aplicación de simulaciones de Monte Carlo a los análisis de incertidumbre recomendados por el IPCC, en los cuales la incertidumbre se cuantifica utilizando intervalos de confianza, llevará a incertidumbres mínimas. Estas bajas incertidumbres probablemente reflejen los firmes resultados producidos por las simulaciones, que se basan en la forma y en el alcance de los datos subyacentes (es decir, las FDPs ajustadas), en los cuales se combinan y modelan las incertidumbres. Esto es especialmente cierto cuando se aplica Monte Carlo para derivar cálculos de emisión además de solo incertidumbre. No obstante, la razón fundamental de las bajas incertidumbres es el gran número de simulaciones (por ejemplo, 10.000) ejecutadas en general, ya que el alto número de simulaciones (el número de muestra) lleva inevitablemente a pequeños intervalos de confianza (una función de n en el cálculo de error estándar y valores t).

Monte Carlo en muchas situaciones, si no en la mayoría, especialmente cuando solo se aplica a la incertidumbre, subestimarán por lo tanto la incertidumbre general. La solución más simple a este problema sería limitar el número de simulaciones. Sin embargo, los autores de este informe no recomiendan esto, ya que los números bajos de simulación (por ejemplo, 100 o incluso 1,000) probablemente no conduzcan a distribuciones estables y confiables. Y el número de simulaciones (y, por lo tanto, el tamaño de la muestra) sería más arbitrario que el valor más comúnmente seleccionado de 10,000.

Otra opción sería presentar la incertidumbre en términos que sean más independientes del tamaño de la muestra, por ejemplo, presentando solo la desviación estándar de la estimación final de emisiones. Esto capturaría el rango de los datos en el modelo de simulación, pero no permitiría que la incertidumbre reportada sea impulsada por el

número de simulaciones seleccionadas. La desventaja, especialmente en los acuerdos de informes globales (tales como REDD+), es que no existe un método para incluir medidas o criterios del tipo de desviación estándar para indicar la aceptabilidad o las deducciones relativas.

Asimismo, se podría informar la incertidumbre en términos de intervalos cuantiles específicos de la distribución final de Monte Carlo (que no se debe confundir con la distribución de remuestreo). Los valores del percentil 2.5 y 97.5 capturarían el 95% de los valores simulados. La Figura 11 muestra el cálculo de los valores del percentil 2.5 y 97.5 de la misma distribución utilizada como un ejemplo de la sección de bootstrapping. La amplitud del intervalo de los cuantiles se podría aplicar en la ecuación de incertidumbre final.

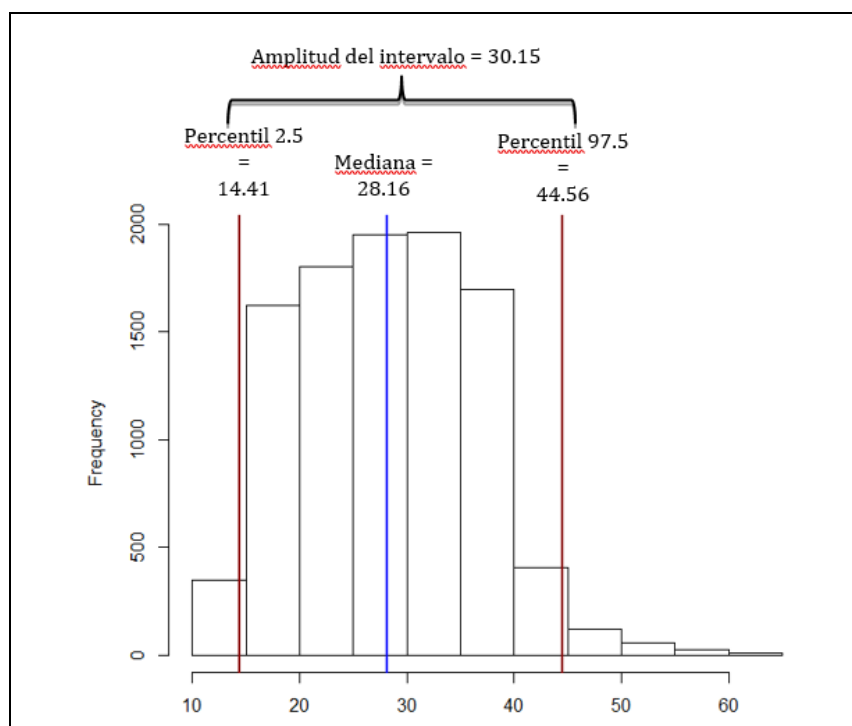


Figura 11. Ejemplos de uso de cuantiles de distribución de emisión final para calcular la incertidumbre

