



Curso 7

Versión no interactiva

Elementos del análisis de datos

La versión interactiva de este curso está disponible gratuitamente en la siguiente dirección

<https://elearning.fao.org/>



Algunos derechos reservados. Esta lección está bajo una licencia CC BY-NC-SA 3.0 IGO (https://creativecommons.org/licenses/by-nc-sa/3.0/igo/deed.es_ES).

Lección 1: Introducción al análisis de datos	6
Introducción de la lección	6
Análisis de datos en las diversas fases de un inventario forestal	6
Principios generales del análisis de datos	7
Depuración de datos	8
Resumen.....	10
Lección 2: Estimación	12
Introducción de la lección	12
Observaciones generales sobre las estimaciones	12
Principios generales de la estimación estadística	13
Estimaciones puntuales y de intervalo: generar estimaciones sobre la localización y la dispersión	14
Distribución de las estimaciones puntuales	15
Bootstrapping y jack-knifing para estimaciones de intervalo en diseños complejos	16
Datos auxiliares en la estimación de inventarios forestales	18
Resumen.....	21
Lección 3: Modelos estadísticos para el monitoreo forestal.....	22
Introducción de la lección	22
¿Qué es un modelo estadístico?	22
Principales características de los modelos estadísticos	28
Cómo construir su propio modelo de biomasa	30
Identificar los modelos estadísticos adecuados	36
Resumen.....	38
Lección 4: Errores en el monitoreo forestal	40
Introducción de la lección	40
Observaciones generales sobre los errores en los inventarios forestales.....	40
Tipos de errores en el monitoreo forestal y su función	43
Resumen.....	51
Lección 5: Productos característicos del análisis de datos	52
Productos de los análisis de datos del monitoreo forestal: Observaciones generales	52
Tipos de productos.....	53
Principales características de los productos del análisis de datos	62
La función del análisis de datos en la elaboración de informes sobre inventarios forestales	63
Resumen.....	63

Este curso ofrece lineamientos sobre los enfoques y cálculos que suelen emplearse en los análisis de datos forestales y temas relacionados.

¿A quién va dirigido este curso?

El curso está dirigido principalmente a quienes se dedican a analizar datos de monitoreo forestal, pero puede realizarlo cualquier persona interesada en el tema. Específicamente, este curso está dirigido a:

1. Técnicos forestales responsables de la ejecución de los IFN de su país.
2. Profesionales dedicados al análisis de datos forestales.
3. Equipos de monitoreo forestal nacional.
4. Estudiantes e investigadores, como parte del material curricular en escuelas forestales silvicultura y en cursos académicos.
5. Jóvenes y nuevas generaciones de forestales.

Estructura del curso

Este curso consta de cinco lecciones.

Lección 1: Introducción al análisis de datos

Esta lección presenta las cuestiones que son relevantes para el análisis de datos característico después de la recopilación y depuración de los datos, pero que también se deben considerar durante todo el proceso de planificación y ejecución del inventario.

Lección 2: Estimación

Esta lección ofrece una visión general del proceso que genera resultados (o estimaciones) a partir de los datos de la muestra. Recuerde, sin embargo, que esta lección sólo ofrece una visión muy básica - no cubre el tema de la estimación estadística de forma exhaustiva. Si usted es un experto interesado en el tema con mayor profundidad, o se ocupa regularmente del análisis de datos del IFN, le recomendamos que complemente esta lección con libros de texto, y/o que discuta sus enfoques con estadísticos experimentados en inventarios forestales.

Lección 3: Modelos estadísticos para el monitoreo forestal

Esta lección proporciona información sobre los usos de los modelos estadísticos y profundiza en las

cuestiones que se deben considerar al utilizarlos.

Lección 4: Errores en el monitoreo forestal

Esta lección profundiza en los diversos errores aleatorios que se producen a lo largo del proceso del IFN. También describe la propagación de errores: cómo las diferentes fuentes de error se propagan al error total del resultado final.

Lección 5: Productos característicos del análisis de datos

Esta lección trata sobre los productos característicos de los análisis de datos en el monitoreo forestal y detalla los principales productos generados a partir del análisis de datos de los IFN.

Acerca de la serie

Este curso es el séptimo de una serie de ocho cursos a su propio ritmo que cubren diversos aspectos de un IFN. Aquí puede ver la serie completa.

Curso	Aprenderá sobre el curso
Curso 1: ¿Por qué un inventario forestal nacional (IFN)?	Objetivos y propósito de un IFN, y cómo los IFN contribuyen a la formulación de las políticas y a la toma de decisiones en el sector forestal.
Curso 2: Preparación de un inventario forestal nacional (IFN)	La planificación y el trabajo necesarios para establecer un IFN eficiente o un Sistema nacional de monitoreo forestal (SNMF).
Curso 3: Introducción al muestreo	Aspectos generales del muestreo en los inventarios forestales.
Curso 4: Introducción al trabajo de campo	Consideraciones para el trabajo de campo, variables a nivel de parcela y mediciones a nivel de árbol.
Curso 5: Gestión de datos en un inventario forestal nacional	Recopilación de información y gestión de datos para los IFN.
Curso 6: Garantía de calidad y control de calidad en un inventario forestal nacional	Procedimientos de GC y CC en la recopilación y gestión de datos de inventarios forestales

 Curso 7: Elementos del análisis de datos	(Este es el curso que está estudiando actualmente).
Curso 8: Resultados de los inventarios forestales nacionales Presentación de informes y difusión	Presentación de informes de los IFN y la importancia de la presentación de informes en el contexto de las acciones de REDD+.

Lección 1: Introducción al análisis de datos

Introducción de la lección

En esta lección, aprenderá sobre temas que son relevantes para el análisis de datos característico, no sólo después de la recopilación y depuración de datos, sino aquellos que se deben considerar durante todo el proceso de planificación y ejecución del inventario.

Objetivos

Al final de esta lección, usted podrá:

1. Describir la importancia del análisis de datos en las distintas fases de un inventario forestal.
2. Explicar los principios generales del análisis de datos.
3. Describir la depuración de datos y las consideraciones asociadas a la misma.

Análisis de datos en las diversas fases de un inventario forestal

Aunque el análisis de datos es pertinente a lo largo de las diversas fases de un inventario forestal, el **análisis real de los datos** se lleva a cabo **entre la recopilación de datos y la elaboración del informe**, es decir, una vez que los datos han sido registrados, organizados y depurados. El resultado final del análisis de datos está destinado a resolver las cuestiones planteadas en la Evaluación de las Necesidades de Información (ENI).

Sin embargo, las consideraciones relativas al análisis de los datos son pertinentes para todo el proceso de inventario, ya que uno de los objetivos generales de todo inventario forestal es generar una base de datos pertinente y fiable que sirva de insumo para el análisis. Al final, la **calidad de los datos** **codetermina la calidad de los resultados**.

Veamos ahora el papel del análisis de datos durante la planificación y la recogida de los datos.

Consideraciones durante la fase de planificación

- Es necesario asegurarse de que todas las variables (necesarias para obtener los resultados previstos) formen parte del protocolo del inventario.
- A menos que se prevea que los datos se puedan utilizar en el futuro para cuestiones potencialmente emergentes, se recomienda evitar el registro de variables que no sean

necesarias en el análisis.

- Deben observarse y registrarse las variables necesarias para que se puedan cumplir los requisitos de precisión y generar resultados para las unidades de referencia objetivo.
- El diseño del muestreo y de la parcela se debe definir de modo que se garantice la disponibilidad de estimadores para el análisis estadístico.
- Son necesarios protocolos de garantía de calidad de los datos. Estos incluyen:
 - la organización de medidas de formación adecuadas para los equipos de campo, tanto antes como durante la recogida de datos;
 - la definición clara y transparente de las normas de calidad de los datos; y
 - mecanismos de control adecuados.

Principios generales del análisis de datos

El análisis de los datos en los proyectos de inventarios forestales sigue los mismos principios que todo el proceso del inventario: tiene que estar **bien fundamentados metodológicamente**, ser **coherente**, **completo** y estar **documentado de forma transparente**.

Todos los pasos del análisis deben ser justificables y estar en línea con el diseño del inventario (en términos de diseño del muestreo, diseño de las parcelas y modelos utilizados).

Cumplimiento de las necesidades de información

El análisis de los datos debe abordar todas las necesidades de información que se hayan formulado antes de la recopilación de datos y, en la medida de lo posible, los problemas que hayan podido surgir durante ese tiempo.

Análisis para la futura optimización del diseño

El análisis de los datos también puede extenderse a cuestiones metodológicas y prácticas que apoyen la planificación eficiente de los inventarios de seguimiento continuos.

Éstas podrían incluir estudios longitudinales (es decir, de seguimiento en el tiempo), la evaluación del consumo de tiempo en diferentes pasos del inventario, y/o la evaluación del diseño del muestreo y de la parcela con el objetivo de identificar posibles optimizaciones en la realización de inventarios continuos.

Doble verificación de todos los análisis

Al analizar los datos para los resultados previstos, es importante volver a comprobar que todos los resultados sean correctos, incluidos los resultados intermedios.

El principio básico correspondiente puede formularse según Sutherland (1996) "Nunca creas en tus resultados", lo que significa que más que creer, hay que estar seguro y comprender plenamente los resultados y los supuestos en los que se basan.

Cualquier duda -por pequeña que sea- debe ser objeto de seguimiento, y esto es válido tanto para los resultados que parecen sospechosos como para los que parecen totalmente plausibles y creíbles.

Documentación

Cada paso del análisis se debe documentar adecuadamente en el informe del inventario, por lo general, como un volumen adicional. En el caso ideal, la documentación tiene que contener todos los estimadores, los cálculos paso a paso, la descripción de los modelos, los factores de conversión y los sistemas de indicadores utilizados, y tiene que abordar las dificultades del análisis.

Para concluir, recuerde que la documentación estará en línea con la metodología inicial definida junto con el protocolo de muestreo.

Depuración de datos

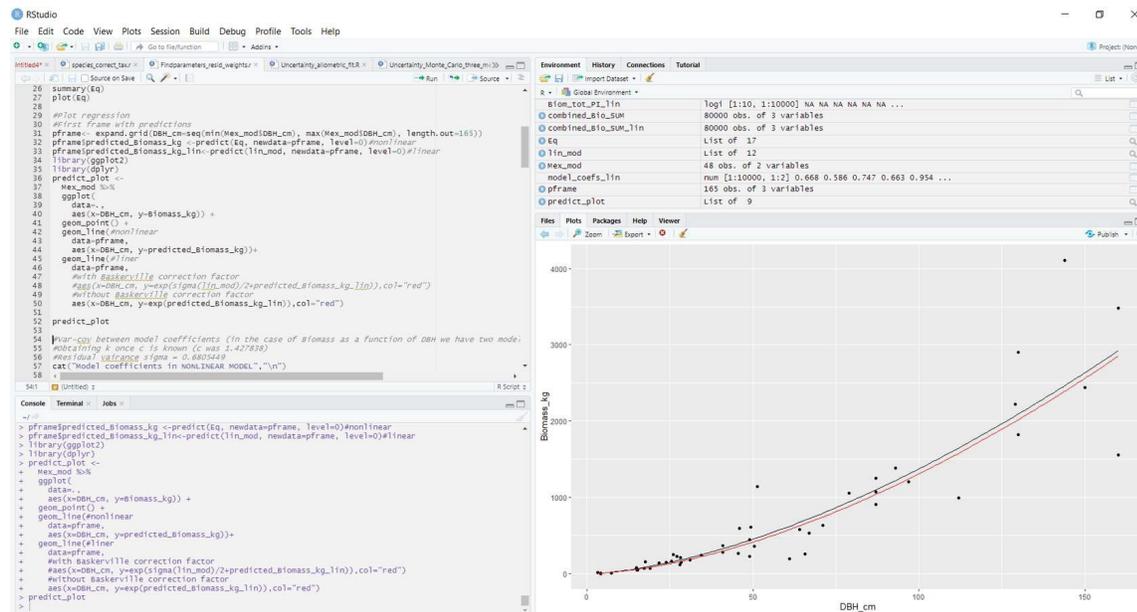
El análisis de datos sólo puede comenzar si los datos son coherentes y están depurados, y si se eliminan todos los errores e incoherencias identificables. En última instancia, es la calidad de los datos la que determina la calidad final de los resultados.

Sin embargo, a veces los errores en los datos no se pueden identificar en el proceso común de depuración y sólo se hacen evidentes cuando los resultados no parecen plausibles. Entonces es necesario volver a revisar el proceso de depuración de datos para identificar posibles errores. Conviene recordar aquí que la falta de plausibilidad no siempre es un error, y que siempre pueden producirse variabilidades inesperadas.

Consideraciones sobre el software para el análisis de datos

En la última década se ha producido un rápido desarrollo de las soluciones de software, tanto en general como, más concretamente, para los IFN. Mientras que antes los inventarios forestales se evaluaban

mediante programas hechos a medida, ya fuera en un lenguaje de programación o en un paquete de software estadístico, actualmente la tendencia es desarrollar scripts R y utilizar en la mayor medida posible paquetes R ya existentes (bibliotecas de código especializadas en la realización de tareas concretas).



Ejemplo de script R generado utilizando RStudio, un entorno de desarrollo para R, un lenguaje de programación para cálculo estadístico y gráficos.

En el caso de inventarios más pequeños (con conjuntos de datos más reducidos), el análisis también se puede llevar a cabo mediante cálculos en hojas de cálculo (como Microsoft Excel), secuencias de comandos R y software específico, como los desarrollados por la FAO. Para estas tareas se necesitan programadores cualificados que puedan trabajar en colaboración con expertos en inventarios que definen los resultados previstos.

Los pasos de análisis individuales o los flujos de trabajo completos también pueden resolverse implementando los estimadores estadísticos en un modelo de datos adecuado mediante un moderno software de Inteligencia Empresarial (BI) o procesarse directamente en un sistema de gestión de bases de datos.



¿Sabía que?

La función del software de BI en el análisis de datos forestales

La mayoría de los programas estándar de BI no incluyen los estimadores correctos que se deben aplicar para obtener resultados. Por lo tanto, si se va a utilizar software de BI, estos estimadores deben incluirse adecuadamente como fórmulas. Hasta ahora, sin embargo, esto no ha sido habitual en la mayoría de los países. Actualmente, el IFN alemán calcula todas las estimaciones utilizando la sintaxis de SQL Server directamente en la base de datos.

En cualquier caso, no existe un software que sirva para todo, sino sólo procedimientos específicos que puedan aplicarse a cada inventario forestal y que reflejen exactamente el diseño del inventario. Como cada inventario requiere un software complejo nuevo (o al menos adaptado), es necesario realizar verificaciones minuciosas para garantizar que los resultados sean correctos. Es una buena idea pedir a dos diferentes analistas de datos que hagan los mismos análisis de forma independiente, y luego comparar los resultados.

En algunos casos, estos resultados pueden parecer buenos, plausibles y coherentes, y cumplir las expectativas de los expertos en inventario, pero seguirán siendo erróneos. Recuerde que todos los resultados deben verificarse dos veces.

Resumen

Antes de finalizar, aquí están los puntos clave de aprendizaje de esta lección:

- Aunque las consideraciones relativas al análisis de datos son pertinentes para el proceso del inventario, el análisis real de los datos se da tras la recopilación de datos y es un requisito previo para elaborar informes.
- El análisis de datos en los proyectos de inventarios forestales sigue los mismos principios que todo el proceso de inventario: debe estar bien fundamentado metodológicamente, ser coherentes, completos y estar documentado de forma transparente.
- A veces los errores en los datos no se pueden identificar en el proceso común de depuración y sólo se hacen evidentes cuando los resultados no parecen plausibles. Esto implica volver a

revisar el proceso de depuración de datos para identificar posibles errores.

- Mientras que antes los inventarios forestales se evaluaban mediante programas hechos a medida, hoy (2023) la tendencia es desarrollar scripts R y utilizar en la mayor medida posible paquetes R existentes (bibliotecas de código especializadas en tareas concretas).
- No existe un software que sirva para todo, sino sólo procedimientos específicos que puedan aplicarse a cada inventario forestal y que reflejen exactamente el diseño del inventario.