Tal como se muestra en la Figura 11, las incertidumbres resultantes calculadas a través del método cuantil producen el problema opuesto: la incertidumbre resultante es muy grande. Aplicar la amplitud de intervalo entre el percentil 2.5 y los valores del percentil 97.5 en la Figura 11 lleva a una incertidumbre del 53.5%.

Por lo tanto, recomendamos que los analistas apliquen el método del intervalo de confianza para calcular la incertidumbre, reconociendo la posible subestimación de la incertidumbre.

Anexo

Anexo 1: Ejemplo simplificado de aplicación del enfoque Monte Carlo

El País X está desarrollando un nivel de referencia de sus emisiones provenientes de la deforestación como parte de su programa REDD+. Los analistas a cargo del desarrollo del nivel de referencia han identificado las diferentes fuentes de incertidumbre en los datos y, dado que algunos de los datos tienen distribuciones no normales y que la incertidumbre es grande, han decidido utilizar el Enfoque 2 para calcular las incertidumbres: las simulaciones de Monte Carlo. En este ejemplo simplificado, se supone que las dos únicas reservas de carbono consideradas son biomasa aérea y subterránea. A continuación se detallan los pasos que dieron para estimar la incertidumbre de las emisiones anuales de deforestación de *un estrato forestal*.

A. Ajustar las distribuciones y ejecutar las simulaciones de Monte Carlo

Para los datos de actividad, la fuente de incertidumbre es el error en el mapeo del cambio en el uso de la tierra, específicamente el cambio de los bosques a otros usos de la tierra. Para estimar este error, los analistas aplican el enfoque presentado en Olofsson y otros. (2013)¹⁰. Se estimó la superficie deforestada en 50,000 hectáreas por año con un error estándar de 3,000 hectáreas. El enfoque de Olofsson supone una distribución de probabilidad normal y el analista ejecuta diez mil simulaciones de Monte Carlo utilizando SimVoi¹¹. La Figura A1 muestra la distribución de las simulaciones mencionadas.

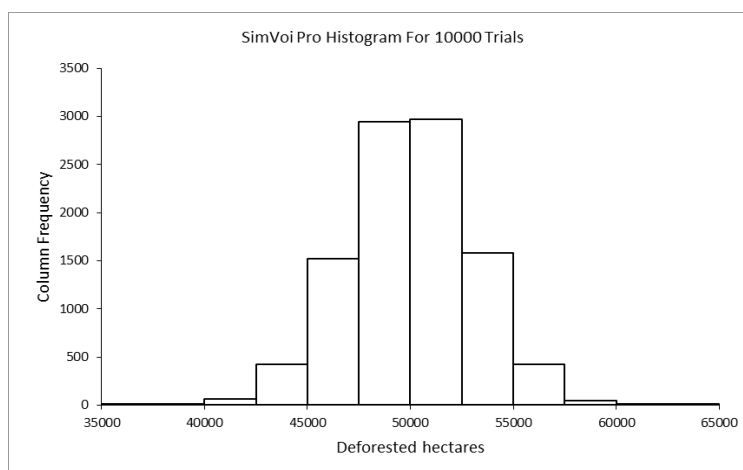


Figura A1. Simulaciones de Monte Carlo para el área deforestada

Para los factores de emisión, las fuentes de incertidumbre son el error de muestreo del inventario forestal utilizado para calcular las toneladas de carbono en la biomasa aérea, así como el error de la relación raíz:vástago utilizado para calcular la biomasa subterránea. Para la biomasa aérea, el conjunto de datos consta de 150 observaciones. Los

¹⁰ Olofsson, P., Foody, G. M., Stehman, S. V., & Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129, 122-131.

¹¹ <http://simvoipro.com/>

analistas realizan pruebas de bondad de ajuste utilizando el software EasyFit¹² y encuentran que la FDP que mejor se ajusta a los datos es lognormal, como se muestra en la Figura A2. En base a los parámetros de la distribución ajustada que proporciona EasyFit ($\mu=3.8967$; $\sigma=0.15607$)¹³, el analista puede ejecutar 10,000 simulaciones Monte Carlo en SimVoi (Figura A3).

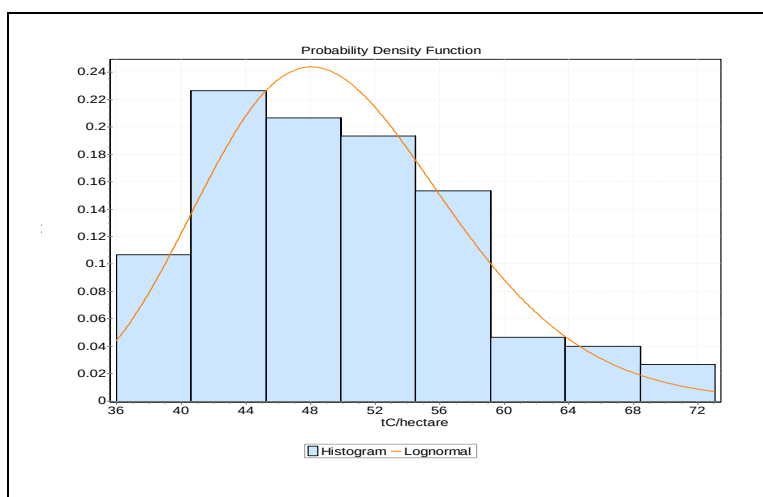


Figure A2. Distribución lognormal ajustada a los datos de biomasa aérea

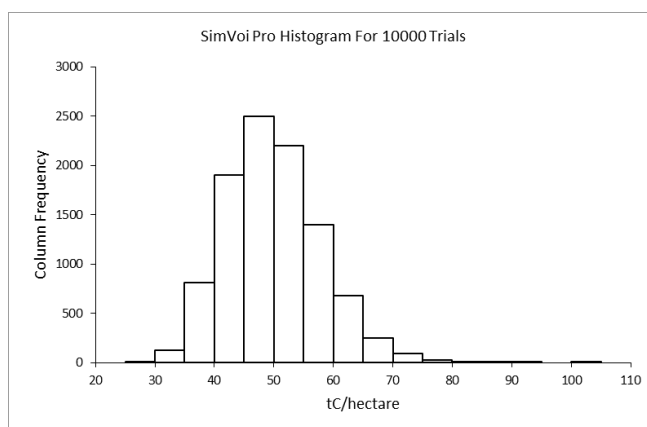


Figure A3. Simulaciones de Monte Carlo para toneladas de carbono en biomasa aérea por hectárea

No obstante, el analista advierte que algunas de las simulaciones producen valores muy altos. Los expertos en el país consideran que cualquier valor superior a 90 tC/hectárea es poco realista. Como resultado, el analista vuelve a ejecutar la simulación, esta vez truncando la distribución estableciendo el valor máximo en 90, como se muestra en la Figura A4.

¹² <http://www.mathwave.com/>

¹³ El analista debe asegurarse de que los parámetros brindados por el software de bondad de ajuste sean los mismos que los parámetros necesarios para ejecutar las simulaciones de Monte Carlo. De lo contrario, se deben convertir a los parámetros requeridos para las simulaciones. En este caso, μ y σ se deben convertir a la desviación media y estándar. Las ecuaciones para hacer esto se pueden encontrar en <https://www.mathworks.com/help/stats/lognstat.html>.