Lección 2: Estimación

Introducción de la lección

La recogida de datos de campo (y algunos análisis basados en la teledetección) depende en gran medida del muestreo estadístico. La estimación es el proceso que genera resultados (o estimaciones) a partir de datos de muestreo. Como tal, la estimación es fundamental en los análisis de datos.

Dependiendo del diseño del muestreo y del diseño de la parcela utilizados, la estimación puede ser simple y también muy compleja. Esta lección ofrece sólo una introducción a la estimación basada en muestreo; si está interesado en conocerla en profundidad, consulte los libros de texto o hable con los técnicos de inventarios forestales.

Encontrará más detalles sobre este tema en el **Curso 3: Introducción al muestreo**.

Objetivos

Al final de esta lección, usted podrá:

1. Describir el papel de las estimaciones en los análisis de datos de los IFN.
2. Señalar algunos principios generales de la estimación estadística.
3. Explicar las estimaciones puntuales y de intervalo.

Discutir el papel de los datos auxiliares en la estimación de los inventarios forestales.

Observaciones generales sobre las estimaciones

Los IFN utilizan diversas fuentes de datos, desde el monitoreo de campo por muestreo (que es la base del monitoreo forestal) hasta la tecnología de teledetección en constante evolución.

Todos los productos de los análisis de datos de los IFN son, por lo tanto, estimaciones, y proceden de:

1. observaciones de las variables objetivo en la muestra de las parcelas de campo (las denominadas estimaciones basadas en el diseño);
2. respaldadas por datos auxiliares (normalmente procedentes de SIG o teledetección; en este caso se denominan estimaciones asistidas por modelos; consulte la Lección 3 de este curso para obtener una explicación completa); o bien
3. enteramente basadas en modelos (estimaciones basadas en modelos)..



Nota

Recuerde que **todos los resultados de los estudios de muestreo son estimaciones**, ya se trate de medias o varianzas, o de intervalos de confianza, regresiones y correlaciones. Por lo tanto, es mejor utilizar una terminología clara. Por ejemplo, en lugar de concluir "*El análisis del IFN mostró que la cubierta forestal del país es del 43,5 %*", es más exacto decir "*El IFN estima que la cubierta forestal del país es del 43,5 %*".

Las estimaciones son variables aleatorias. Esto indica que son contrarias a los valores paramétricos flujos de la población, que son constantes. Todas las estimaciones siguen una distribución con un valor medio de esa distribución (el valor esperado para el que la estimación puntual es la aproximación basada en la muestra a partir de esa distribución) y una desviación estándar que describe la variabilidad en esa distribución de valores medios estimados (estimada mediante la estimación por intervalos).

En el caso hipotético de repetir un IFN con exactamente el mismo diseño, pero diferente aleatorización, se obtendrían diferentes resultados numéricos de las estimaciones, tanto para la estimación puntual como para la estimación por intervalos.

Las estimaciones sirven para aprender sobre la población, de modo que el mayor interés no está tanto en los datos de la muestra en sí, sino en la inferencia sobre el valor verdadero de la población que ofrece la muestra. Cuanto menor sea el error estándar, más cerca se puede suponer que las estimaciones están, en promedio, del verdadero valor paramétrico. En ese caso, se percibirá la estimación como fiable.

Dado que inferimos a partir de la muestra (estimada) el valor (verdadero) de la población, los términos estimaciones basadas en el diseño, estimaciones asistidas por modelos y estimaciones basadas en modelos suelen denominarse también inferencia basada en el diseño, inferencia asistida por modelos e inferencia basada en modelos.

Principios generales de la estimación estadística

Cuando se realiza una estimación sobre bases estadísticas (en contraposición a las evaluaciones subjetivas), los cálculos deben corresponder estrictamente al diseño del muestreo y al diseño de la

parcela utilizados, lo que significa que distintos expertos tenderán a llegar al mismo resultado. Las fórmulas utilizadas para la estimación son los estimadores. Hay casos en los que se dispone de más de un estimador alternativo, pero normalmente no hay elección.

Una característica importante de los diseños de muestreo de inventarios en los IFN es que los estimadores utilizados deben ser no sesgados (o al menos aproximadamente no sesgados, como en el caso del estimador de razón). Esto significa que el valor esperado de nuestro diseño muestral debe ser idéntico al valor poblacional buscado. El valor esperado es el valor que resulta como media en el caso (hipotético) de que repitiéramos nuestro estudio de muestreo muchas veces, con el mismo diseño, pero con diferente aleatorización.

Los diseños y estimadores que hemos presentado en otros cursos de esta serie (sobre todo en el **Curso 3: Introducción al muestreo**) son casi todos no sesgados, con tres excepciones notables:

1. El estimador de razón es aproximadamente no sesgado en algunas circunstancias.
2. No existe un estimador no sesgado para la varianza del error en el muestreo sistemático, mientras que sí existen estimadores no sesgados para la media.
3. No existe un enfoque no sesgado para analizar parcelas de muestreo que incluyan los k árboles más cercanos (no tratado en este curso, pero explicado en libros de texto especializados) - un diseño de parcela utilizado a menudo en estudios ecológicos, pero no así en inventarios forestales.

En caso de que se utilice uno de estos elementos de diseño en un inventario, deberán abordarse de forma transparente los problemas que surgen al utilizar estos estimadores sesgados. Esto es especialmente importante en el caso del muestreo sistemático, ya que es el diseño de muestreo más utilizado en los IFN. En este caso, sin embargo, optamos por la vía segura al aplicar el marco de estimación del muestreo aleatorio simple (MAS) para la estimación de la varianza del error, porque sabemos que dicho enfoque arroja una estimación conservadora y siempre sobrestima la verdadera varianza del error (aunque en una medida desconocida).

Estimaciones puntuales y de intervalo: generar estimaciones sobre la localización y la dispersión

Al analizar los datos de los IFN, nos interesan principalmente las estimaciones puntuales y de intervalo.

En este apartado nos centraremos en ellas.

La **estimación puntual informa sobre el punto del eje numérico en el que se encuentra la estimación** (por ejemplo, para la biomasa aérea en función de la superficie 200 Mg/ha), y la **estimación de intervalo informa sobre la variabilidad estimada de esta estimación puntual** (por ejemplo, SE% =5 %). Utilizando la terminología de la estadística descriptiva, también podemos decir: la **estimación puntual es una medida de la localización de la estimación**, mientras que la estimación por **intervalos es una medida de la dispersión de las estimaciones**.

Cuando se habla de estimaciones puntuales, a menudo se hace referencia al valor medio, pero también a la estimación de un coeficiente de regresión (b_1) o de un coeficiente de correlación (r), r , o de la varianza poblacional (s^2) que son estimaciones puntuales.

Las estimaciones puntuales proporcionan la información básica a los usuarios de los datos. En la mayoría de los casos, los no expertos centran su interpretación de los resultados en estas medidas de localización de las estimaciones. Es importante reiterar que estas estimaciones puntuales no son la verdad, sino sólo estimaciones.

El hecho de que cualquier estimación puntual no sea idéntica al verdadero parámetro poblacional deseado no es una expresión de sesgo, sino una expresión de la variabilidad en el muestreo, o del error de muestreo. Sin embargo, se desconoce hasta qué punto esta estimación en particular (obtenida a partir de nuestro único estudio de muestreo, o IFN) se aleja, en términos numéricos, del verdadero parámetro poblacional.

Distribución de las estimaciones puntuales

Para poder realizar una inferencia probabilística sobre el valor verdadero de la población, es necesario conocer la distribución de las estimaciones puntuales. Por ejemplo, en el caso de los valores medios, se sabe que siguen la distribución t para muestras pequeñas y la distribución normal para muestras grandes; donde grande, en estadística, suele definirse como $n \geq 30$ (para un n grande, la distribución t se aproxima a la distribución normal).

Cuando sabemos que las medias estimadas varían según la distribución normal en torno al valor esperado (la media real en caso de un estimador no sesgado), se pueden utilizar las densidades de probabilidad bajo esta distribución normal para estimar la probabilidad de que dicho valor

verdadero se encuentre dentro de un intervalo definido en torno a la media estimada que procede de nuestro estudio muestral (IFN).

Para la media estimada, dicho intervalo es simétrico alrededor de la media estimada y tiene una desviación estándar que corresponde al error estándar. Aquí, debemos ser conscientes de que el error estándar se calcula a partir de la varianza poblacional estimada, s^2_y .

Esta varianza poblacional estimada es una estimación por sí misma (la estimación muestral de la varianza poblacional): puede considerarse una estimación puntual (el valor de la varianza poblacional estimada) que conlleva una estimación de intervalo (la variabilidad de las varianzas poblacionales).

También se podrían estimar los intervalos de confianza para la varianza poblacional estimada y para las varianzas estimadas, el intervalo de confianza será asimétrico, ya que las varianzas estimadas siguen la distribución (asimétrica) F.

$$S_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}$$

Las estimaciones de intervalo también son pertinentes en los IFN porque son una medida de la precisión de la estimación y, por tanto, de la incertidumbre, que a su vez suele interpretarse como una medida de la fiabilidad de las estimaciones.

Bootstrapping y jack-knifing para estimaciones de intervalo en diseños complejos

En algunos diseños de inventario en los que un estimador es excesivamente complejo, está indicado realizar un remuestreo. El remuestreo es una simulación en la que los datos de la muestra se aprovechan para simular muchas muestras (submuestras) y a partir de los resultados correspondientes se hacen inferencias a las estadísticas de todo el estudio de muestreo.

El bootstrapping es la técnica más utilizada cuando se determinan intervalos de confianza en diseños complejos en los que no se dispone de estimadores directos no sesgados. Se remonta a 1979, cuando Bradley Efron lo acuñó e introdujo como una modificación de la técnica jack-knifing que mucho antes había sido introducida por Quenouille (1956), el llamado "remuestreo dejando uno fuera".

Estas técnicas se basan en simulaciones y no en suposiciones sobre los parámetros de una distribución

específica de las estimaciones; por ello, también se denominan técnicas no paramétricas.

La idea de bootstrapping sigue la idea abordada anteriormente de volver a realizar un muestreo a partir de la muestra de tamaño n que se había tomado. Se puede hacer "con o sin reemplazo".

Bootstrapping "sin reemplazo" significa que se extrae un gran número de veces una submuestra de tamaño n de la muestra de tamaño original, y las que se han seleccionado no se vuelven a colocar en el grupo para volver a seleccionarlas (en otras palabras, sólo se pueden seleccionar una vez).

Sin embargo, el bootstrapping "con reemplazo" significa que la misma observación puede seleccionarse varias veces en la muestra bootstrap. Qué el término "grande" significa que este muestreo se repite varios miles de veces, digamos 10 000 veces.

Tabla 1

Muestra aleatoria del número de árboles en 25 1 ha. Parcelas en un bosque

63	84	92	92	124	145	152	154	162	164	174	189	192
223	269	294	323	344	354	358	368	372	463	477	900	

Tabla 2

Remuestreos bootstrap de la Tabla 1

Muestra 1

63	84	92	92	152	152	154	154	162	174	189	192	223
269	294	354	354	358	358	368	372	463	477	477	900	

Muestra 2

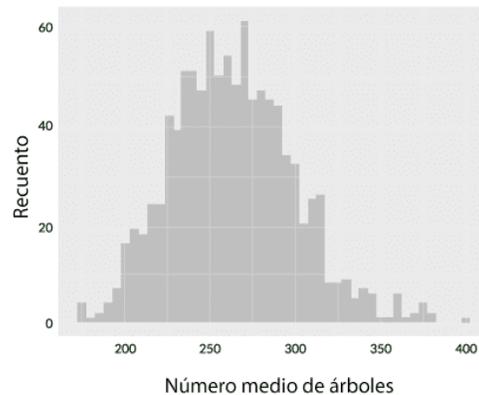
84	91	92	154	154	164	174	174	174	189	192	223	294
294	323	358	358	358	368	372	372	463	477	900	900	

Muestra 1000

63	92	145	152	152	154	154	154	162	164	164	189	192
192	192	223	223	294	344	368	368	463	477	900	900	

Remuestreo

1000 muestras



Histograma de 100 medias muestrales bootstrap a partir de un muestreo aleatorio con sustitución de los datos de la Tabla 1

Para cada una de estas muestras bootstrap se calcula el estadístico objetivo (por ejemplo, el valor medio), de modo que al final hay, digamos, 10 000 valores medios bootstrap y éstos pueden representarse gráficamente como una distribución. Esta distribución tiene un valor medio (que corresponde, por supuesto, al valor medio de la muestra original de tamaño n) y una distribución de densidad de probabilidad particular de la que, para cualquier probabilidad, pueden obtenerse intervalos de confianza.

Si, por ejemplo, se deben calcular los límites de un intervalo de confianza del 95 %, se buscan los puntos de corte en los que el 2,5 % de las medias bootstrapped se truncan en el extremo superior y el 2,5 % en

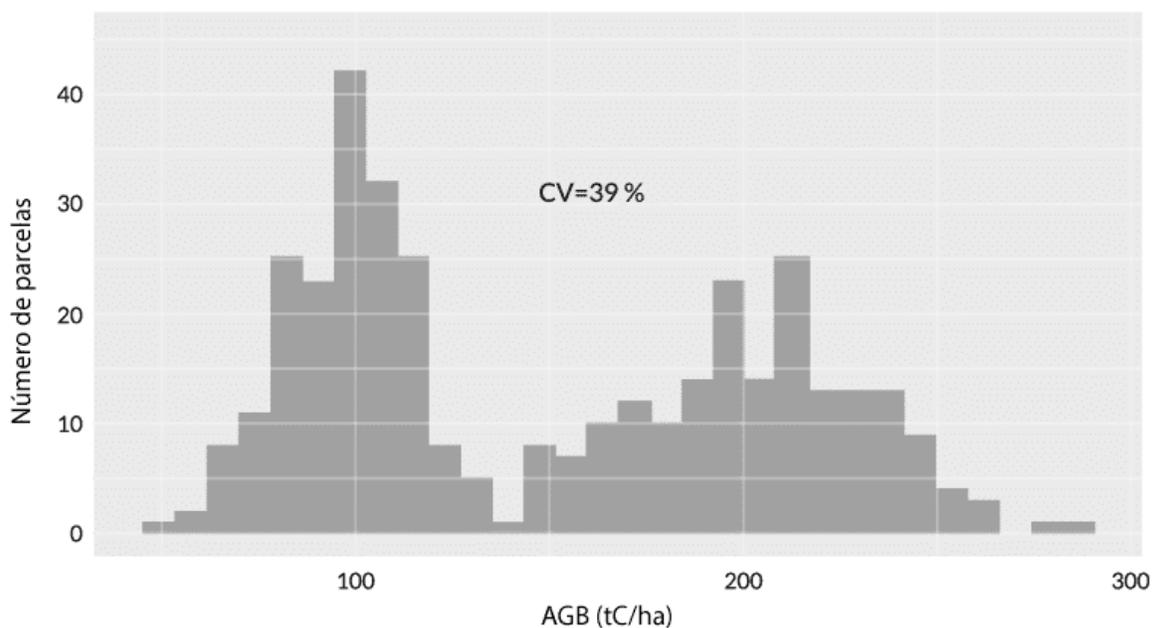
el extremo inferior. Los dos "puntos de corte" se toman como los límites superior e inferior del intervalo de confianza del 95 %.