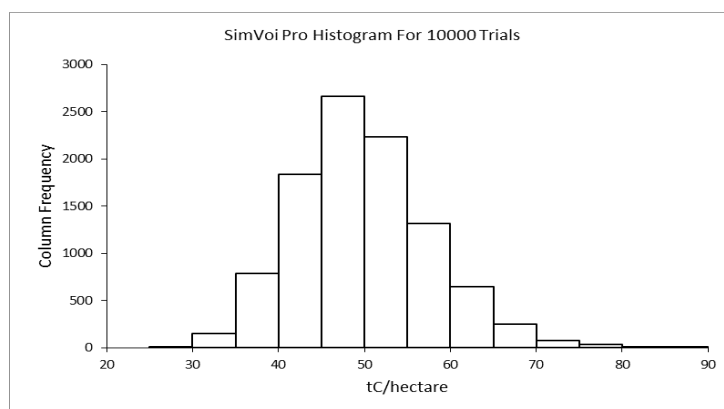


Figure A4. Simulaciones de Monte Carlo truncadas con un valor máximo de 90 toneladas de carbono en biomasa aérea por hectárea

Debido a que no hay datos específicos del país sobre la biomasa subterránea, los analistas aplican la relación raíz:vástago, 0.205, para los bosques tropicales húmedos (en qué categoría cae el estrato forestal que se analiza) que Mokany y otros (2006)¹⁴ identifica. Debido a que el estudio proporciona el error estándar, el analista puede ejecutar una simulación de Monte Carlo en base a la suposición de que la distribución es normal. La simulación de Monte Carlo resultante de la relación raíz:vástago se muestra en la figura A5.

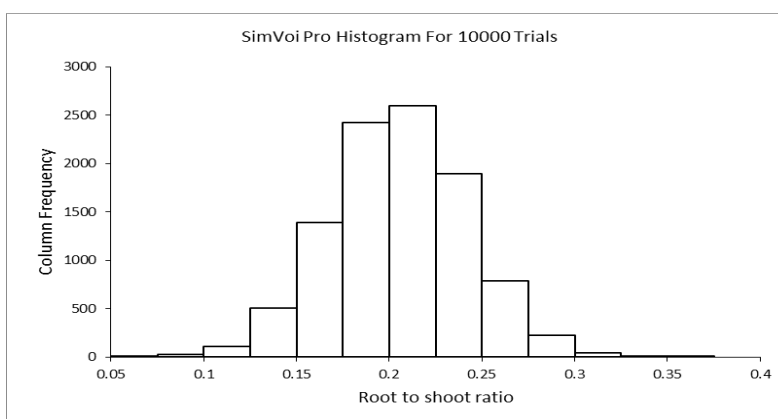


Figure A5. Simulación Monte Carlo de la relación raíz:vástago

¹⁴Mokany, K., Raison, R., & Prokushkin, A. S. (2006). Análisis crítico de las relaciones raíz: vástago en biomas terrestres. *Biología del cambio global*, 12(1), 84-96.

B. Aplicación de simulaciones de Monte Carlo a ecuaciones para calcular la incertidumbre de las emisiones totales

Con el fin de identificar la distribución de probabilidad del carbono en la biomasa subterránea, el analista multiplica las simulaciones del carbono en la biomasa aérea por las simulaciones de la relación raíz: vástago, como en la siguiente ecuación en la Figura A6.