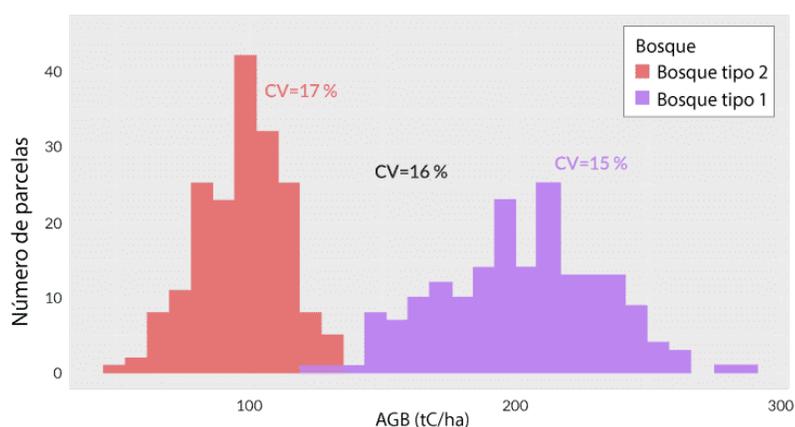
Datos auxiliares en la estimación de inventarios forestales

Los datos auxiliares proceden de la observación de variables auxiliares en algunos diseños de inventarios para mejorar la precisión de la estimación de las variables objetivo. Las variables auxiliares se denominan a veces covariables (latín: *auxilium*=ayuda, *ancilla*= sirviente). Hasta ahora hemos visto variables auxiliares en los estimadores de razón y de regresión, en los que aprovechamos una alta correlación entre las variables objetivo y las auxiliares para extraer e integrar información de la variable auxiliar en la estimación de la variable objetivo.

En un sentido más general, al distinguir únicamente entre variables objetivo y auxiliares, podemos considerar otras variables como auxiliares (de apoyo a los análisis). Esta observación se refiere, por ejemplo, a todas las variables topográficas que sirven para desglosar los resultados de nuestras variables objetivo en clases, por ejemplo, existencias en formación por clase de elevación o biomasa por clase de pendiente. En este caso, dichas variables auxiliares definen criterios para una post-estratificación, permitiendo análisis específicos y evaluando las relaciones entre las variables objetivo y las auxiliares.

Veamos ahora dos figuras sobre cómo mejorar las estimaciones de error a través de la post-estratificación con la ayuda de datos auxiliares.



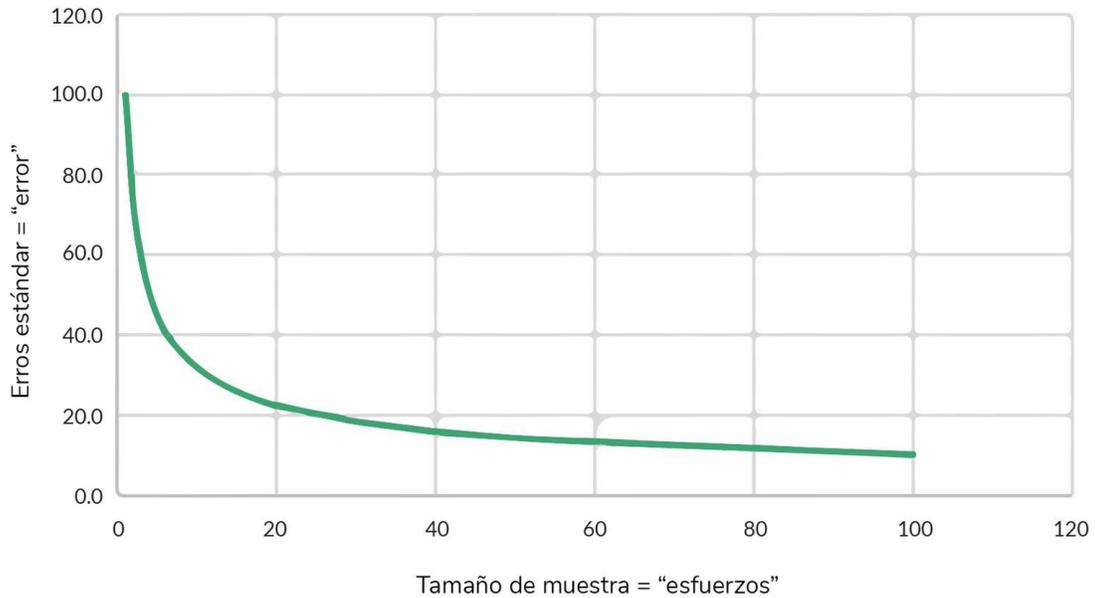


Estimaciones para diferentes unidades de referencia/subpoblaciones

La unidad básica de referencia para la estimación a partir de los IFN es todo el país. El tamaño de la muestra suele definirse de forma que la precisión de la estimación a escala nacional cumpla las expectativas. El tamaño de la muestra suele ser bastante grande y las estimaciones serán precisas, con errores estándar bajos. Dependiendo del tamaño de la muestra, los errores estándar relativos son en algunos casos inferiores al 1 %. Pero una precisión tan elevada sólo se produce cuando tenemos a todo el país como unidad de referencia = área de reporte.

A menudo, también son de interés las estimaciones para unidades subnacionales, como provincias, estados o territorios. Por supuesto, al utilizar la cuadrícula sistemática común de puntos de muestreo, el tamaño de la muestra para estas unidades de referencia más pequeñas será menor y los errores estándar de las estimaciones correspondientes serán mayores. **Cuanto más pequeña sea la unidad de referencia y menor el tamaño de la muestra, menos precisas serán las estimaciones.** Por ejemplo, en el IFN alemán, mientras que la superficie forestal de todo el país tiene un error estándar relativo de $SE\%=0,7\%$ con un tamaño de muestra de aproximadamente $n=21\ 000$ conglomerados, es de $SE\%=1,6\%$ para el Estado Federal de Baviera ($n=2\ 815$) y de $SE\%=25,8\%$ para los Estados Federales combinados de Hamburgo y Bremen con sólo $n=15$.

Resulta instructivo considerar aquí la relación simple entre el tamaño de la muestra y el error estándar en el MAS, representado en la figura siguiente; donde la forma básica de la relación se mantiene para todos los diseños de muestreo: ¡la ganancia marginal en precisión para tamaños de muestra grandes es pequeña; pero pequeños cambios en el tamaño de la muestra tienen un impacto mucho mayor en el error estándar para tamaños de muestra más pequeños!



Cuando se dispone de datos de teledetección, pueden generarse estimaciones para áreas más pequeñas con una precisión mucho mayor utilizando la llamada estimación de áreas pequeñas, en la que los datos de teledetección se utilizan como apoyo para generar estimaciones para unidades de referencia (casi) arbitrariamente pequeñas.

Aunque se puede ver una definición de la estimación de áreas pequeñas en la Lección 5 de este curso, podemos ofrecer aquí un ejemplo sencillo: Supongamos que tenemos una gran superficie forestal de 1000 hectáreas y queremos estimar la densidad promedio de árboles (número de árboles por acre) en un área pequeña de sólo 10 hectáreas dentro del bosque. Sin embargo, sólo disponemos de datos de 5 parcelas de muestreo dentro del área pequeña, lo que no es suficiente para obtener una estimación precisa de la densidad de árboles.

Mediante la estimación de áreas pequeñas con datos de teledetección, podemos utilizar la información de la superficie forestal más grande para generar una estimación más precisa de la densidad de árboles en el área pequeña. Podemos utilizar datos de teledetección, como imágenes satelitales y el sistema LiDAR, para obtener información adicional sobre la densidad de árboles y otras características del bosque en el área más grande.

A continuación, podemos utilizar estos datos de teledetección como apoyo para generar una estimación para la pequeña área de interés. Por ejemplo, los datos de teledetección podrían sugerir que la densidad media de árboles en la superficie forestal más extensa es de 400 árboles por acre. Utilizando la

estimación de áreas pequeñas, podemos ajustar esta estimación basándonos en los datos del área pequeña para obtener una estimación más precisa del área pequeña. Por ejemplo, el modelo podría estimar que la densidad promedio de árboles en el área pequeña es de 450 árboles por acre, con un margen de error menor que el que podríamos obtener utilizando sólo los datos de la muestra.

Resumen

Antes de finalizar, aquí están los puntos clave de aprendizaje de esta lección:

- Todos los resultados de los estudios de muestreo son estimaciones, ya se trate de medias o varianzas, o de intervalos de confianza, regresiones y correlaciones.
- Las estimaciones sirven para aprender sobre la población: el mayor interés no está tanto en los datos de la muestra en sí, sino en utilizar estos datos muestrales para inferir el verdadero valor de la población.
- Cuando se realiza una estimación sobre bases estadísticas, los cálculos deben corresponder estrictamente al diseño del muestreo y al diseño de la parcela utilizados, lo que significa que distintos expertos debieran llegar al mismo resultado y no se puede hacer la estimación con cualquier enfoque arbitrario.
- Al analizar muestras del IFN, deseamos utilizar estimadores no sesgados del diseño, si es posible. Para algunos diseños, tales estimadores no existen, incluidos los estimadores de la varianza del error en el muestreo sistemático.
- La estimación puntual es una medida de la localización, mientras que la estimación por intervalos es una medida de dispersión de una distribución. Esta terminología es válida tanto para las variables como para las estimaciones.
- En algunos diseños de inventario en los que un estimador es excesivamente complejo, puede estar indicada una simulación en la que los datos de la muestra se aprovechan para simular muchas muestras, lo que se conoce como "remuestreo".
- En determinadas condiciones, resulta eficaz observar también las variables auxiliares junto con las variables objetivo para mejorar la precisión de la estimación de las variables objetivo.

Lección 3: Modelos estadísticos para el monitoreo forestal

Introducción de la lección

Los modelos estadísticos están omnipresentes en el monitoreo forestal. Esta lección ofrece una visión general y una perspectiva de los usos de los modelos estadísticos en los IFN y profundiza en las cuestiones que se deben considerar a la hora de utilizarlos.

Objetivos

Al final de esta lección, usted podrá:

1. Describir la función de los modelos estadísticos en el monitoreo forestal.
2. Identificar las principales características de los modelos estadísticos.
3. Demostrar cómo construir un modelo de biomasa.
4. Identificar un modelo estadístico adecuado para una situación en particular.
5. Explicar algunos problemas habituales de información en los modelos estadísticos.

¿Qué es un modelo estadístico?

Los modelos estadísticos pretenden establecer una relación cuantitativa entre una variable predicha y una o más variables predictoras. En otras palabras, al haber medido/observado la(s) variable(s) predictor(a)s, el modelo se utiliza para generar un valor para la variable predicha.

Esencialmente, un modelo estadístico predice un valor para una variable objetivo.

En el contexto de los inventarios forestales, los modelos estadísticos se utilizan cuando:

1. una variable objetivo no puede medirse en un inventario forestal (por ejemplo, la biomasa no puede medirse pesándola). Si tala todos los árboles del bosque para pesarlos, ¡ya no tendrá más un bosque! Sin embargo, se puede modelar a partir de la medición de otras variables (denominadas variables predictoras); o bien
2. una medición requiere mucho tiempo o es costosa (por ejemplo, la altura requiere mucho tiempo y a menudo sólo se mide para un subconjunto de árboles y luego se predice mediante un modelo para los demás en función del dap).

Los modelos estadísticos describen la relación entre datos/observaciones de dos variables, frecuentemente observadas a partir de los mismos objetos (como los árboles). Los modelos estadísticos no sirven para establecer una relación causa-efecto, algo que sería el objetivo de los llamados modelos de proceso: éstos pretenden incluir explícitamente las causas que subyacen a los procesos biológicos para predecir resultados específicos y situaciones diferentes. Aunque tal relación puede existir también para los modelos estadísticos, no es objeto del ejercicio de modelación, ni un modelo estadístico puede interpretarse de ese modo.

Ejemplos de modelos estadísticos utilizados en el monitoreo forestal

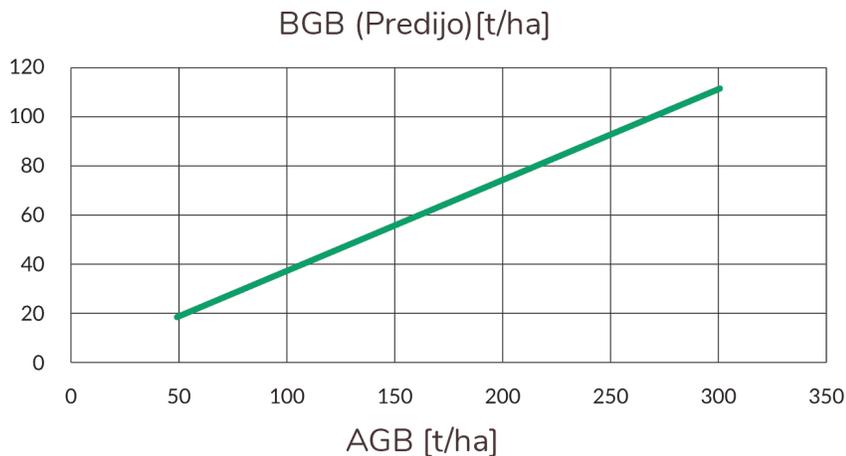
Existen diferentes tipos de modelos -de complejidad variable- que se utilizan en el monitoreo forestal y, en algunos casos, es difícil darse cuenta de que se ha utilizado un modelo.

Ejemplo 1: Determinar el área basal a partir de la medición del diámetro

Un modelo muy básico, por ejemplo, es la suposición de que la sección transversal del árbol a la altura del pecho será siempre un círculo: el área basal del árbol se calcula según el modelo circular simple. or supuesto, siendo realistas, ningún árbol tiene un círculo matemáticamente perfecto ni cuya sección transversal a la altura del pecho sea un círculo matemáticamente perfecto, pero la aproximación ha servido razonablemente bien hasta ahora, y no hay mejores soluciones.

Ejemplo 2: Factores de conversión como modelos estadísticos

Otros modelos básicos son los factores simples, ya que se utilizan con frecuencia para convertir, por ejemplo, la biomasa del tronco en biomasa total o la biomasa aérea (AGB, por su sigla en inglés) en biomasa subterránea (BGB, por su sigla en inglés). Los factores de forma -un indicador numérico resumido de la forma de un tronco- también son modelos simples, utilizados para determinar el volumen individual de un árbol. Podemos decir que estos factores de conversión no son más que modelos de regresión lineal simple "reducidos", en los que el intercepto es cero y el factor en sí es el coeficiente de la pendiente. En la siguiente figura se muestra un ejemplo



Un factor de conversión simple puede considerarse un modelo básico que permite predecir una variable a partir de una variable predictora. El IPCC (2006, Tabla 4.4), por ejemplo, recomienda el factor de conversión de 0,37 para los bosques húmedos tropicales a fin de determinar la biomasa subterránea a partir de la biomasa aérea. Ese factor de conversión se traduce en un modelo de regresión lineal simple con un coeficiente de intercepción de cero y un coeficiente de la pendiente de 0,37: $BGB=0,37*AGB$.

Obsérvese que, en este caso, el IPCC da una fuente para este factor de conversión, pero no publica un error estándar ni otras medidas de incertidumbre. En su lugar, especifica, para algunos tipos de bosque, el rango de valores del factor de conversión.

Por supuesto, con factores de conversión tan simples surge una incertidumbre considerable, pero en muchos casos, como el de la biomasa radicular, difícilmente sería conveniente tomar muestras propias.

Ejemplo 3: Modelos de regresión comunes

Los modelos de regresión comunes, como los utilizados en los inventarios forestales, incluyen:

1. predecir la altura a partir del *dap* (curvas de altura); y
2. volumen/biomasa/carbono a partir del *dap* o a partir del *dap* y la altura o a partir del *dap*, la altura y un diámetro superior (funciones de volumen, funciones de biomasa o funciones de carbono respectivamente).

Por supuesto, se utilizan otros modelos para fines específicos, por ejemplo:

- en inventarios de tocones, predecir el dap a partir del diámetro del tocón (y la altura del tocón); o para los árboles con raíces de contrafuerte, predecir el dap a partir del diámetro por encima de las raíces del contrafuerte.
- para los árboles con raíces de contrafuerte, predecir el dap a partir del diámetro por encima de las raíces del contrafuerte.

Para las funciones de biomasa, se utiliza con frecuencia el término **funciones alométricas de biomasa** o **modelos alométricos de biomasa**. El término alométrico deriva del griego antiguo y del latín, donde "ἄλλος" (allos) significa otro y métrico significa medida. Por lo tanto, alométrico significa determinar la biomasa a partir de otras variables. Siguiendo este significado original, el término es esencialmente redundante a la hora de especificar un modelo-porque simplemente describe lo que hacen todos los modelos: producir el valor de una variable a partir de mediciones de valores de otras variables.

Los tipos de modelos mencionados hasta ahora suelen generarse a partir de estudios de investigación previos al inventario. Para inventarios forestales específicos, los modelos de biomasa suelen tomarse de la bibliografía tras comprobar su idoneidad para el inventario correspondiente (tema que se trata más adelante en esta lección).

Sin embargo, también hay modelos que se generan con y a partir de los propios datos del inventario: un ejemplo típico es una curva de altura utilizada para predecir las alturas de los árboles. Las mediciones de altura requieren mucho tiempo y, por lo tanto, son caras, de modo que las alturas se miden sólo en un subconjunto (bien definido) de árboles de muestra. A partir de estas mediciones se construye un modelo que se utiliza para predecir las alturas de los árboles no medidos. En este caso, las mediciones de altura mostrarán una mayor variabilidad que las alturas predichas, ya que estas alturas predichas representan valores medios para una clase de dap determinada.

Ejemplo 4: Modelos calculados a partir de variables objetivo y auxiliares

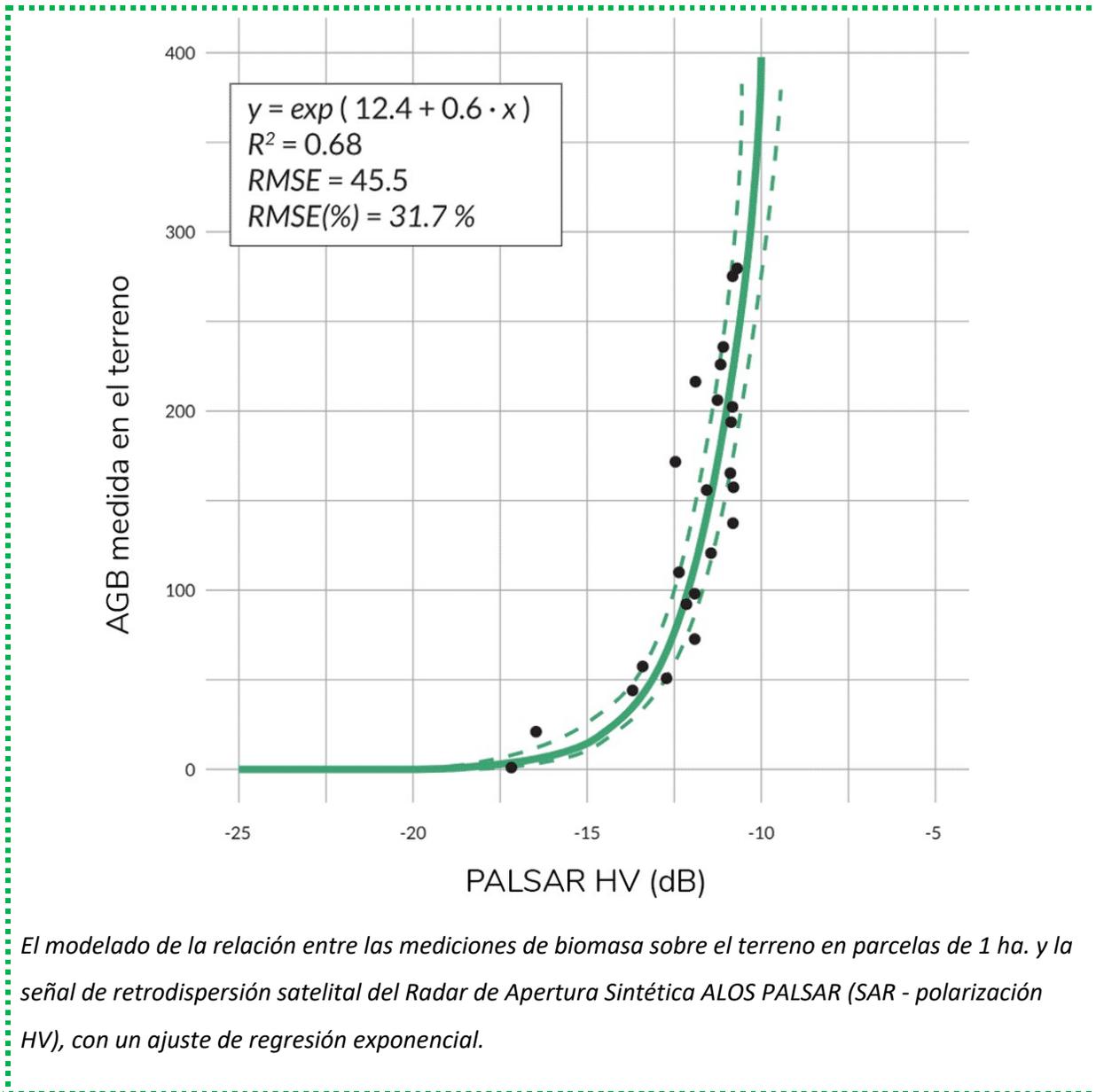
Otro ejemplo en el que se construyen modelos durante la ejecución del inventario es cuando se observan variables auxiliares para utilizarlas con los estimadores de razón o de regresión, o en el

muestreo doble con el estimador de razón o de regresión: entonces, se calcula un modelo (ya sea una razón simple o una regresión) a partir de las parcelas de muestreo en las que se han registrado ambas variables, objetivo y auxiliares. El modelo ayuda a la estimación y permite - cuando hay suficiente correlación entre las variables objetivo y auxiliares- una estimación más precisa de la variable objetivo. Es decir, el modelo ayuda al proceso de estimación, por lo que el estimador utilizado también se denomina estimador asistido por el modelo.

Aunque en este caso el proceso de estimación se apoya en el modelo, el carácter no sesgado de la estimación sigue proviniendo de la aleatorización de la selección de la muestra, que garantiza que la muestra es representativa de la población. Sin embargo, nótese que el estimador de razón sólo es no sesgado cuando se cumple el modelo simple. Esto significa que el modelo debe capturar con exactitud la relación entre las variables objetivo y auxiliares, y que no hay otros factores que puedan afectar al proceso de estimación. Si el modelo simple no funciona, el estimador de razón puede estar sesgado y los resultados de la estimación pueden no ser exactos.

La estimación y la inferencia también se pueden basar completamente en modelos, y los enfoques correspondientes se denominan estimación basada en modelos o inferencia basada en modelos. En este caso, el carácter no sesgado de la inferencia depende por completo de la validez del modelo.

Un ejemplo típico es la modelación de la relación entre la biomasa observada sobre el terreno y los datos de teledetección. Cuando se ha establecido un modelo de este tipo, es posible predecir la biomasa forestal de cada píxel. Con esta predicción, se está en condiciones no sólo de producir una estimación de la biomasa para toda la región del inventario (sumando las predicciones de biomasa por píxel), sino también de generar un mapa de biomasa.



El modelado de la relación entre las mediciones de biomasa sobre el terreno en parcelas de 1 ha. y la señal de retrodispersión satelital del Radar de Apertura Sintética ALOS PALSAR (SAR - polarización HV), con un ajuste de regresión exponencial.



Nota

De la lista relativamente larga de modelos utilizados en el monitoreo forestal se desprende que éstos desempeñan un papel crucial en este ámbito: esto se debe principalmente a que los bosques son objetos complejos de monitorear y a que diversas variables de interés no pueden observarse

directamente. Para que un sistema de monitoreo sea operativo, es esencial trabajar con valores que se predicen a partir de modelos. Sin embargo, es importante distinguir claramente entre los valores que se registran mediante observaciones o mediciones inmediatas y los que se predicen a partir de modelos.

Los dos puntos principales que debemos recordar son los siguientes:

- ➡ las observaciones inmediatas conllevan una sola fuente de error: el error de medición, mientras que las predicciones de los modelos conllevan errores de medición (de las variables predictoras necesarias) y errores de modelo; y
- ➡ las predicciones de los modelos presentan menor variabilidad que las observaciones inmediatas: para el mismo conjunto de valores de variables predictoras se genera siempre el mismo valor predicho, mientras que, en la realidad, por ejemplo, árboles diferentes con el mismo dap pueden tener biomásas bastante diferentes.

Principales características de los modelos estadísticos

Es importante comprender que, dado un conjunto de variables predictoras, los modelos no proporcionan el valor verdadero de una variable predicha. Las predicciones deben entenderse como valores medios. Al utilizar un modelo para predecir, por ejemplo, la altura de los árboles a partir de las mediciones del dap, asignamos alturas medias a los árboles de la muestra: todos los árboles con un dap específico, digamos 40 cm, tendrán la misma altura predicha.

Esto, por supuesto, no se corresponde con la situación real, en la que árboles con el mismo dap varían en altura: eso significa que la varianza de las alturas que se predicen de los árboles de la muestra es siempre menor que las alturas de los árboles si se hubieran medido.

Los modelos estadísticos se basan en observaciones de muestras y los coeficientes de los modelos son en sí mismos estimaciones

Todos los modelos son estimaciones en sí mismos: se basan en observaciones sobre un conjunto (muestra) de árboles y no representan el modelo paramétrico (verdadero), sino sólo una aproximación (estimación) del mismo. Cuando diferentes equipos de campo toman una muestra aleatoria de 100 árboles -cada uno de la misma región de inventario, pero con diferentes aleatorizaciones- para calcular

un modelo de biomasa, cada uno utilizando el mismo modelo matemático, todos ellos obtendrán diferentes coeficientes del modelo.

Del mismo modo que ocurre con el muestreo común para los valores medios, la precisión de la estimación suele mejorar cuando el tamaño de la muestra es más grande: para cada valor individual predicho o valor medio, se pueden determinar intervalos de confianza. Para ello, por supuesto, es necesario conocer las medidas de variabilidad del modelo.

Los modelos estadísticos se caracterizan por medidas estadísticas

Al igual que con la estimación de valores medios basada en muestras, también es importante para los modelos estimados acompañar las estimaciones puntuales con estimaciones de intervalo. Las estimaciones puntuales de los modelos de regresión son los coeficientes de regresión estimados. Para cada coeficiente de regresión estimado, puede estimarse un error estándar, y cuanto más pequeño sea, más precisa será la estimación.

Una característica importante es la significancia de la regresión: si, en el ejemplo de una regresión lineal simple, el coeficiente de la pendiente no es estadísticamente distinto de cero, decimos que la regresión no es significativa. Esto significa que la línea de regresión es paralela al eje x y, en consecuencia, que el valor pronosticado para la variable objetivo, y , será el mismo para todas las variables predictoras, x . Este valor pronosticado es el valor medio de y . Así, en lugar de calcular una regresión que determine un valor medio específico por clase de valores x , se puede utilizar la media global, como valor pronosticado para todas las clases x .

Los modelos estadísticos emplean puntos de datos variables

Otra estadística importante que caracteriza la precisión de la predicción del modelo de regresión es la variabilidad de los puntos de datos utilizados para la construcción del modelo en torno a la línea de regresión. Del mismo modo que los valores de una única variable varían en torno al valor medio, los puntos de datos de un modelo de regresión varían en torno a la línea de regresión, donde la línea de regresión representa un valor medio cambiante que **adopta valores diferentes** para las distintas clases x . Llamamos a esta variabilidad la varianza residual. Cuando esta varianza de los residuos es pequeña, los puntos de datos se agrupan estrechamente alrededor de la línea de regresión, y podemos suponer que las predicciones del modelo son bastante precisas. Recuerde que no sólo es importante conocer los

coeficientes de regresión para poder calcular los valores predichos, sino también conocer las estadísticas de varianza del modelo para poder evaluar su calidad.

Recuerde que no sólo es importante conocer los coeficientes de regresión para poder calcular los valores predichos, sino también conocer las estadísticas de varianza del modelo para poder evaluar su calidad. La precisión de la estimación a partir de un modelo de regresión lineal simple, por ejemplo, es máxima en torno al valor medio de las variables predictoras y se reduce hacia los extremos del intervalo de valores predictores incluidos; más allá de este intervalo, el modelo no debe utilizarse y, en tal caso, la precisión será baja.

Los modelos estadísticos son válidos para "condiciones específicas"

Los datos de base utilizados para construir un modelo definen la validez del modelo para un propósito de inventario específico. Cuando nos referimos a los datos de base, nos referimos a factores como:

Región geográfica	En el mejor de los casos, los datos de base proceden de la región geográfica donde se realiza el inventario. Si no es el caso, hay que asegurarse de que el modelo es adecuado (véase la sección sobre Identificación de modelos estadísticos adecuados en esta lección).
Especies arbóreas	Algunos modelos son específicos para una especie o para un grupo de especies; habría que comprobar si podrían ser aplicados a otras especies.
Rango de las variables de entrada	En general, los modelos sólo deberían utilizarse para valores de variables de entrada que estén cubiertos por los datos de base. Puede ser un riesgo, por ejemplo, utilizar un modelo de biomasa que se ha construido para valores de entre 30 y 150 cm también para árboles más pequeños por debajo de 30 cm; puede ser que tal extrapolación produzca valores poco plausibles, ya que la función del modelo no está esencialmente definida fuera del rango de los datos de base.

Cómo construir su propio modelo de biomasa

Construir su propio modelo de biomasa no suele ser tarea de un proyecto del IFN. Se puede recurrir a modelos ya publicados. A veces, estos modelos se construyeron en el marco de investigaciones

académicas o informes técnicos y son difíciles de encontrar. Pero es probable que existan modelos adecuados para todas las situaciones.

La FAO, junto con el CIRAD, ofrece una base de datos para alojar modelos de biomasa en la [iniciativa GlobAllomeTree](#) (en inglés), que puede servir de referencia a la hora de buscar o construir modelos de biomasa. Consultar un manual detallado sobre el desarrollo de ecuaciones alométricas es una buena idea para los interesados (Picard *et al.*, 2012).

En esta sección, describimos brevemente los pasos que hay que seguir para construir su propio modelo de biomasa. El proceso es igualmente válido para cualquier otro modelo estadístico. Por ejemplo, cuando se desea estimar la biomasa de árboles eliminados ilegalmente a partir de un inventario de tocones, se aplicarían modelos normales de biomasa, y para ello es necesario predecir el dap a partir del diámetro del tocón, y es posible que desee construir aquí su modelo específico.



Nota

Existen varios modelos publicados para diferentes especies arbóreas que permiten predecir el dap a partir del diámetro del tocón. Se trata de factores simples (por ejemplo, Bones, 1960) o de modelos de regresión, algunos de los cuales incluyen también la altura del tocón (por ejemplo, Pond y Froese, 2014).

Cabe señalar que en estas publicaciones no encontramos ninguna definición del diámetro del tocón ni de la altura del tocón, ni indicaciones sobre cómo medirlos. Sin embargo, el diámetro del tocón puede ser muy irregular (ya que a menudo los árboles son contrafuertes a alturas muy bajas) y requeriría definiciones claras e inequívocas.

Esto ilustra la importancia de las definiciones inequívocas, no sólo en el monitoreo forestal, sino también a la hora de referirse a las variables de entrada de los modelos estadísticos.

Paso 1: Empezar con las definiciones

Siguiendo las buenas prácticas de monitoreo forestal que recomiendan tener una terminología, definiciones y mediciones claras y transparentemente documentadas, se debe empezar por las

definiciones. Las definiciones se extienden a la población de la que se van a tomar los árboles de muestra, incluyendo:

- la definición geográfica del área de procedencia de los árboles de muestra;
- una definición de especie o grupo de especies; y
- una definición del rango de dimensiones (normalmente el rango del dap) para el que se mantendrá el modelo.

Además, es necesario definir la biomasa en términos de compartimentos de biomasa (tronco, ramas grandes, ramitas más pequeñas, subterránea, hojas, etc.), así como los diámetros mínimos.

Paso 2: Determinar el número de árboles de muestra que se talarán

Hay que determinar el número de árboles de muestra que se van a talar, lo que suele hacerse en función de los recursos disponibles. Dado que talar y pesar los árboles es costoso, el número de árboles suele ser bastante limitado, aunque es bueno trabajar con un mayor número de árboles de muestra para que los modelos predigan con mayor precisión.