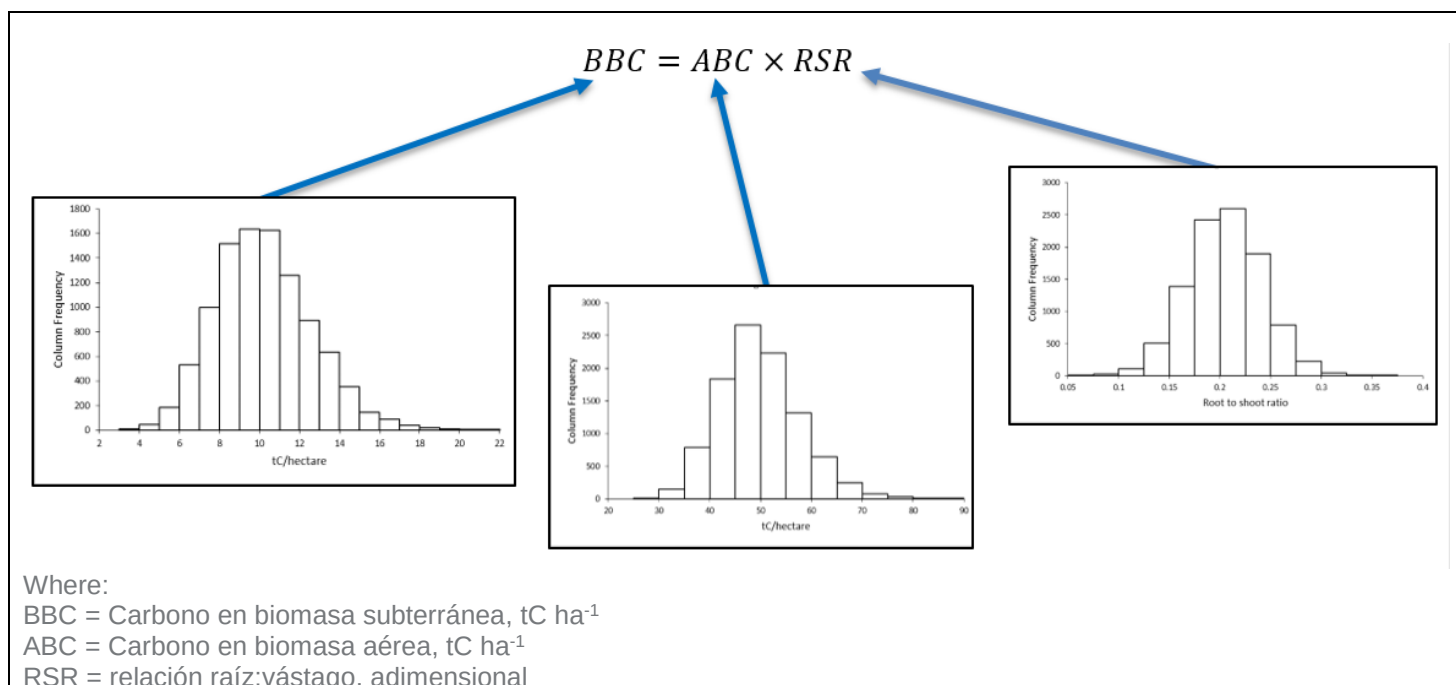


Figure A6. Aplicación de simulaciones de Monte Carlo en la ecuación para identificar la distribución de probabilidad del contenido de carbono en la biomasa subterránea

Para calcular la cantidad total de dióxido de carbono en los dos depósitos de carbono que se analizan (es decir, el factor de emisión), se aplicaron las simulaciones de carbono en biomasa aérea y subterránea en la ecuación de la Figura A7.