Hay una gran diferencia entre muestrear un árbol grande o un árbol pequeño, ya que la tala, el troceado y el pesaje de los árboles más grandes serán desproporcionados y mucho más caros. Esto lleva a una situación en la que la mayoría de los estudios de biomasa tienen muchos árboles pequeños y sólo unos pocos árboles grandes.

Se trata de una especie de dilema, porque la variabilidad de la biomasa de los árboles más pequeños es mucho menor que la variabilidad de los árboles más grandes, y la conclusión sería que necesitamos árboles más grandes para tener una mejor base para un modelo en el que la variabilidad es mayor (y en el que con frecuencia la mayor parte de la biomasa de la masa forestal está en los árboles grandes).

Paso 3: Seleccionar los árboles de muestra

Aunque la forma general de los modelos de biomasa es bastante conocida y sigue algunas leyes físicas, el objetivo es intentar determinar la forma de la función de biomasa en todo el rango de valores de entrada (normalmente un rango del dap). Esto significa que se deben seleccionar árboles de muestra para todas las clases diamétricas y (teóricamente) más árboles de muestra donde la variabilidad de la variable objetivo (biomasa) sea mayor.



Comprobación de la realidad

La selección de los árboles de muestra suele estar dominada por consideraciones prácticas como la accesibilidad, los permisos de tala y otras. Se trata de otro ejemplo en el contexto del monitoreo forestal en el que la teoría se enfrenta a la práctica, en otras palabras, en el que la ciencia se encuentra con la vida real.

Paso 4: Medir los árboles de muestra en pie

Primero hay que medir los árboles de muestra en pie, como cuando se miden las variables predictoras en un inventario forestal por defecto. Por ejemplo, la altura de los árboles difícilmente pueda determinarse después de la tala, porque entonces lo que medimos es la longitud del árbol. Además, el dap se mide mejor en el árbol en pie porque después de la tala es más difícil determinar la altura del pecho.

Paso 5: Talar los árboles de muestra

Una vez que se han medido los árboles de muestra en pie, hay que talarlos con cuidado, porque es necesario asegurarse de que todas las partes de biomasa pertinentes pueden atribuirse a esa muestra. A continuación, hay que separar y pesar los compartimentos de biomasa pertinentes.



Comprobación de la realidad

Durante la medición y la tala de los árboles de muestra, es habitual que los errores de medición no se tengan en cuenta ni se incluyan en el modelo a la hora de determinar las medidas de incertidumbre. Las mediciones, tanto en los árboles de muestra en pie como en los talados, se toman simplemente como valores verdaderos. Sin embargo, sabemos que no es así y que los errores de medición pueden influir, sobre todo cuando el número de árboles de muestra es bajo. En los IFN, normalmente tenemos muestras de gran tamaño y número de árboles de muestra registrados; entonces, podemos asumir que los errores aleatorios de medición tienen un peso relativamente pequeño ya que el gran número de observaciones mantendrá baja la varianza del error.

Paso 6: Gestionar los datos de los árboles de muestra

Los datos de los árboles de muestras deberán gestionarse en una base de datos en la que se almacenen todos los resultados de los compartimentos y, finalmente, se sumen a los valores objetivo de cada árbol de muestras en particular. Al final, y como aporte para análisis posteriores, se elabora una lista con datos por árbol de la variable objetivo y las variables predictoras. Dicha lista es la matriz de entrada para la estimación de los coeficientes del modelo.

Paso 7: Identificar un modelo matemático que se ajuste bien al conjunto de datos

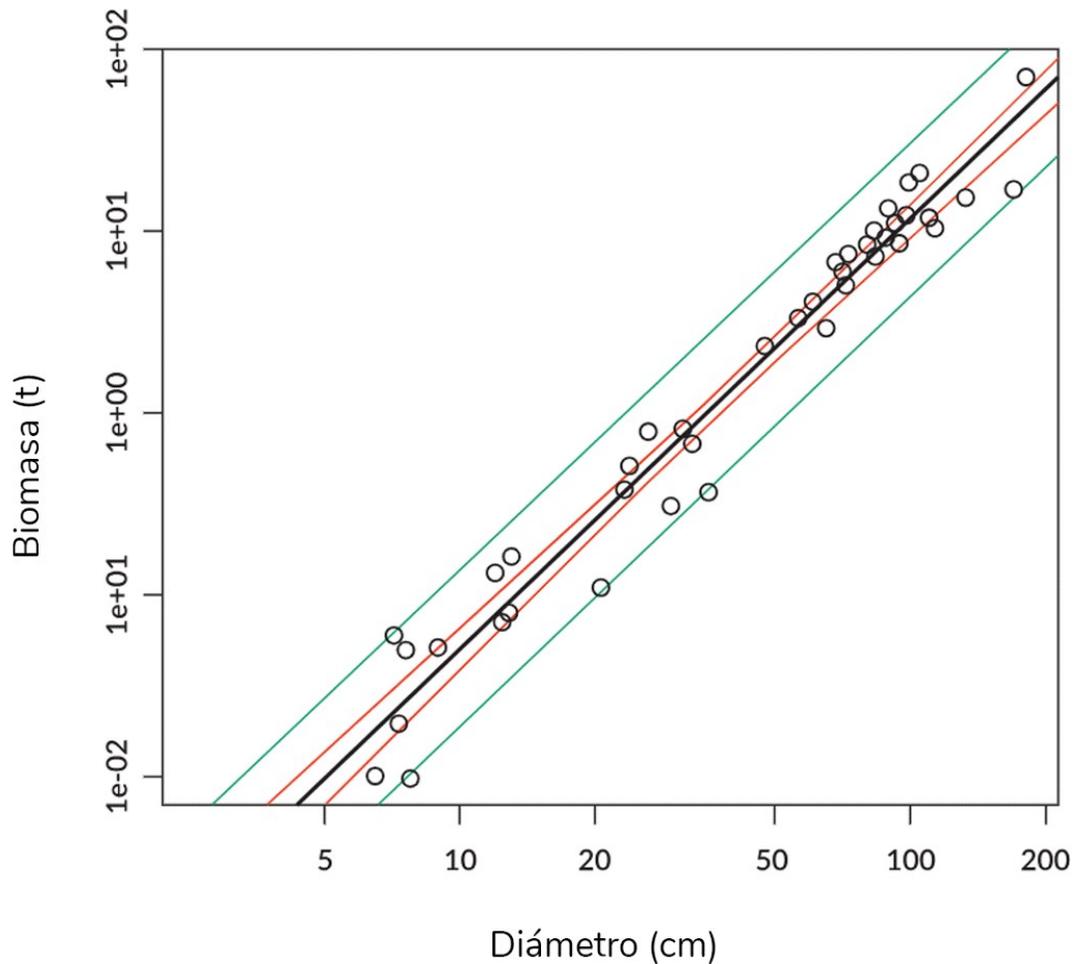
Los siguientes pasos son de estadística aplicada: hay que identificar un modelo matemático que se ajuste bien al conjunto de datos. Para los modelos de biomasa, como para otros modelos, se conocen modelos matemáticos típicos. Es habitual comparar el desempeño de diferentes modelos matemáticos y elegir el que permita las predicciones más precisas.

A common approach is to create a model using a random selection of 75 percent of the sample trees, and then evaluate the model's performance using the remaining 25 percent of the sample trees. Here, we consider the 25 percent check-trees as independently selected trees for model validation. En este caso, consideramos los árboles de control del 25 % como árboles seleccionados de forma independiente para la validación del modelo. Es importante distinguir entre:

1. la incertidumbre del modelo, que resulta del análisis del 75 % de los árboles de muestra utilizados para construir el modelo; no hace falta decir que el modelo se ajusta bastante bien a este conjunto de datos porque este conjunto de datos es la base del modelo; y
2. la incertidumbre de la predicción, que suele ser mayor y se refiere a la predicción de los árboles que no se han utilizado para construir el modelo. Considere aquí que los árboles de muestra utilizados para construir el modelo son una muestra de la población de interés y, por supuesto, nuestra muestra no podrá capturar toda la variabilidad presente en la población, pero es sólo una estimación, como se muestra en la siguiente figura.

Entonces, todos los árboles registrados en un inventario pertenecen a este conjunto de árboles que no se habían utilizado para construir el modelo. De ahí que la incertidumbre en la predicción sea un punto importante en el desarrollo de modelos. Picard *et al.* (2012) proporcionan información sobre estas

cuestiones.



Una vez que esté conforme con la incertidumbre del modelo y la incertidumbre de la predicción, puede volver a todo el conjunto de datos de árboles de muestra y estimar los coeficientes finales del modelo a partir del 100 % de los árboles de muestra.



Nota

Es necesario tener en cuenta una serie de cuestiones estadísticas al construir el modelo estadístico, por lo que tiene sentido consultar con un modelador estadístico a la hora de construir y elegir un modelo específico. Las cuestiones estadísticas incluyen:

1. la correlación entre las variables predictoras (la llamada colinealidad); y
2. una característica típica de los modelos de biomasa (y de volumen y carbono): la

variabilidad de la biomasa que es pequeña para los árboles más pequeños y grande para los árboles más grandes (la llamada heterocedasticidad).

Esto último afecta a la estimación de las varianzas y los intervalos de confianza para las predicciones de los valores medios y los valores individuales; por ejemplo, para los modelos de biomasa en función únicamente del dap, los intervalos de confianza serán estrechos para los árboles más pequeños y se ampliarán a medida que aumente el dap.

Paso 8: Documentar el modelo

Una vez decidido el modelo final, el último paso consiste en elaborar un informe completo y transparente del modelo; no sólo hay que documentar el modelo y sus coeficientes, sino también las características de la incertidumbre, incluido el coeficiente de determinación, el error estándar de la regresión y la incertidumbre de la predicción, así como otras posibles incertidumbres.

Puede que se desee poner a disposición del público el conjunto de datos original, porque esos conjuntos de datos pueden ser útiles combinados con otros nuevos para generar modelos más precisos, más adaptados al lugar o más generalizables.

Identificar los modelos estadísticos adecuados

Existen muchos modelos estadísticos disponibles para el monitoreo forestal. El IPCC, por ejemplo, ofrece una larga lista de diferentes factores de conversión y funciones de biomasa (véanse, por ejemplo, las Publicaciones de las Directrices del *IPCC de 2006 - IPCC-TFI* (iges.or.jp) o las publicaciones de sus valores perfeccionados de *2019 - IPCC-TFI* (iges.or.jp)) (en inglés).

En muchos casos, está claro desde el principio qué modelo utilizar, porque se ha utilizado antes en el mismo contexto geográfico o temático, o se sabe que funciona bien en las circunstancias del proyecto de inventario en particular. En los IFN que se extienden por superficies extensas e incluyen muchas especies y grupos de especies, puede ser adecuado aplicar diferentes modelos en función del grupo de especies, y/o de la región geográfica y/o de las condiciones del lugar.

Un primer paso en la selección del modelo consiste en comprobar qué modelos se han utilizado anteriormente, la procedencia de los datos de árboles de muestra utilizados para la construcción del modelo y evaluar las estadísticas de incertidumbre de los modelos; por lo general, cuantos más árboles

de muestra se hayan procesado, más preciso será el modelo. A veces hay que decidir, por ejemplo, si se utilizan varios modelos de biomasa específicos para cada especie o un modelo general para todas las especies. En los IFN en los que el tamaño de las muestras suele ser grande, la recomendación de investigaciones recientes es fijarse en el número de árboles de muestra utilizados para la construcción del modelo más que en las especies específicas para las que se había construido.

Se prefiere un modelo general para todas las especies basado en árboles de muestra en vez de varios modelos específicos para cada especie construidos a partir de pocos árboles cada uno.



Consejos prácticos

Se han identificado algunas reglas empíricas en relación con el desarrollo de modelos de biomasa (McRoberts y Westfall, 2014): el número de árboles de muestra debe ser de al menos 100, y los modelos deben tener un coeficiente de determinación superior a 0,95. Luego, en los IFN, los errores de los modelos en la biomasa suelen ser relativamente pequeños en comparación con el error estándar.

Sin embargo, es importante subrayar que el papel menor de los errores de modelo, en los IFN con muestras grandes, se refiere solo a errores aleatorios, mientras que los posibles errores sistemáticos y sesgos se propagarán como sesgos en las estimaciones finales.

En situaciones en las que la elección del modelo no está clara desde el principio, la tarea consiste en comprobar la idoneidad y comparar el desempeño de los diferentes modelos. Eso solo se puede hacer con un número suficientemente grande de árboles de muestra, y es costoso, sobre todo cuando se trata de modelos de biomasa, porque determinar/medir la biomasa de los árboles de muestra siempre es costoso.

La CMNUCC (2011) ofrece una guía básica y práctica sobre cómo realizar dicha prueba de idoneidad para modelos de biomasa forestal aérea. En Pérez-Cruzado *et al.* (2015) se encuentra una descripción exhaustiva y basada en el análisis de la idoneidad de los modelos.

Cuestiones relativas a la elaboración de informes en modelos estadísticos

Cuando se habla de modelos estadísticos en el monitoreo forestal, existen esencialmente dos tipos de cuestiones relativas a la elaboración de informes:

- ↳ *Informar sobre el modelo en sí mismo y permitir que el usuario potencial comprenda plenamente la procedencia y las características del modelo*

Aquí es importante documentar todos los detalles de la construcción del modelo de forma transparente y completa, como se ha mencionado antes: de dónde proceden los datos de los árboles de muestra en términos de región geográfica, el enfoque de muestreo y las posibles restricciones, cuántos árboles de muestra se utilizaron y cómo se distribuyeron en el rango de las variables predictoras. Esencialmente, se deben notificar todos los detalles de la construcción del modelo que sean necesarios para que un usuario potencial del modelo lo entienda y comprenda su trasfondo. Esto también incluye medidas estadísticas de la exactitud del modelo, aunque no sea una práctica habitual por defecto en los IFN. Sin embargo, las medidas de incertidumbre son importantes cuando un usuario potencial compara diferentes modelos; entonces el usuario puede tender a preferir el modelo más exacto.

- ↳ *Notificación de las predicciones a partir de un modelo en el contexto de la ejecución de un IFN*

Aquí, las predicciones del modelo se tratan en los proyectos de inventarios forestales como observaciones normales, la correspondiente incertidumbre del modelo no se suele comunicar, ya que se ha demostrado que su contribución al error final es menor en los estudios empíricos y teóricos. No obstante, es una buena práctica documentar qué modelos se han utilizado e indicar su fuente y las características en el informe del inventario.

Resumen

- Antes de finalizar, aquí están los puntos clave de aprendizaje de esta lección:
- Los modelos estadísticos pretenden establecer una relación cuantitativa entre una variable predicha y una o más variables predictoras.
- Los modelos estadísticos no sirven para establecer una relación causa-efecto; esto se abordaría con modelos de procesos, cuyo objetivo es incluir explícitamente las causalidades inferidas por los procesos biológicos.
- Hay diferentes tipos de modelos de complejidad variable que se utilizan en el monitoreo forestal

y, en algunos casos, es difícil percatarse de que se haya utilizado un modelo.

- En muchos casos, está claro desde el principio qué modelo utilizar, porque se ha utilizado antes en el mismo contexto geográfico o temático, o se sabe que funciona bien en las circunstancias del proyecto de inventario en particular.
- En los IFN que se extienden por superficies extensas e incluyen muchas especies y grupos de especies, puede ser adecuado aplicar diferentes modelos en función del grupo de especies, y/o de la región geográfica y/o de las condiciones del lugar.

Lección 4: Errores en el monitoreo forestal

Introducción de la lección

Esta lección profundiza en los tipos y funciones de los errores aleatorios que se producen a lo largo del proceso del IFN. También se analiza la propagación de errores, es decir, cómo las fuentes de error se propagan al error total del resultado final.

Objetivos

Al final de esta lección, usted podrá:

1. Definir el término "error" en los estudios de muestreo empírico.
2. Describir por qué las consideraciones de error son importantes en el monitoreo forestal.
3. Explicar la relación entre error y esfuerzo.
4. Comprender los tipos y las funciones de los errores en el monitoreo forestal.
5. Explicar cómo hacer frente a los errores en el monitoreo forestal.

Observaciones generales sobre los errores en los inventarios forestales

Definición del término error en los estudios empíricos

Los inventarios forestales son estudios empíricos de muestreo, en los que es más exacto referirse a los errores como variabilidad residual y no como equivocaciones. Si bien las equivocaciones se pueden evitar mediante un trabajo cuidadoso y una orientación permanente hacia la calidad, los errores no, sólo podemos intentar que sean pequeños. Los errores a los que nos referimos aquí son de carácter aleatorio y tienden a seguir la distribución de errores gaussianas, también conocida como distribución normal.

Como ya se ha mencionado, y a diferencia de los errores aleatorios, los errores sistemáticos o sesgos suelen poder evitarse, ya que se basan en calibraciones erróneas, en el uso de estimadores sesgados o en otras aplicaciones equivocadas de los enfoques para generar datos. Como los errores aleatorios son omnipresentes, los equipos de campo y otras partes que generan datos pueden trabajar con cuidado y coherencia para mantener a un nivel mínimo las fuentes de error correspondientes. Por lo general, en el muestreo estadístico, los errores sistemáticos definen la exactitud, mientras que los errores aleatorios definen la precisión.



¿Sabías que?

La precisión y la exactitud son dos términos fundamentales en el muestreo estadístico, y vienen determinados por los errores sistemáticos (sesgo) y los errores aleatorios. A menudo, en la presentación de informes se utiliza el término incertidumbre para referirse a los errores, ya que es un término menos técnico y más accesible. Pero también está menos claramente definido. Por lo tanto, cuando se utiliza el término incertidumbre en el contexto del muestreo estadístico, es una buena práctica aclarar lo que se quiere decir específicamente.

¿Por qué es tan importante considerar los errores en el monitoreo forestal?

La presencia y la magnitud de los errores son factores importantes que contribuyen en gran medida a la credibilidad de los resultados de los inventarios. Si el error es del 50 %, se tendrá menos confianza en los resultados que si el error es del 1 %. Por lo tanto, es imperativo informar de los errores de todos los resultados, cuantificado los errores que se pueden cuantificar y abordando/discutiendo los errores que no se pueden cuantificar directamente.

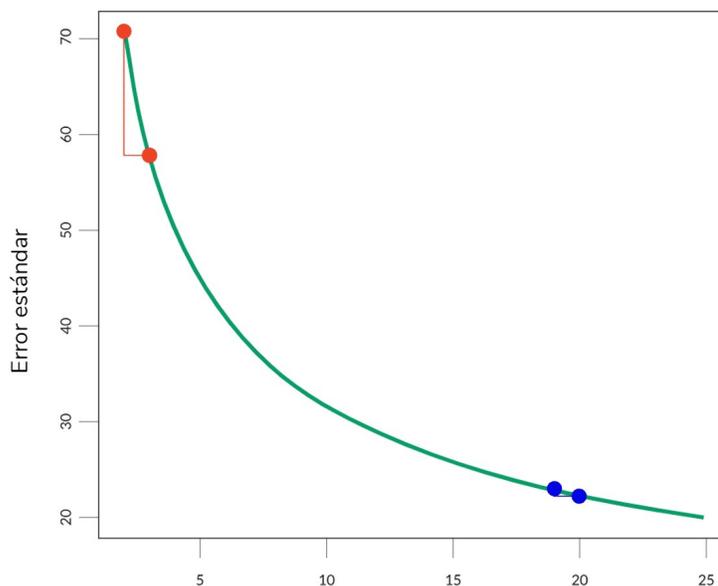
Al planificar el diseño del inventario, el objetivo es siempre utilizar los recursos disponibles para optimizar la precisión de la estimación y evitar errores sistemáticos. Por lo tanto, es importante comprender las funciones de los distintos elementos del diseño del inventario y cómo contribuyen las fuentes de error correspondientes a los errores totales. Si se dispone de recursos adicionales que se pueden utilizar para mejorar el diseño del inventario, habrá que identificar cómo asignarlos, para que aumenten la precisión = reducir la incertidumbre = reducir el error total para la(s) variable(s) objetivo principal(es).

La relación entre errores y esfuerzos

El error estándar se suele mostrar en función del esfuerzo. El esfuerzo se define como el tamaño de la muestra que se puede establecer. El aumento marginal de la precisión es menor cuanto mayor es la precisión. Esto significa que, para el mismo esfuerzo, se obtendrá un aumento menor de la precisión si se parte de un nivel inicial de precisión elevado. Aquí, el error se refiere al error estándar y el esfuerzo

se refiere al tamaño de la muestra que se puede establecer. Por lo tanto, al optimizar los diseños de los inventarios forestales para reducir los errores, puede resultar sencillo fijarse primero en las fuentes de errores relativamente grandes, aunque su impacto en el error final no parezca el más importante a primera vista.

Por poner un ejemplo: en la figura a continuación, suponemos un MAS y una desviación estándar estimada de $s=100$. El aumento del tamaño de la muestra en 1, de $n=2$ a $n=3$, reduce el error estándar estimado de $SE=70,7$ a $SE=57,7$, en torno a un 18 %. Si invertimos los mismos recursos adicionales en términos absolutos y aumentamos el tamaño de la muestra en 1 de $n=19$ a $n=20$, el aumento de la precisión = reducción del error estándar de $SE=22,9$ a $SE=22,4$, lo que supone una reducción relativa de tan sólo un 2,2 %.



La relación representada en esta figura es válida para el error estándar en los diseños del MAS. Sin embargo, se puede suponer que existen relaciones similares para otras fuentes de error, por ejemplo, aumentar los esfuerzos de formación en una cantidad absoluta fija para obtener mejores observaciones de las variables tendrá por lo general un mayor efecto para aquellas variables cuyo error de medición sea relativamente grande (como la medición de la altura de los árboles), y no tendrá grandes efectos cuando se trate de variables cuyo error de medición ya sea bastante pequeño (como la medición del dap).

¿Qué es una buena precisión en los IFN y a qué nivel de error se debe apuntar?

No existe ninguna regla válida en general sobre la precisión objetivo en los IFN. El enfoque más común

consiste en definir los recursos (presupuesto) y diseñar el inventario de forma que se logre la mejor precisión dentro de las restricciones de estos recursos. La precisión alcanzada depende del diseño del muestreo y, en particular, del tamaño de la muestra.

En algunos IFN, el tamaño de la muestra es del orden de magnitud de los 10 000, de modo que la precisión de la estimación para todo el país es elevada -en algunos casos, el error estándar relativo es inferior al 1 %- pero en otros, el tamaño de la muestra puede ser dos órdenes de magnitud inferior, produciendo errores estándar relativos superiores al 10 %. Sin embargo, cuando se producen estimaciones para subpoblaciones más pequeñas, en las que el tamaño de la muestra es mucho menor, el error estándar puede aumentar y alcanzar niveles para los que la fiabilidad de los resultados puede verse comprometida.

Tipos de errores en el monitoreo forestal y su función

Los sistemas de monitoreo forestal son complejos y en su ejecución participan muchas personas. Esto tiene el potencial de aumentar las diversas fuentes de errores que hay que observar a la hora de planificar y ejecutar estos sistemas. Existen tres tipos de errores que se producen en el monitoreo forestal y todos ellos desempeñan una función importante, aunque tienen una relevancia variable

- ① error de medición;
- ② error de modelo, y
- ③ error de muestreo.

En esta sección profundizaremos en lo que constituye cada uno de ellos.

Error de medición

Cuando se realiza una observación, esta tiene una variabilidad residual. Cuando, por ejemplo, el dap se mide con una forcípula de escala muy fina, digamos con el 5º decimal de un milímetro, las mediciones repetidas -realizadas todas ellas de forma muy diligente y correcta- arrojarán prácticamente todas diferentes mediciones. La variabilidad de estas mediciones indica la existencia de errores de medición.

Estos errores de medición pueden producirse más allá de las variables cuantitativas. Las variables categóricas que se observan en, digamos, 10 clases diferentes, pueden presentar asignaciones incorrectas: dicho error de medición se denominaría entonces también error de clasificación. O, al

observar variables nominales, como las especies arbóreas, pueden producirse confusiones/identificaciones erróneas, también consideradas errores de medición. Es importante darse cuenta de que se producirán errores de medición en la observación de cualquier variable.

Un caso interesante es la medición de la altura de los árboles empleando el principio trigonométrico. De hecho, la altura no se mide, sino que se calcula a partir de tres mediciones: la distancia horizontal al árbol y las medidas angulares a la parte superior e inferior del árbol. Dependiendo del dispositivo de medición, la distancia horizontal también puede basarse en dos mediciones: la distancia de la pendiente y el ángulo de la pendiente. Todas estas mediciones individuales conllevan sus errores de medición específicos y el error en la altura procede de los errores propagados de todas estas mediciones individuales.



Consejos prácticos

En el monitoreo forestal, no suele haber mucha información disponible sobre los errores de medición, y las mediciones se utilizan en los cálculos como si estuvieran libres de errores. Para muchos árboles de muestra, como suele ser en los IFN, se puede justificar ignorar los errores aleatorios de medición y no informarlos explícitamente, ya que serán pequeños comparados con el error de muestreo.

Si fuera de interés, se pueden establecer pequeños estudios de investigación en los que diferentes equipos de campo realicen todas las observaciones en parcelas de muestreo. La varianza de las mediciones de las distintas variables se puede considerar entonces como una aproximación de los errores de medición/observación; esto se puede extender a las mediciones del dap y la altura, a la identificación de las especies arbóreas y al número de árboles muestra existentes en cada parcela.

Error de modelo

Los modelos se usan con frecuencia en el monitoreo forestal para establecer relaciones que permitan predecir variables que no pueden medirse o que resultan laboriosas de medir (véase la Lección 3 de este curso). Lo que se lee de un modelo no es, por supuesto, el valor verdadero del objeto en cuestión.

Cuando, por ejemplo, el volumen del tronco se lee a partir de un modelo de volumen como una función

del dap, entonces lo que se lee como el volumen del tronco del árbol es en realidad el volumen del tronco de todos los árboles en esta clase de dap - y el verdadero volumen del árbol en particular se desviará de eso. Esto es lo que se denomina error de modelo.

En el monitoreo forestal, los errores del modelo no se suelen comunicar ni se tienen en cuenta en las estimaciones de intervalo. Las predicciones que se hacen a partir de los modelos se toman como valores verdaderos, o libres de errores. En los modelos para variables arbóreas como los de biomasa, volumen o altura, si los modelos utilizados se basan en un número relativamente grande de árboles de muestra, y si el número de árboles de muestra en el inventario es grande, se justifica no informar de los errores del modelo, ya que serán mínimos comparados con el error de muestreo.

Error de muestreo

El error de muestreo tiene su origen en el hecho de que no estamos observando todos los elementos de una población, sino sólo una muestra. Por lo tanto, todos los resultados son estimaciones y se desvían del valor real de una forma que se describe mediante el error estándar.

En el monitoreo forestal, el error estándar se estima por lo general de forma no sesgada para la mayoría de los diseños de muestreo, excepto para el muestreo sistemático, en el que tenemos que recurrir a aproximaciones o a estimaciones conservadoras de la varianza del error/error estándar al aplicar el marco del MAS.

La función de estos errores

Los tres tipos de errores descritos anteriormente existen en todos los estudios de monitoreo forestal. En la última década se ha investigado mucho, sobre todo en el contexto de la estimación de la biomasa forestal, para averiguar cuáles son las contribuciones relativas de estos errores a la estimación final de la biomasa. Enfocamos en el monitoreo forestal nacional, y entre las principales características de los IFN está una muestra grande y, por tanto, un número muy elevado de árboles muestra.

Un hallazgo muy relevante ya se publicó en uno de los primeros artículos sobre la propagación de las fuentes de error al error final: Gertner y Köhl (1992) acuñaron el término "**presupuesto de error**" y descubrieron que en los IFN suizos el mayor peso, por mucho, lo tiene el error de muestreo, con cerca del 98 % del error total. Los errores de modelo y los errores de medición sólo representan el pequeño porcentaje restante. Se trata de una conclusión importante, ya que en los inventarios forestales sólo se suele informar del error estándar inducido por el muestreo, mientras que las demás fuentes no suelen

cuantificarse ni notificarse.

Esta conclusión es válida para los IFN con muestras de gran tamaño. En estudios de inventarios más pequeños con tamaños de muestra más reducidos, el peso relativo de los errores de medición y de modelo puede ser considerablemente mayor.

Cómo hacer frente a los errores en el monitoreo forestal

Aunque podemos evitar los errores sistemáticos mediante una planificación y ejecución cuidadosas, se siguen produciendo errores aleatorios. Por lo tanto, el objetivo es que estos errores aleatorios (= la variabilidad residual) sean pequeños. Uno de los principales objetivos del monitoreo forestal es producir estimaciones con errores razonablemente pequeños. Decimos "razonablemente pequeños" porque cualquier aumento de la precisión costará recursos y es especialmente costoso cuando la precisión ya es relativamente alta. En los IFN, el error de muestreo es el componente de varianza del error más importante y la elección de un diseño de muestreo adecuado y, a continuación, de un tamaño de muestra apropiado son los recursos disponibles para ajustar el **error estándar** (es decir, el **error de muestreo**).

El trabajo cuidadoso, la formación y el control periódico pueden ayudar a reducir los **errores de medición**. En este caso, lo más importante es evitar errores sistemáticos, por calibración incorrecta, y mantener motivados a los equipos de campo para mantener el afán de producir buenos datos; los extensos periodos sobre el terreno son agotadores y pueden empeorar la motivación y la calidad de los datos.

Sin embargo, en cuanto a los **errores de modelo**, todo depende del rigor de la relación estadística entre la variable objetivo y los predictores, así como de la elección del modelo y de las características de calidad de este, que están codeterminadas por el número de observaciones que subyacen al modelo. En general, cuantos más datos se utilicen para construir un modelo, más fiable podrá considerarse.

Los proyectos de inventario forestal y los programas de monitoreo forestal son complejos. Los tres tipos de errores abordados anteriormente se pueden producir en cualquiera de las etapas de estos sistemas, y el principal interés reside finalmente en el error total en las variables objetivo. Intuitivamente, los errores de las distintas fases del proceso afectarán el error total de la variable objetivo, es decir, su impacto se propaga por las distintas fases del proceso del inventario a la variable objetivo final.

En este contexto de propagación de errores, hay que tener en cuenta un par de puntos:

1. **Cómo funciona el mecanismo de propagación de errores y cómo se determina el error total.** Esto es importante para reportar el error total como indicación de la precisión y la fiabilidad general de los resultados.
2. **En qué medida los distintos errores contribuyen al error total.** Esto es importante en el contexto de la optimización del diseño del inventario para inventarios continuos: uno se esforzará por reducir aquellos errores que tengan el mayor impacto en el error de las variables objetivo (a un costo aceptable).

Principios básicos de la propagación de errores

Los principios básicos de la propagación de errores son sencillos y dependen de cómo se relacionen las distintas variables de entrada y sus errores.

Una introducción buena y comprensible a la propagación de errores es Taylor (1997), que aborda las reglas para combinar variables aleatorias según la operación utilizada para combinarlas. Se pueden encontrar conferencias introductorias más breves que incluyen sumas, productos, razones y otras operaciones en los siguientes lugares (en inglés):

[*Guide to Uncertainty Propagation and Error Analysis: Stony Brook Introductory Physics Labs*](#)

[*Propagation of Uncertainty through Mathematical Operations*](#)

Lo que se describe en Taylor (1997) es un enfoque analítico de la propagación de errores que puede aplicarse directamente a funciones de variables aleatorias. Sin embargo, si se debe realizar una propagación de error en un diseño de inventario complejo, en el que se deben considerar muchas fuentes de error diferentes, dicha propagación de error analítica se vuelve extremadamente difícil. Entonces, un estudio de simulación (también llamado Simulación Monte- Carlo), podría ser más apropiado.