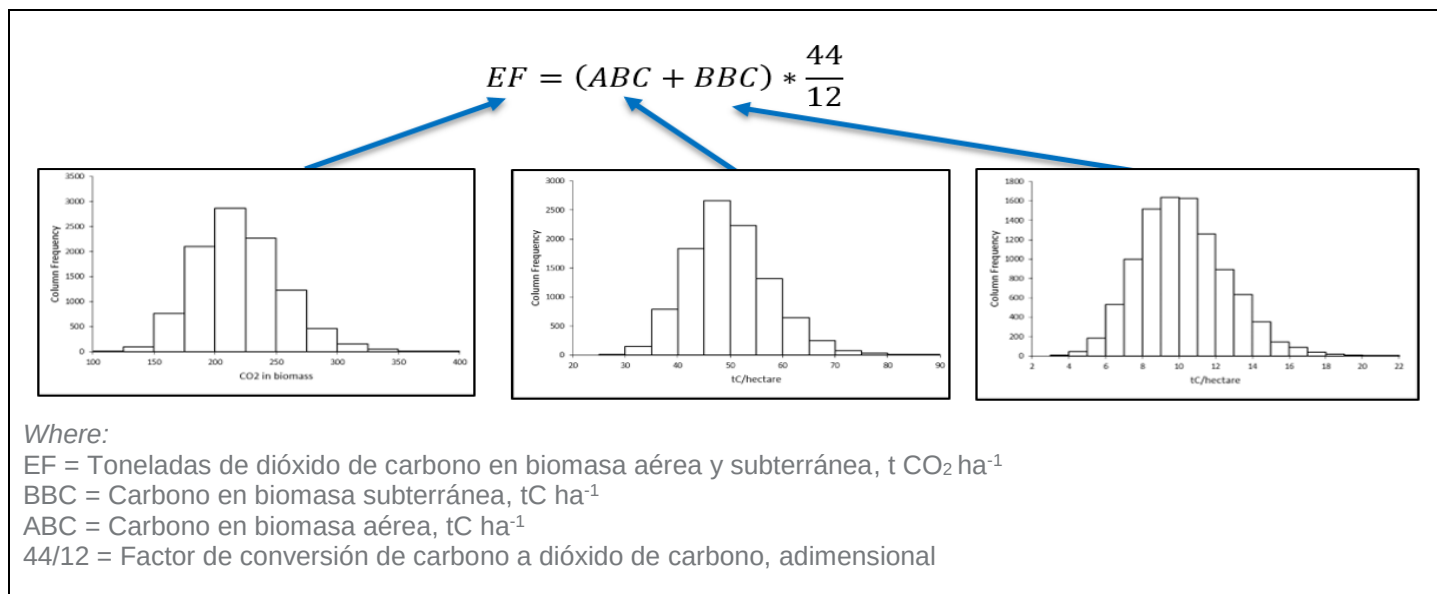


Figure A7. Cálculo de la distribución del factor de emisión en base a las distribuciones de biomasa aérea y subterránea

Con el fin de identificar la distribución de probabilidad de las emisiones totales de la deforestación para el estrato forestal en cuestión (Figura A8), las distribuciones para el factor de emisión calculadas en la Figura A7 se multiplican por la distribución de las simulaciones de Monte Carlo de los datos de actividad (superficie deforestada por año).

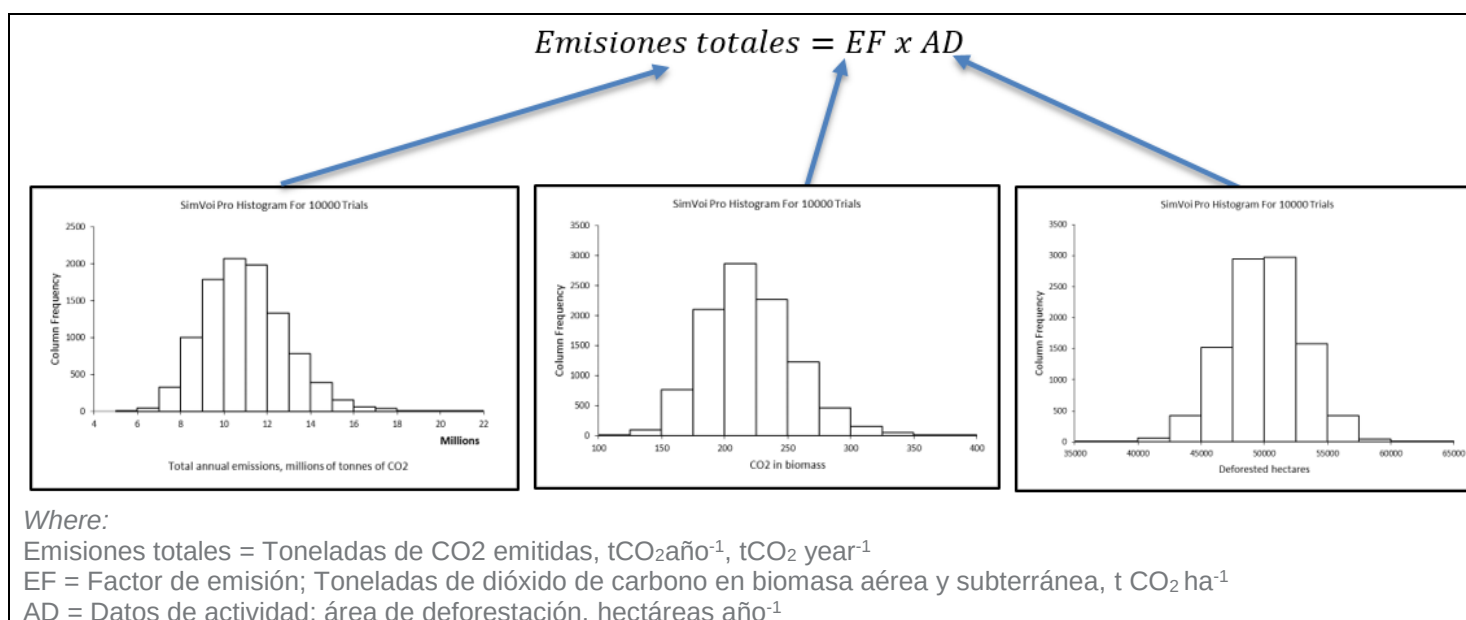


Figure A8. Calculating the distribution of annual emissions from deforestation in one forest stratum