Para realizar un estudio de este tipo, se debe disponer de información sobre los distintos componentes del error. Por lo general, se asumen distribuciones normales de estos errores. A continuación, se calcula la variable objetivo (por ejemplo, la biomasa) a partir de todos los datos de entrada, y para cada estimación puntual se determina una desviación aleatoria de su error distribuido normalmente. Esta

operación se repite muy a menudo, unas 10 000 veces, y la varianza de los valores finales resultantes de la variable objetivo es la varianza total del error propagado. Se pueden encontrar ejemplos instructivos en Molto *et al.* (2013), McRoberts y Westfall (2016) y Lin *et al.* (2023).

Tanto la simulación como el cálculo analítico de la propagación del error permiten evaluar el peso de los distintos componentes del error en el error de estimación final, por lo que estos ejercicios son instructivos para optimizar el diseño del inventario.

McRoberts y Westfall (2016) ofrecen un ejemplo instructivo de cómo se puede realizar un estudio de simulación cuando lo que interesa es propagar diversas fuentes de error en el monitoreo forestal a la variable objetivo final. Si la estimación de la biomasa es la variable objetivo, los dos autores integraron las siguientes fuentes de error en su simulación:

- Si la estimación de la biomasa es la variable objetivo

Los dos autores integraron las siguientes fuentes de error en su simulación: Variabilidad de las estimaciones de los parámetros del modelo, (β), en el modelo alométrico; Variabilidad de las mediciones del dap; Variabilidad de las mediciones de la altura (y otras variables de entrada al modelo); Variabilidad residual (lo que se predice a partir del modelo no es la biomasa real); Agregación de la biomasa de árboles individuales a la biomasa de la parcela; y Estimación de la biomasa total para el área de estudio a partir de una muestra de n parcelas (con un diseño de muestreo definido).

- A continuación, la simulación se realizó del siguiente modo:

Para cada simulación, se determina un conjunto de valores de entrada para las variables mencionadas, en el que se añaden a la estimación puntual componentes de error elegidos aleatoriamente entre errores distribuidos normalmente. Luego, se determina la biomasa total para este ajuste específico. La simulación se repite (m) veces, cada una con valores de entrada que se determinan a partir de las estimaciones puntuales más un componente de error aleatorio. La variabilidad de los m valores de biomasa total resultantes es entonces una aproximación empírica del error total propagado.

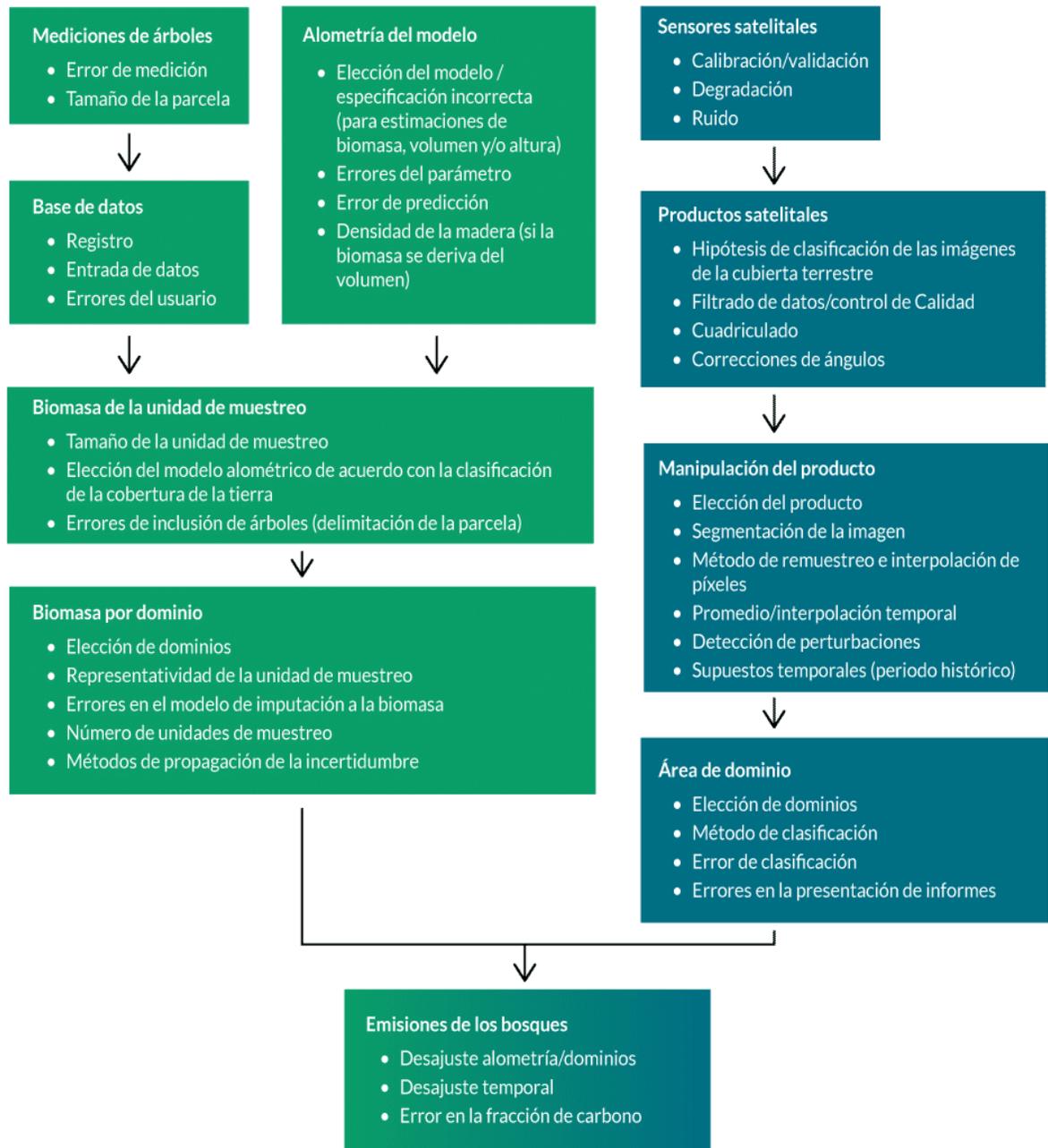
Algunos comentarios finales sobre la propagación de errores

Los países presentan una lista exhaustiva de las posibles fuentes de error que suelen surgir en la cadena de procesamiento para calcular los factores de emisión, los datos de actividad y el carbono total. La lista contiene posibles conceptos demasiado engorrosos y fuera del alcance de estos cursos.

Tenga en cuenta que las rutas verdes se definen exclusivamente mediante datos del inventario. Las

rutas azules se siguen del procesamiento de datos satelitales. El recuadro azul-verde procede de la combinación de factores de emisión basados en inventarios y estimaciones de datos de actividad obtenidos por satélite.

Flujo de errores en los factores de emisión basados en inventarios (en verde) y en los datos de actividad obtenidos por satélite (en azul)



Resumen

Antes de finalizar, aquí están los puntos clave de aprendizaje de esta lección:

- Los inventarios forestales son estudios empíricos de muestreo, y cuando hablamos de errores en los estudios de muestreo empírico nos referimos a la variabilidad residual y no a equivocaciones.
- La presencia y la magnitud de los errores son factores importantes que contribuyen en gran medida a la credibilidad de los resultados de los inventarios.
- No existe ninguna regla válida en general sobre la precisión objetivo en los IFN - la precisión se decide por la combinación más eficaz de recursos (presupuesto) y el diseño del inventario.
- Existen tres tipos de errores que se producen en el monitoreo forestal y todos desempeñan un papel importante, pero tienen una relevancia variable en función del diseño del inventario: error de medición, error de modelo y error de muestreo.

Lección 5: Productos característicos del análisis de datos

El principal objetivo del análisis de datos en los IFN es transformar los datos de los IFN en información significativa para las partes interesadas.

Objetivos

Al final de esta lección, usted podrá:

1. Describir los posibles productos de los análisis de datos de los IFN.
2. Identificar la función de los análisis de datos para la elaboración de informes de los inventarios forestales.

Productos de los análisis de datos del monitoreo forestal: Observaciones generales

El análisis de datos genera resultados de los IFN que finalmente cumplen con las expectativas establecidas en la evaluación de las necesidades de información. En esta sección, abordaremos brevemente los tipos generales de productos que se espera obtener a partir de los datos de los IFN.

Tener una idea clara de los potenciales productos del análisis de datos de los IFN en la fase de planificación y durante la ENI siempre ayuda. También es instructivo -en la fase de la ENI- presentar todos los posibles resultados de los análisis de los datos de los IFN y reducirlos a lo que puede esperarse de forma realista.



Consejos prácticos

Durante la fase de evaluación de las necesidades de información no es necesario considerar todas las implicaciones del análisis, ya que esto es tarea de los analistas de datos. Sin embargo, es sin duda útil contar con la presencia de expertos con experiencia en el análisis de datos de los IFN en la evaluación de las necesidades de información para evitar expectativas completamente irreales. Un ejemplo típico es la expectativa de que los datos de los IFN se pueden utilizar directamente para fines de planificación forestal a nivel de distrito o incluso de rodal; en este caso, un experto en inventarios con experiencia deberá aclarar las posibilidades y limitaciones de los conjuntos de datos de los IFN.

Tipos de productos

Teniendo en cuenta la amplia gama de resultados y productos que se pueden generar, los IFN producen conjuntos de datos exhaustivos que incluyen diversas opciones de análisis. En general, es una buena práctica presentar los resultados utilizando dos estrategias distintas:

1. una para las partes interesadas y los responsables de la toma de decisiones (que debe ser técnico-científica y seguir los requisitos expresados en la ENI); y
2. una para el público en general (que debe resumir las principales conclusiones en un lenguaje accesible, aunque preciso).

Por lo que respecta a los productos técnico-científicos, podemos desglosar la información en las cuatro siguientes categorías:

1. estadística estándar;
2. mapas;
3. optimización del diseño del IFN; y
4. uso en el ámbito académico.

Estadística estándar

Los resultados del análisis de datos de los IFN no pueden abarcar la lista completa de resultados estándar. Por lo tanto, es mejor limitarse a los que se producen comúnmente en los IFN, junto con los productos adicionales específicos que surgen de las necesidades particulares de información para un país específico.

Por ejemplo, en un país con una cubierta forestal baja, puede ser extremadamente relevante producir también resultados sobre los árboles fuera del bosque (TOF), mientras que, en los países con una cubierta forestal alta, este recurso arbóreo puede tener una relevancia menor (por supuesto, el análisis de los datos sobre los TOF sólo es posible cuando la evaluación de los TOF está integrada en el diseño del inventario). Las unidades básicas de referencia para los análisis suelen ser el país entero y las unidades subnacionales -provincias, estados o zonas ecológicas definidas. En la mayoría de los casos, el tamaño de la muestra de los IFN no permite ir más abajo y producir estimaciones para unidades más pequeñas, a menos que se utilicen técnicas especiales de análisis científico, como las estimaciones de

áreas pequeñas que son usadas como enfoque avanzado.



¿Sabías que?

¿Qué es la estimación de áreas pequeñas?

Los IFN suelen utilizar muestras sistemáticas con un tamaño de cuadrícula del orden de kilómetros. Por ello, no se pueden producir estimaciones razonables para unidades relativamente pequeñas (por ejemplo, unos pocos km²) porque el tamaño de la muestra es demasiado pequeño.

Sin embargo, se está investigando cómo aprovechar la información de grandes áreas producida por el muestreo de IFN y utilizarla para producir resultados también para unidades geográficas más pequeñas. Esto se puede hacer vinculando las observaciones de campo a datos de teledetección de cobertura total, que luego se utilizan como datos auxiliares para establecer modelos que permitan predecir las variables objetivo para cualquier píxel, es decir: sobre toda el área de inventario. Entonces, los datos de las variables objetivo están disponibles no sólo en los puntos de campo observados, sino para cualquier ubicación (píxel) en la región del inventario.

Este enfoque para producir resultados a partir de la muestra de campo gruesa de área grande para cualquier unidad de área pequeña dentro de la región del inventario se denomina **estimación de área pequeña**.

Por supuesto, la incertidumbre de la predicción para las áreas pequeñas depende exclusivamente de la calidad del modelo que se elabore y utilice, vinculando las observaciones de campo de área grande y los datos de teledetección; y esto depende, entre otros factores, de la resolución espacial y espectral de los datos de teledetección, del diseño de las parcelas del inventario de campo y de la calidad del co-registro de los datos de campo y de teledetección.

Las **estimaciones de áreas** incluyen: área de bosque, área de tipos de bosque, áreas o porcentaje de área para especies arbóreas concretas, áreas por tipos de gestión, de estado de protección, de estado de degradación, de propiedad o por características topográficas. A continuación, se producen resultados para cada una de estas unidades de información, por ejemplo, área de bosque por unidad subnacional, áreas de tipos de bosque por país y por unidad subnacional, área de bosque en diferentes elevaciones,

etc. La presentación suele ser en forma de tablas de contingencia, como por ejemplo el área de, digamos, 5 tipos de bosque dentro de 10 unidades subnacionales.

Tierras	Medida	Especificación del bosque				
		Tierras forestales con existencias	Temporalmente sin existencias	Tierra forestal	Tierras forestales sin existencias	Bosque
Baden-Württemberg	[ha]	330 625	1 301	1 331	39 922	1 371 847
	Prim	4 600	13	9264 601	372	4 620
	SE [%]	1.2	27.7	1.2	5.2	1.2
Bayern	[ha]	2 534 232	3 796	2 538 028	67 535	2 605 563
	Prim	2 795	11	2 797	194	2 815
	SE [%]	1.6	33.7	1.6	7.7	1.6
Brandenburg+Berlin	[ha]	1 096 101	2 369	1 098 470	32 378	1 130 847
	Prim	907	6	907	67	909
	SE [%]	2.7	40.8	2.7	12.8	2.7
Hessen	[ha]	845 792	7 598	853 390		
	Prim	706	19	706	40 790	894 180
	SE [%]	2.9	22.8	2.9	91	715
Mecklenburg-Vorpommern	[ha]	538 651	2 186	540 836	17 286	558 123
	Prim S	2 038	19	2 041	148	2 055
	E [%]	2.1	24.0	2.1	8.8	2.0
Niedersachsen	[ha]	1 158 459	2 985	1 161 444	43 147	1 204 591
	Prim	1 552	12	1 555	135	1 571
	SE [%]	2.4	30.5	2.4	9.2	2.4
Nordrhein-Westfalen	[ha]	880 082	3 997	884 059	25 452	909 511
	Prim	861	10	863	59	867
	SE [%]	3.1	31.6	3.1	13.3	3.1
Rheinland-Pfalz	[ha]	812 818	2 290	815 108	24 688	839 796
	Prim	2 828	22	2 831	236	2 848
	SE [%]	1.5	21.7	1.5	6.5	1.4
Saarland	[ha]	101 459	783	102 242	392	102 634
	Prim	100	2	100	1	100
	SE [%]	8.0	70.5	8.0	100.0	8.0

	[ha]	517 858	2 392	520 249	12 956	533 206
					56	951
					14.5	2.9
Sachsen	Prim	943	12	946		
	SE [%]	2.9	28.8	2.9		
Sachsen-Anhalt	[ha]	493 920	9 067	502 987	29 494	532 481
	Prim	1 829	79	1 845	264	1 884
	SE [%]	2.2	12.1	2.1	6.3	2.1
Schleswig-Holstein	[ha]	168 426	199	168 626	4 787	173 412
	Prim S	775	2	776	45	778
	E [%]	3.8	70.7	3.8	15.2	3.7
Thüringen	[ha]	520 944	2 799	523 743		
	Prim	895	14	902	25 345	549 088
					118	912
					9.4	2.6
	SE [%]	2.7	26.6	2.7		
Hamburg + Bremen	[ha]	13 054	---	13 054	791	13 846
	Prim	152		15	2	15
	SE [%]	5.6		25.6	70.4	25.8
Germany (all Lander)	[ha]	11 012	41 742	11 054 162	364 962	11 419 124
	Prim	420	221	20 885	1 778	21 040
	SE [%]	20 844	8.0	0.7	2.9	0.7
		0.7				

Ejemplo de una tabla de contingencia que ofrece la superficie forestal por Estado Federal en Alemania ("Territorio") desglosada en diferentes categorías de tierras forestales. Esta tabla ha sido elaborada a partir de la [herramienta en línea del IFN alemán](#) (en inglés); no sólo presenta el área estimada, sino también el error estándar relativo estimado SE% y el número de conglomerados = unidades primarias de muestreo (que corresponde al tamaño de la muestra por unidad subnacional) que cayeron en las combinaciones de Estado Federal y tipo de tierra forestal. Aquí se aprecia claramente que la precisión de la estimación es función del tamaño de la muestra.