C. Cálculo del intervalo de confianza

Una vez que los analistas tienen las distribuciones finales, primero evalúan si la distribución es normal o no. Mediante una prueba de bondad de ajuste, descubren que la distribución no es normal y, por lo tanto, aplicarán el método de bootstrapping para obtener el intervalo de confianza de la mediana del valor de las emisiones (que debería ser la mediana de la distribución). La Figura A9 muestra la distribución final de las medianas, así como el intervalo de confianza calculado mediante bootstrapping. El ancho del intervalo de confianza final (97,484) es la diferencia entre el valor del percentil 97.5th (10,940,035) menos el valor del percentil 2.5th (10,842,551) de la distribución mediana.

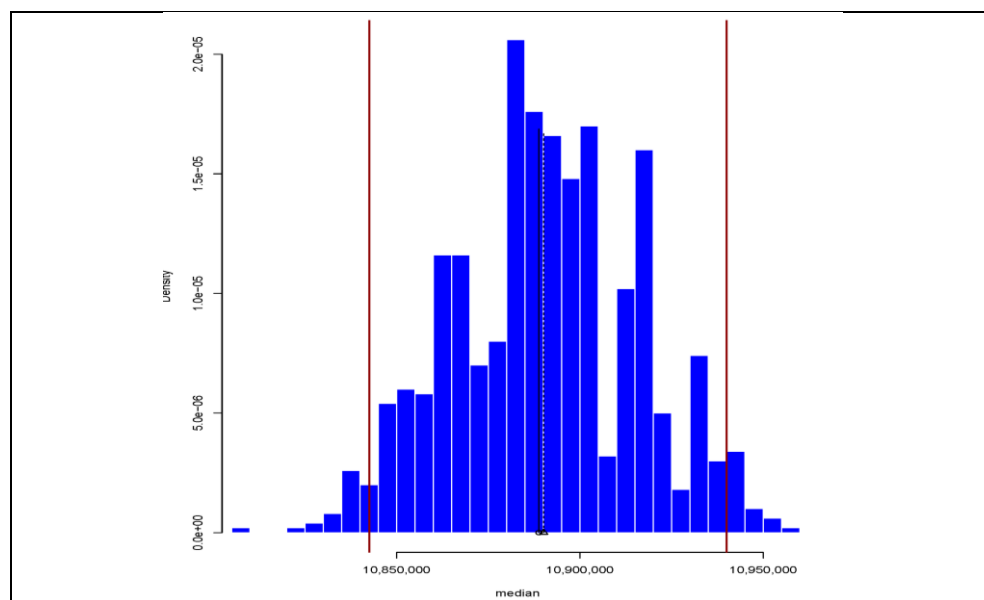


Figure A9. Distribución de las medianas de las emisiones finales calculadas mediante el bootstrapping

Las líneas rojas indican el percentil 2.5 y el percentil 97.5 de la distribución de las medianas utilizadas para identificar el intervalo de confianza de la distribución final de las emisiones.

D. Cálculo de incertidumbre

Para calcular la incertidumbre porcentual, la mitad del intervalo de confianza se divide por la estimación de emisión (la mediana de la distribución de emisiones final) y se multiplica por 100.

$$\% \text{ de incertidumbre} = 0.45\% = \frac{\frac{1}{2} \times (97,484)}{10,888,815} \times 100$$

Por lo tanto, el % de incertidumbre final para las emisiones derivadas de la deforestación en un estrato forestal es del 0.45%. Tal como se analizó en la Sección 5 del informe, este pequeño valor de incertidumbre es el resultado de la gran cantidad de simulaciones.