Recuerde que cuando se desglosan las áreas, todos los criterios de desglose deben estar claramente definidos para que los resultados se puedan interpretar correctamente según estas definiciones: hay que definir claramente "bosque" y contrastarlo con lo no forestal, y hay que tener criterios claros para distinguir dentro del bosque las diferentes clases de "degradación", "tipos de bosque", "tipos de gestión", etcétera.

Además, es importante tener en cuenta que no todas las categorías se pueden identificar sobre el terreno o utilizando imágenes de teledetección. En algunos casos, estas categorías deben extraerse de

documentos oficiales. Por ejemplo, la propiedad y el estado de protección se deben extraer de los mapas catastrales y de los mapas de áreas protegidas, respectivamente.

Estimaciones de características por área, incluyendo: volumen/biomasa/reservas de carbono por hectárea, número de árboles por hectárea, número de árboles grandes, densidad de regeneración, existencias de madera muerta en diferentes clases de dimensión, etc.

De nuevo, podemos ilustrar esto con un ejemplo de una tabla unidireccional que da la masa total en biomasa y carbono en Bután (DFPS. 2019). También se presenta el error estándar relativo (calculado a partir del intervalo de confianza) para cada distrito.

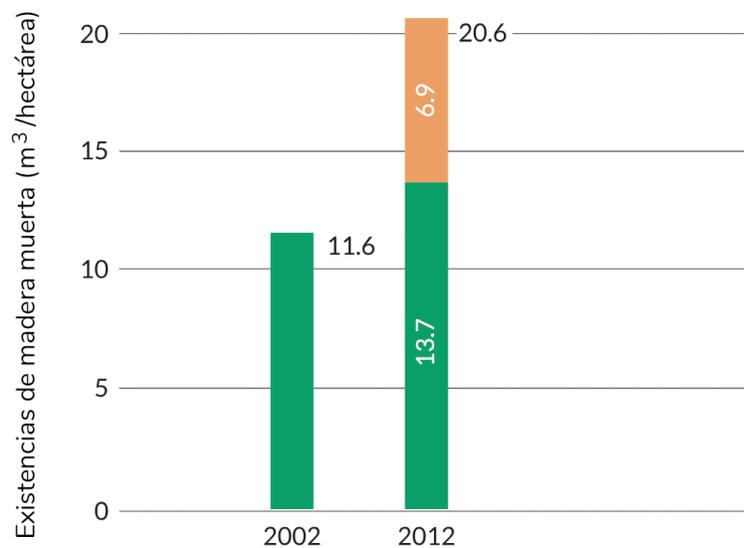
Distrito	Biomasa (millones de toneladas)	MoE (%)	Carbono (millones de toneladas)	MoE (%)
Bumthang	80 ± 16	20	37 ± 7	20
Chhukha	91 ± 21	23	43 ± 10	23
Dagana	50 ± 8	15	24 ± 4	15
Gasa	7±2	31	3 ± 1	31
Haa	57 ± 10	18	27 ± 5	18
Lhuntse	77 ± 19	24	36 ±9	24
Mongar	95 ± 18	19	45 ±9	19
Paro	30 ± 8	27	14 ±4	28
Pemagatshel	18 ± 4	21	8±2	21
Punakha	54 ± 9	36	25 ± 9	36
Samdrup Jongkhar	72 ± 14	20	34 ± 7	20
Samtse	20 ± 5	22	10 ± 2	22
Sarpang	37 ± 6	17	18 ±3	17
Thimphu	48 ± 25	51	23 ± 12	52
Trashigang	96 ± 16	16	45 ± 7	16
Trashiyangtse	41 ± 20	48	19 ± 9	48
Trongsa	52 ± 11	21	25 ±5	21
Tsirang	29 ± 11	39	14 ± 5	39
Wangduephodrang	91 ± 16	18	43 ± 8	18
Zhemgang	56 ± 7	13	26 ±4	13

Estimaciones de los cambios en las variables objetivo, cuando los análisis se refieren a inventarios repetidos. Al analizar estos resultados, es importante observar si las definiciones pueden haber cambiado. Puede ocurrir que el análisis revele un cambio, pero que parte del cambio se pueda atribuir a cambios en las definiciones.

Un ejemplo es el de los cambios en las estimaciones de existencias de madera muerta en el IFN alemán entre 2002 y 2012. El análisis arrojó un cambio muy grande en las estimaciones de las existencias de madera muerta. Parte de este inesperado gran cambio en las estimaciones se debió a la adaptación del diámetro mínimo de las piezas de madera muerta registradas, que anteriormente era de 20 cm, a la norma internacional del IPCC de 10 cm.

En este caso, se puede analizar fácilmente qué parte del cambio se puede atribuir al cambio de definición, porque toda la información necesaria estaba en los datos (para aplicar la antigua definición sólo tenían que dejar fuera todos los trozos de madera muerta con un diámetro inferior a 20 cm). Esto resultaría más difícil a la hora de analizar otras modificaciones de las definiciones, como el cambio de la cobertura mínima de copas en la definición de bosque.

Es importante entonces que el análisis deje claro cuáles son los componentes de estos cambios: en este caso, manteniendo la antigua definición, el cambio sería un aumento de 2,1 m³/ha de 11,6 m³ a 13,7 m³, pero en el gráfico se muestra como 4 veces mayor (9 m/ha de 11,6 m³ a 20,6 m³) debido a la modificación hacia una definición más inclusiva (BMEL 2014).



Si el análisis también produce **estimaciones sobre los árboles fuera del bosque**, deberá referirse a

diferentes tipos de uso de la tierra no forestales, presentando esencialmente los resultados por superficie, pero para los tipos de uso de tierras no forestales.

Es importante reiterar aquí que, para todas las estimaciones puntuales, el análisis también debe producir estimaciones de intervalo (error estándar o intervalos de confianza), de modo que el análisis muestre el estatus de la estimación, así como la incertidumbre de dicha estimación.

Por supuesto, los intervalos de confianza también son de interés crucial para las estimaciones de cambio. Si el valor cero está contenido dentro de los intervalos de confianza superior e inferior, se puede suponer que los cambios no son significativos. Aquí, por supuesto, a la hora de obtener afirmaciones sobre la significación estadística, la interpretación de los intervalos de confianza debe tener en cuenta que en los IFN se suele utilizar el muestreo sistemático.

Mapas

Los mapas se utilizan con frecuencia para presentar los resultados de los IFN y, a menudo para muchas personas no expertas, resultan más convincentes que las estadísticas. Los mapas continuos de cobertura total sólo pueden elaborarse cuando se han analizado las imágenes de teledetección de toda la superficie y se han desarrollado los respectivos modelos.

Los mapas forestales/no forestales son un producto base que, acompañado también de mapas de biomasa, son de gran interés. Al igual que ocurre con todos los demás productos del inventario forestal, los analistas del inventario deben enfatizar que los mapas pueden presentar imprecisiones, al igual que todos los productos pertenecientes a estudios empíricos. En consecuencia, estas incertidumbres se deben documentar e informar adecuadamente junto con los mapas.

Si no se utilizan datos de teledetección en un IFN, sólo se pueden generar mapas con la resolución espacial de la cuadrícula de muestreo sistemático utilizada. Estos mapas no pueden producir una representación continua de una variable objetivo, sino que sólo dan una idea aproximada de las distribuciones espaciales a una escala bastante grande. Por lo general, se da una información por punto muestral, a veces a varios kilómetros de distancia

Utilización de los datos de los IFN para la optimización del diseño de futuros IFN

El análisis de los datos de los IFN también puede ser beneficioso para la optimización del diseño del inventario en la planificación de inventarios posteriores. Por supuesto, al hacerlo, hay que ser cauteloso

a la hora de cambiar el diseño de los IFN entre ciclos posteriores, ya que la coherencia es necesaria en las series temporales. Aun así, las adaptaciones del diseño realizadas de vez en cuando pueden aumentar la eficiencia y, con frecuencia, es necesario integrar nuevas variables objetivo para que los IFN puedan responder de forma significativa a los nuevos problemas que surjan.

También puede optimizarse el consumo de tiempo introduciendo nuevos dispositivos de medición. Las mediciones de control se pueden reorganizar y las precisiones objetivo se pueden volver a definir. Además, es importante señalar que las nuevas tecnologías pueden implicar reducciones en el tamaño de los equipos de campo.

En este contexto, puede ser interesante comprobar si una reducción del número de subparcelas en un diseño de conglomerados daría lugar a una reducción significativa de la precisión de la estimación. Puede resultar que para algunas variables objetivo sea suficiente un número menor de subparcelas. Este análisis de optimización se puede llevar a cabo fácilmente realizando el análisis de datos para un número menor de subparcelas por conglomerado, reduciendo así el tamaño de la parcela por unidad de muestreo seleccionada.

Uso de los datos de los IFN en el mundo académico

La principal tarea (y por defecto) del análisis de los datos de los IFN es generar los resultados básicos que las partes interesadas y los responsables de la toma de decisiones exigieron en la evaluación de las necesidades de información. Los datos de los IFN son una fuente para otros fines como la investigación y la enseñanza académica; a veces, los IFN son los únicos proyectos que generan una base de datos científicamente sólidos sobre bosques o territorios de un país.

Se pueden analizar muchos temas a partir de los datos de los IFN, incluidas las cuestiones metodológicas (como las evaluaciones sobre la optimización de los diseños de los inventarios o sobre la aplicación o adaptación de modelos) y las cuestiones temáticas (cuestiones de rendimiento, comparación de la composición y ubicación de las especies, causas de la pérdida de bosques, etc.). Estos tipos de análisis no son una tarea habitual del equipo de inventario, pero requieren que las bases de datos se pongan a disposición de los investigadores.

La investigación académica que utiliza los datos de los IFN contribuye también a formar a futuros expertos en IFN, de modo que los planificadores y analistas de los IFN deberían fomentar de forma proactiva el uso de los datos de los IFN en el mundo académico, en la investigación y en la enseñanza

(Liang y Gamarra 2020). En particular, en los sistemas de monitoreo forestal a más largo plazo se dispone de valiosas series temporales que permiten monitorear el desarrollo forestal y la sostenibilidad de las políticas forestales nacionales.

Existen estudios de investigación que utilizan **datos de IFN repetidos para actualizar tablas de rendimiento** (por ejemplo, Staupendahl y Schmidt 2016) o que **identifican la idoneidad de las especies arbóreas bajo el cambio climático** (por ejemplo, Prasad et al., 2020).

Dos ejemplos de estudios de investigación específicos son la **estimación de la longitud del borde de los bosques** a partir de los datos de los IFN alemanes a diferentes escalas que el diseño de la parcela ofreció en Kleinn et al., 2011; y la **comparación de las existencias de árboles fuera del bosque (TOF)** de 12 IFN apoyados por la FAO en el Sur del Mundo (Schnell et al., 2015).

Los investigadores pueden estar interesados en los conjuntos de datos de grandes superficies y a largo plazo de los programas nacionales de monitoreo forestal cuando deseen examinar cuestiones metodológicas o temáticas específicas. Por ejemplo, las muestras sistemáticas en conglomerados contienen información sobre la fragmentación del paisaje: si los bosques están más fragmentados, habrá más intersecciones con los conglomerados y un número menor de conglomerados estará totalmente contenido dentro o fuera del bosque, sin intersecciones. A partir del número y la proporción de conglomerados que se intersectan, se pueden derivar estimaciones del estado general de fragmentación -como, por ejemplo, las presentadas por Kleinn (2000) para un inventario de gran superficie en Costa Rica- o estimaciones de la longitud del borde del bosque para toda la región o unidades de referencia subnacionales, como se hizo en Kleinn et al., (2011) para Alemania.



¿Sabías que?

En una conferencia impartida por uno de los autores, un estudiante planteó la pregunta de si las existencias en crecimiento son mayores en las clases de mayor pendiente porque las copas de los árboles reciben más luz y la superficie es mayor que en el terreno plano. El profesor tenía acceso a la base de datos del IFN alemán, que permite realizar análisis flexibles combinando variables.

Combinando las existencias en crecimiento como variable de respuesta y las clases de pendiente como categorías, se pudo elaborar rápidamente el siguiente gráfico, que permite dar una respuesta

preliminar a esta pregunta.



Existencias en crecimiento [m³/ha] sobre clases de pendiente [%] - gráfico elaborado rápidamente a partir del análisis de los datos del IFN alemán en respuesta a la pregunta de un estudiante: ¿Tienden las existencias en crecimiento a ser mayores en las pendientes más pronunciadas (al menos hasta un límite superior de pendiente) porque la superficie del terreno es mayor?

Principales características de los productos del análisis de datos

Las principales características que caracterizan al análisis de datos son esencialmente las mismas que también caracterizan la elaboración de informes y que se han formulado en términos generales como principios rectores en el Marco de Transparencia Reforzado (CMNUCC 2020): **transparencia, exactitud, coherencia, exhaustividad, comparabilidad**, y todo ello manifestado en una documentación integral. Es esencialmente la combinación de estas características lo que hace que el análisis de los datos de los IFN sea creíble para las partes interesadas y los usuarios de los datos.

Siempre hay que ser consciente de que los resultados de los IFN entran en el ámbito de las políticas forestales de un país y de que hay diferentes intereses en juego: puede que no todas las partes interesadas estén satisfechas con los resultados por múltiples razones.

Por lo tanto, es de suma importancia que:

- el análisis de los datos sea "impermeable" y correcto y pueda defenderse sobre la base de una documentación completa y transparente; y
- la interpretación de las conclusiones sea compatible con los resultados de los análisis. La

interpretación de los distintos actores puede entonces variar bajo los mismos resultados y estadísticas, basándose en ideas y valores particulares de distintos grupos de interés, pero esto queda fuera del alcance del analista de datos.

La función del análisis de datos en la elaboración de informes sobre inventarios forestales

Probablemente ya haya quedado claro que el análisis de datos es anterior a la elaboración de informes. El análisis de datos tiene lugar entre la recopilación y gestión de los datos y la elaboración de los informes. Por lo tanto, a la hora de realizar el análisis de datos es importante tener en cuenta ambos aspectos: de dónde provienen los datos (recopilación/gestión de datos) y cómo se pretende utilizar y procesar los resultados del análisis (elaboración de informes).

Por ello, es imperativo que el análisis de datos y la elaboración de informes estén estrechamente interrelacionados y, si hay distintos expertos trabajando en estos ámbitos, es necesario que colaboren estrechamente. De este modo, los resultados iniciales e intermedios podrán debatirse, interpretarse y compararse para detectar posibles incoherencias en una fase temprana. Esta sería especialmente útil, ya que tales incoherencias pueden muy bien ser expresión de resultados inesperados o, más sencillamente, estar causadas por errores en los cálculos o en la recopilación de datos.

Resumen

Antes de finalizar, aquí están los puntos clave de aprendizaje de esta lección:

- Tener una idea clara de los potenciales productos de los análisis de datos de los IFN en la fase de planificación y durante la evaluación de las necesidades de información (ENI) siempre ayuda.
- La información del análisis de datos forestales puede servir para generar estadísticas estándar y para generar mapas basados en la teledetección y también para optimizar futuros diseños del IFN, y para la investigación y el mundo académico.
- Los mapas son una presentación común y convincente de los resultados de los IFN; a menudo y para muchos más fácilmente accesibles y más convincentes que las estadísticas.
- El análisis de datos debe tener en cuenta tanto la fuente de datos (como parte de la recopilación/gestión de datos) como el lugar en el que se utilizarán y procesarán los resultados del análisis (elaboración de informes).